Is only one of my selves authentic? An empirical approach

Sebastian Krügel, Matthias Uhl [*]

Technische Hochschule Ingolstadt, Germany

ARTICLE INFO	ABSTRACT
	In behavioral economics, intrapersonal conflict is predominantly interpreted hierarchically. A "present-biased" intrapersonal doer may spoil the goal of a rational planner. This often suggests paternalistic interventions that help the "true" or "authentic" self to overwhelm its present-biased alter ego. Game theorist Schelling proposed a reciprocal interpretation of intrapersonal conflict that interprets both selves as strategic players, which Elster contradicted by claiming that in any conflict only one self is capable of strategic behavior and therefore authentic. Previous empirical studies, however, cannot test this interpretation, because their design provides commitment devices unilaterally to only one self. In an experiment, we provided commitment devices to both selves and find similar inclinations to use this strategic tool. Given this, the symmetric view on intrapersonal conflict seems no less plausible than the hierarchical one. Our results might contribute to a richer debate on intrapersonal conflict by feeding in some skepticism about the self-evidence with which paternalists take sides.

1. Introduction

Hard and libertarian paternalists alike assume that individuals make inferior choices that are caused by a lack of willpower (Thaler and Sunstein, 2003). Examples for this very mundane behavior seem to abound. Think of a dieter giving in to a dessert or of a saver making a pleasure purchase, of a student skipping classes to watch a movie or of an early morning jogger sleeping late. Are these choices the revealed proclivities of a present-biased intrapersonal self that overrides the informed preferences of the "true" or "authentic" intrapersonal self? This view represents the most prominent behavioral economic interpretation of intrapersonal conflict: present-biased preferences (or hyperbolic discounting, more generally). Although the decision maker is represented by a single utility function, the usual narrative is that the intrapersonal planner forms a plan, and the intrapersonal doer spoils the plan by succumbing to temptations along the way. This modelling philosophy is exemplified by the works of Strotz (1955) and Laibson (1997).

Game theorist Thomas Schelling was more critical about concepts of the "true" or "authentic" self (see, e.g., Schelling 1984a,b). He transferred the skepticism of modern economics about soci(et)al welfare judgments to the conflicted individual with "alternating preferences" (Schelling 1984a,b). Schelling made the point that our intuitions about authenticity might change if we interpret intrapersonal conflict

differently. Schelling's suggestion was to consider both selves in situations of intrapersonal conflict as equally authentic with conflicting but equally legitimate interests. Hard and libertarian paternalists might differ in the means that they are willing to employ to help weak-willed individuals reach their "true" goals, but they agree with self-confidence on *for whom* to take sides.[1] For agnostics such as Schelling, this confidence is rather an ideological than a rational one. Formally, this philosophy is expressed in the work of Karni et al. (1983) on state-dependent preferences that permit that the decision maker's preferences are not independent of the prevailing state of nature. In this model, behavior in all states is locally rational. Sugden (2018) argues that many behavioral "anomalies" have a context-dependent structure. In the Allais paradox, for instance, there is an inconsistency in the choice that people make between two gambles (Allais, 1953). Yet, there is no rational answer to the question of which of the two gambles one should prefer. The same, Sugden (2018) argues, applies to systematically fluctuating choices that people make in different states of nature.

Jon Elster (1985) argued that the question of the "true" or "authentic" self is ultimately empirical. For him, the authentic self (i.e., the self that we should side with as friends or policy makers) is the self that engages in forward-looking and strategic behavior. Ironically, however, Elster (1985) found it sufficient to decide this empirical question by "introspective evidence". For him, in any domain, there is

[1] At least some people may be less confident about paternalists' superior decisions. Arad and Rubinstein (2018) found that a substantial proportion of study participants prefer the government to simply provide information than to implement libertarian-paternalistic policies like choice-optimizing architectures.

only one self that is capable of acting strategically in binding its short-sighted alter ego. The other self is a lack of willpower and therefore not even a proper "self" as it is only negatively defined and not even attributed its own will. We argue that Schelling's intuition of a conflict between two emancipated selves and Elster's counterintuition of exclusively hierarchical conflicts between a forward-looking (planning) self and a present-biased (doing) self should be put to a more literal empirical test. This is all the more important, because previous evidence on commitment choices only *seemingly* suggests the strategic superiority of a planning self over a present-biased doing self. This is because the respective studies predetermine intrapersonal conflict as a strategic one-way street (see, e.g., Ariely and Wertenbroch, 2002; Ashraf et al., 2006; DellaVigna and Malmendier, 2006). Their design does simply not allow to identify anything but a one-way conflict (or no conflict at all). In this sense, part of the reason that we tend to interpret one self as a strategic planner and the other self as a present-biased doer might be due to the fact that previous evidence stems from the asymmetric provision of commitment devices to only one of the competing selves. How can the seemingly "present-biased" self get rid of its disparaging label, if it is never given the chance to act strategically?

Consider, for instance, the case of saving and spending money. Experimental evidence shows that the saver voluntarily restricts future options to bind his or her overspending alter ego (Madrian and Shea 2001; Ashraf et al. 2006). Prima facie, this emphasizes the forward-looking rationality of the saver and the spender's lack of will-power. Moreover, this may be accommodated by our daily intuitions. Saving is usually seen as a forward-looking virtue, spending as a myopic vice. But what about very economical people? In stark contrast to the just cited commitments against overspending, Kivetz and Simonson (2002) found that consumers who have problems in choosing items that are considered as indulgences over necessities in daily decisions use commitments to indulge. Their participants chose hedonic luxury goods over cash of greater value and explained this by their need to commit to indulge to make sure that the cash does not end up in the pool of money used for plain necessities. Anecdotal evidence for "commitments to indulge" may be provided by workaholics who pay a holiday well in advance in order not to call it off later or who intentionally leave their mobile at home when driving to the lake with their families. Kivetz and Simonson (2002) note that the very same consumer may feel a need to exercise commitment as a means for self-control in one domain (like smoking), while forcing herself to indulge in another domain (like going to a good restaurant). Although the authors provide inspiring evidence for commitments to preferences that are usually attributed to the present-biased self, they did not test whether diametrically opposed or reciprocal self-commitment can actually be shown in the very same domain.

This is precisely the objective of the present paper. First, we seek to observe alternating preferences in the same domain, where none of the preferences can be attributed to a present-biased self in the usual economic sense. Moreover, we give both selves the opportunity to commit to their preferences and explore the extent to which this opportunity is used in each case. Thus, we put Schelling's idea of a reciprocal intrapersonal conflict between two emancipated selves that engage in a mutual strategic conflict without present bias to an experimental test. To the best of our knowledge, we are the first to equally provide commitment devices to both selves in a given situation of intrapersonal conflict. In other words, we are the first to test for intrapersonal conflict outside the realm of time-inconsistent preferences and within the realm of state-dependent preferences. If both selves use commitment in a given situation of intrapersonal conflict, the nominal distinction between the "forward-looking" (planning) self and the "present-biased" (doing) self becomes empty, because both selves could then equally be interpreted as rational and thus strategic selves. This is an interpretation that has hardly had a chance to prove itself empirically. We believe, however, that this is more than an academic exercise. Providing evidence in support of this interpretation might contribute to a more impartial

stance toward intrapersonal conflict and consequently more restraint in intervening in consumer sovereignty.

Our paper proceeds as follows. In Chapter 2, our experiment design is explained. Results are discussed in Chapter 3. Chapter 4 concludes.

2. Experiment

2.1. Preliminary considerations

To be able to distinguish a reciprocal intrapersonal conflict between two emancipated selves from a hierarchical conflict à la present-biased preferences, more than two temporally separated periods are necessary. Consider, for instance, getting up early in the morning, which is often perceived as a typical self-control problem in terms of present-biased preferences. The strategic component is in the control of the evening self, which in wise foresight puts the alarm clock out of reach to force the present-biased morning self out of bed. The evening self plans and the present-biased morning self acts.

To make a reciprocal conflict between two emancipated selves visible to an external observer, the morning self would need a commitment device as well. However, the opportunity for self-commitment is of little help on the day when the morning self is forced to get up. It needs the commitment device prior to that day. This dissolves the categorization of selves into forward-looking planner and present-biased doer, and we might, instead, observe a conflict between two forward-looking planners within a person.[2]

Fig. 1 illustrates our proposed test strategy for a mutual conflict between two authentic selves. The upper line represents a conflict in the usual two-period setup. Consumption takes place in period $t = 2$. For the sake of simplicity, we assume that there are only two goods, "A" and "B," one representing an alleged vice and the other a virtue. For instance, good A might be a salad and B a pizza when choosing food for lunch in a cafeteria; or A might correspond to getting up early and B to sleeping in. At $t = 2$, it shall hold that $B \succ A$, and at some earlier period $t = 1$, it shall be the reverse, $A \succ B$. In the spirit of present-biased preferences, a person who commits at $t = 1$ to the consumption of good A instead of B in $t = 2$ is sophisticated and correctly anticipates his or her weakness in the face of temptation. The true preference is revealed in $t = 1$, and the reversal of preference at $t = 2$ is merely due to a bias for the present. Schelling's idea of the coexistence of several authentic selves would be an alternative interpretation. However, the two approaches cannot be distinguished in such a situation.

To grant the preference at the time of consumption a right to authenticity in Elster's sense, the self with this preference would need an opportunity to plan ahead. This is possible only if both selves appear at least once prior to the time of consumption and can mutually restrict each other. Such a situation is illustrated by the middle and bottom lines in Fig. 1. Consumption still takes place in period $t = 2$. However, now, there are two preceding periods in which the consumption in $t = 2$ can be decided upon. In period $t = 0$, a choice between goods A and B for $t = 2$ is made and in period $t = 1$ this choice can be revised. The commitment device now consists in the possibility that the person in period $t = 0$ can deprive him or herself of the opportunity to revise the plan in $t = 1$ in which no present-biased consumption takes place.

Like the upper line in Fig. 1, the middle line illustrates a commitment of self 1, with $A \succ B$, at the detriment of self 2, with $B \succ A$. However, in this case, the commitment is directed against the planner at time $t = 1$

---

[2] An anonymous reviewer raised the thought-provoking question of how it should be interpreted if the early-morning self in our example destroys the ringing alarm clock in the morning. Would this be a manifestation of "strategic" thinking or a lack thereof? The answer depends on one's own attitude toward intrapersonal conflict and we leave it to the reader to decide. However, if one is inclined to attribute it to a lack of strategic thinking, one must also ask oneself what the early-morning self would have to do to prove its own authenticity.
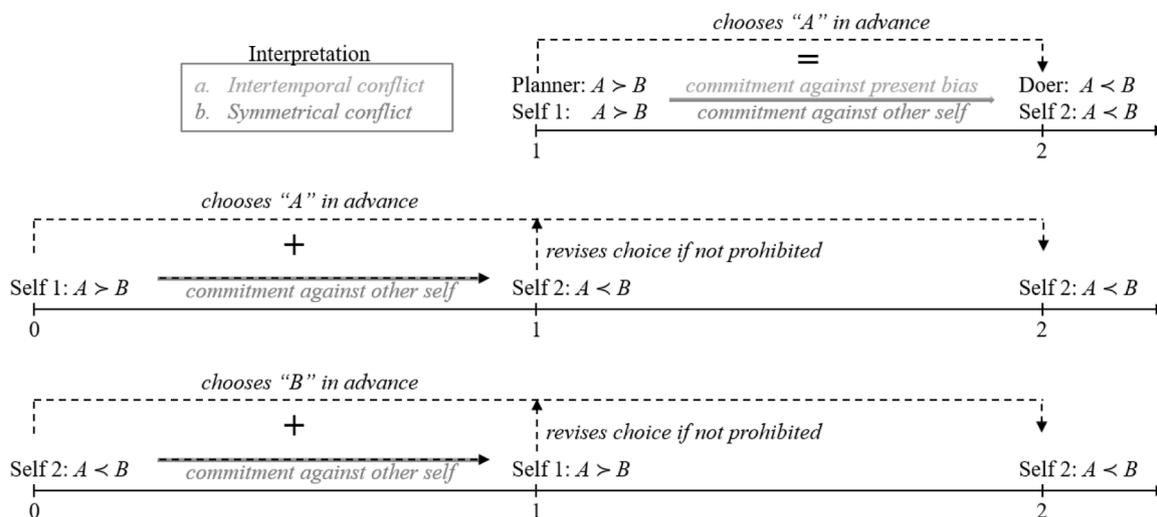
Fig. 1. Comparison of analysis foci between studies on intertemporal preferences or weak will and the present study
Notes: Consumption takes place in period *t = 2*. Dotted lines symbolize decisions in the respective periods. The upper timeline illustrates an intrapersonal conflict in a two-period setup. Here, the planner or self 1 makes a single choice, namely the binding decision regarding consumption in *t = 2*. In the middle and lower timelines, the self in *t = 0* makes two decisions: one regarding consumption in *t = 2* and one commitment decision regarding the self in *t = 1*. The self in *t = 1* can revise the consumption choice of the self in *t = 0* unless prohibited by the self in *t = 0*.

and not against the doer in *t = 2*. The question now is whether self 2 is also capable of planning ahead and making use of the opportunity for self-commitment. This is shown by the bottom line, where the roles of selves 1 and 2 are reversed. Here self 2 is in the position to take measures against self 1 to protect the enforcement of its own preferences in *t = 2*.

2.2. Design

Subsequently, we will explain the design of our experiment. The general setup was based on the reasoning outlined in Section 2.1. The good about which we wanted to observe an intrapersonal conflict was early-morning sleep. The attempt to address different selves within our participants was based on the assumption that people rate the value of early-morning sleep differently depending on the time of day. Specifically, we assumed that people have a greater desire for sleep in the early morning, when they are fatigued, than in the late morning, when they are rested. The experimental manipulation therefore consisted of asking participants about their value of early-morning sleep at different times of the day. The experimental setup required the participants to be informed of the entire procedure prior to making their first decision. Neither deception nor any surprises were involved in the experiment.

2.2.1. Course of the experiment

The course of the experiment and implemented treatments is depicted in Fig. 2. The experiment took place over four consecutive Mondays in November 2014. Every participant had to attend one session each Monday. The sessions on the first and last Mondays took place at the experimental laboratory of the Max Planck Institute of Economics in Jena, Germany. The sessions on the second and third Mondays were conducted online, and the participants could take part in those sessions from home or wherever they had Internet access.

The sessions on the first Monday were used for general instructions and questions (*briefing sessions* in the following). The sessions on the last Monday were used to pay our participants (*payment sessions* in the following). There were two payment sessions: one between 6:15 and 7:15 a.m. (*early payment session* in the following) and one between 12:00 and 4:00 p.m. (*late payment session* in the following). Each participant had to come to exactly one payment session. Whether this was the early or late payment session was announced at the end of the second online session (i.e., on the third Monday of the experiment).

Participants who had to show up to the late payment session earned a

fixed amount of 10 euros for attending the experiment. Participants who had to show up to the early payment session earned between 0 and 40 euros. Participants who came too early had to wait until the relevant time frame started. Participants who came late were sent away without payment. Participants who did not show up at the laboratory during the relevant time frame did not earn anything from the experiment. In both payment sessions, the participants collected their earnings without making any further decision. Thus, once a participant was paid, he or she could leave the laboratory immediately.

2.2.2. Decisions of the participants: reservation wages and commitment

All relevant decisions were made during the online sessions (i.e., on the second and third Monday). In each online session, the participants were asked about their reservation wage for coming to the early payment session instead of the late one. We employed the Becker et al. (1964, "BDM") mechanism to elicit participants' reservation wages. That is, the participants had to state the minimum payment in euros (a whole number between 0 and 41) they wanted to receive for coming to the laboratory during the early instead of late payment session. For each participant, the minimum payment was compared to a random number drawn from a discrete uniform distribution between 0 and 40. If the random number was equal to or higher than the minimum payment, then the relevant participant had to come to the early payment session and his or her earnings were equal to the random number. If the random number was lower than the minimum payment, the relevant participant had to come the late payment session and his or her earnings were the fixed amount of 10 euros.[3]

Since all of the participants had to attend two online sessions, they were asked for their reservation wage twice. For each participant, one of the two reservation wages was chosen at random to be relevant in the BDM procedure. However, the participants could decide whether the reservation wage from their first or second online session would be more likely to be relevant. Specifically, they could choose between the following two options: (i) the reservation wage from the first online session will be relevant with probability 0.8, or (ii) the reservation wage

---

[3] Thus, reservation wages were truncated at 0 and 41. A minimum payment of 0 ensured that the participant had to come to the early payment session. A minimum payment of 41 ensured that the participant had to come to the late payment session.
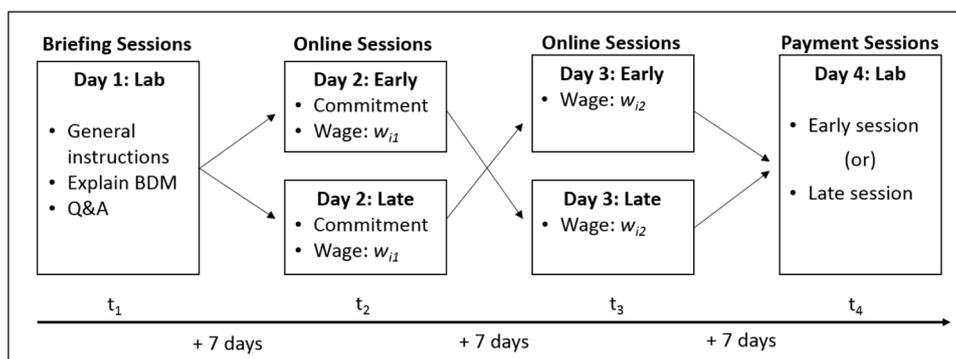
Fig. 2. Course of the experiment.

from the second online session will be relevant with probability 0.8. This decision was made once, namely during the participants' first online session prior to stating their reservation wage.

Notice that this decision granted each participant a stochastic commitment device.[4] Assigning a probability of 0.8 to the second reservation wage was the no-commitment or flexibility option, under which a participant postponed the decisive decision to the second online session. Assigning a probability of 0.8 to the first reservation wage, on the other hand, was the commitment option, under which a participant preponed the decisive decision to the first online session.[5] The stochastic commitment device ensured incentive-compatible statements of reservation wages in both online sessions, independent of the commitment choice.

Regarding reservation wages, we would like to stress two important aspects of our experimental design. First, because reservation wages were the minimum payment for coming to the early *instead* of the late payment session, they can be decomposed into the fixed amount of 10 euros for coming to the laboratory during the late payment session and a markup for doing this task in the early morning.[6] On the last day of the experiment, however, the participants could no longer choose whether they wanted to come to the early or late payment session. The relevant payment session was determined and announced at the end of the second online session. Thus, on the last day of the experiment, the participants could only choose between collecting their earnings during the relevant time frame or not coming at all and being paid nothing. The stated reservation wages were therefore not necessarily equal to the minimum payment required for coming to the early payment session per se. For instance, a participant could demand 20 euros for coming to the early *instead* of the late payment session. However, the same participant might be willing to come to the early payment session for less than 20 euros if the late payment session is no longer available. This design feature rendered possible the visibility of a conflict of selves, where one planning self imposes a lower markup for getting up early, knowing that the future, performing self will still come to the early payment session once the late payment session is excluded from the choice set.

Second, the experimental design assured that present-biased preferences could not affect reservation wages because we did not have an

intertemporal choice problem (see in this context also footnote 10 in O'Donoghue and Rabin, 1999). Participation in the online sessions was mandatory, and in each online session, the participants were asked about their future value of getting up early (where future refers to the last day of the experiment). The costs and rewards for getting up early occurred in the same (future) period, and therefore, the elicitation of reservation wages did not involve any kind of intertemporal trade-off where discounting might have played a role.[7] This made it possible to observe self-commitment in a context other than intertemporal decision-making.

2.2.3. Treatments: early-late and late-early

The online sessions took place within two different time frames: either between 5:00 and 7:00 a.m. (*early online session* in the following) or between 10:00 a.m. and noon (*late online session* in the following). This represents our experimental manipulation to induce state-dependent preferences based on the assumptions that people rate the value of early-morning sleep differently depending on the time of day. Each participant had to attend in one early and one late online session. We implemented two treatments. Half of the participants had to do the early online session first and the late online session the following week (*Early-Late* treatment in the following). The other half of the participants had to do the late online session first and the early online session the following week (*Late-Early* treatment in the following).

One day prior to each online session, the participants received an e-mail with a personalized link to take part in the session. The link was only enabled during the relevant time frame on the following day. Of course, the participants were informed or reminded about that time frame in the e-mail containing the link.[8] During an online session, they could progress at their own pace. On average, it took the participants approximately 5 minutes for the first and 2.5 minutes for the second online session. However, a session had to be finished within the specified time frame. Participants who failed to do so were excluded from the experiment and did not earn anything.

---

[4] To the best of our knowledge, a stochastic commitment device was implemented first in Uhl (2011) and also used in, e.g., Augenblick et al. (2015).

[5] In the experiment, the two options were presented neutrally, without any labels. Also, on the computer screen, the flexibility option was always presented above the commitment option.

[6] The markup can be positive, negative, or equal to zero. A participant who prefers the late over early payment session should state a reservation wage greater than 10 euros (i.e., a positive markup). A participant who prefers the early over the late payment session should state a reservation wage less than 10 euros (i.e., a negative markup). A participant who is indifferent between the early and late payment session should state a reservation wage of precisely 10 euros (i.e., a markup of zero).

[7] Again, both payment sessions took place on the same day: the early payment session between 6:15 and 7:15 a.m., and the late payment session between 12:00 and 4:00 p.m. Thus, there was a time lag of a few hours between payment sessions. Time preferences were not relevant to the elicitation of reservation wages as long as participants ascribed both payment sessions to one and the same consumption period. Considering that the online sessions took place one and two weeks prior to the payment sessions, we believe this assumption is innocuous.

[8] One day prior to their first online session, the participants learned about the time frame for the first time in the e-mail. One day prior to their second online session, the participants were only reminded about the time frame in the e-mail because they had been informed about the entire procedure of the experiment at the beginning of the first online session.

2.2.4. Briefing sessions

There were six 30-minute briefing sessions with approximately 32 students in each. All of the briefing sessions took place in the afternoon of the first Monday of the experiment. In those sessions, each participant received written instructions about the general procedure of the experiment, and as usual, the instructions were read aloud.[9] The participants were reminded that they had to attend one session on each of the three upcoming Mondays to be eligible for their earnings. However, they were not yet informed about the exact time of those sessions. They only knew that they would be informed about the time of the first online session in an e-mail one day prior.

In addition, the instructions informed the participants that the experiment was for them to perform one of two tasks. The specific tasks would be outlined at the beginning of the first online session. In the briefing sessions, the two tasks were labeled neutrally as task A and task B. Participants were told that they would receive a fixed amount of 10 euros for performing task A and a random amount between 0 and 40 euros for task B. Moreover, they were informed that they would be asked about their minimum payment for doing task B instead of task A. The implemented BDM procedure was explained to them carefully and at length. In addition, the BDM procedure was demonstrated several times with a real urn containing numbered balls from 0 to 40 and hypothetical tasks A and B. Afterwards, the participants could ask questions, which were answered publicly. Once all questions were answered, the briefing session was over.

2.2.5. Procedural details

The participants were students of all majors from Friedrich Schiller University and the University of Applied Sciences in Jena, Germany. Students were invited via ORSEE (Greiner 2004). The invitation e-mail made clear that payment for the experiment required participation in four sessions. The invitation only specified the date of each session. Since we wanted to reduce selection effects as much as possible, the students were not informed about the exact time of the sessions on the second, third and fourth Monday. Students were informed, however, that they should only participate in the experiment if they had no appointments before noon on each day of the sessions.

Besides possible selection effects when inviting participants, we were also concerned about the dropout rate.[10] To reduce the dropout rate, we added a lottery drawing for all participants who took part in a session on each of the first three days of the experiment. Out of all participants, five were randomly selected to win an additional prize of 50 euros each.[11] The lottery drawing was announced in the invitation e-mail.

In total, 188 students took part in the briefing sessions. We had seven dropouts before the first online session (five in treatment Early-Late and two in Late-Early) and another five before the second online session (two in treatment Early-Late and three in Late-Early). Thus, of 188 students who started the experiment, 176 completed both online sessions, which implies a dropout rate of only 6.4%. The mean age of those 176 students was 24.5 years, roughly 41% of them were male, and the average payment was approximately 20 euros (excluding the lottery win of 50 euros

for five participants).[12]

3. Results

3.1. Reservation wages

To study an intrapersonal conflict outside the realm of present-biased preferences empirically presupposes the successful triggering of state-dependent preferences (or "alternating" preferences in Schelling's wording). In the experiment, we assumed that people value sleep differently depending on the time of day and that their valuation of sleep is greater in the early than in the late morning. Indeed, this seems to be the case. Participants' reservation wages for coming to the early instead of the late payment session were affected systematically by the time of day when they were asked (see Table 1). Reservation wages in the first online sessions were, on average, significantly higher in the Early-Late treatment than in the Late-Early treatment (18.01 vs. 15.65 euros; $p = 0.04$).[13] This difference is quite substantial, as the required markup for getting up early on the last day of the experiment was approximately 42% ($= \frac{8.01}{5.65} * 100$) greater when participants stated their reservation wage early compared to in the late morning in their first online session. Clearly, participants valued sleep differently depending on the time of day, and this state dependency affected their assumed opportunity costs of getting up early on the last day of the experiment.

The same pattern was observed when comparing participants' reservation wages between the first and second online sessions. Participants in the Late-Early treatment significantly increased their reservation wages from the first to the second online session, by 1.82 euros on average ($p < 0.001$). Participants in the Early-Late treatment significantly decreased their reservation wages from the first to the second online session, by 0.44 euros on average ($p = 0.04$). Thus, participants in both treatments significantly changed their valuations of early-morning sleep, on average, from one online session to the next. This change occurred in the opposite direction between the two treatments, consistent with our assumption that participants value early-morning sleep more when it is early compared to late in the morning.[14]

Of course, when looking at the individual data, not every participant stated different reservation wages. In fact, in both treatments, the majority of participants stated the same reservation wage in the second online session as they did the week before. However, 42% of the participants in the Late-Early treatment and 31% in the Early-Late treatment changed their reservation wage in the second compared to the first

Table 1
Average reservation wages per session.

	Late-Early		Early-Late	
	Obs.	Mean	Obs.	Mean
		(SD)		(SD)
1st session	89	15.65	87	18.01
		(6.56)		(7.99)
2nd session	89	17.47	87	17.57
		(8.20)		(7.52)

Notes: The table shows means (standard deviations) of reservation wages in each online session of each treatment (in euros).

---

[9] For the instructions and screens used in the experiment, see Online Appendix A.2 and A.3.
[10] However, notice that dropouts will most likely be those students who have difficulty getting up early. Thus, if we had a high dropout rate, we may have ended up with a sample of pure early birds. Each dropout from our experiment would therefore bias the results against a conflict of selves.
[11] Since all relevant decisions were made during both online sessions, we were only concerned with dropouts until the third day of the experiment (i.e., the day of the second online session). Whether participants actually collected their payments on the fourth day of the experiment was of no peculiar interest. Also, we did not want to introduce an additional incentive for coming to the payment session beside the wage for performing the relevant task. It was therefore made clear to the participants that they could receive the lottery drawing prize even if they did not come to the payment session.

[12] Of those 176 students, eight did not come to the payment sessions to collect their earnings (five in treatment Early-Late and three in treatment Late-Early).
[13] If not reported otherwise, p-values are based on two-sided Mann-Whitney-U tests for unpaired samples and on two-sided Wilcoxon signed-rank tests for paired samples.
[14] Distributions of reservation wages in both online sessions of each treatment and corresponding distributions of individual differences in reservation wages in each treatment can be found in Online Appendix A.1.

online session. For the participants who changed their reservation wages, getting up early seemed to take more effort in general, as they already stated higher wages, on average, than the other participants in the first session of each treatment (Late-Early: 18.16 vs. 13.87, $p < 0.001$; Early-Late: 19.93 vs. 17.15, $p = 0.106$). The changes themselves then showed the pattern described above (see Table 2). In the Late-Early treatment, 86% of the changes involved an increase of reservation wages, on average by 5.84 euros. In the Early-Late treatment, 78% of changes involved a reduction of reservation wages, on average by 2.71 euros. Thus, if participants' valuations of early-morning sleep changed from one session to the next, the odds were 6.4:1 for an increase (or 0.16:1 for a decrease) in the Late-Early treatment and 3.5:1 for a decrease (or 0.29:1 for an increase) in the Early-Late treatment.

3.1.1. Discussion of results

Even though the participants evaluated the same good in both online sessions of a treatment (i.e., early-morning sleep on the day of the payment), their valuation changed systematically with the time of day. It seems, therefore, that we induced state-dependent preferences and triggered a conflict between selves about the value of early-morning sleep in the two online sessions of each treatment. Of course, not everyone has a problem with getting up early or experiences an inner conflict with him- or herself regarding early rising. Participants for whom the value of early-morning sleep changes with the time of day seem to struggle more to get up early and report higher reservation wages overall.

In terms of potential commitment choices, however, it is not essential that the reservation wage of every participant changed in actuality. The incentive to prepone the decisive decision by one week might have existed, even if a participant ended up stating the same reservation wage twice. It is sufficient that one feared a potential change of valuations in the second online session and wanted to protect oneself against this possibility. Not always when we put the alarm clock out of reach in the evening would it have been necessary in the morning. Often it is merely a precautionary measure for the case that the morning self is not willing to do what it should do from the evening's perspective.[15] This was similar in our experiment. The self a week later in the second online session could have a different valuation of early-morning sleep, and if this happened, the odds were that the change would be in a certain direction.

The only alternative explanation to an intrapersonal conflict about

the appropriate reservation wage for getting up early in our experiment that we are aware of is "projection bias" (Loewenstein et al. 2003). Projection bias captures the idea that people systematically mispredict their own preferences in a different state. Frequently, people appear to be too heavily influenced by their current preferences and they underestimate the magnitude of their preference changes in different states. According to this idea, participants in the first online session in the Late-Early treatment systematically underestimated their discomfort of getting up early, stated reservation wages that were too low, and consequently revised them upward in the second online session in the early morning.

In the Early-Late treatment, however, projection bias alone cannot explain the reservation wage differences in our experiment. In the first online session in this treatment, participants stated their wage for getting up early right after they got up early. In the second online session, late in the morning, they knew this wage and still deviated from it systematically. Therefore, this is more likely to be a disagreement than (an unconscious) misprediction about the reservation wage for getting up early.

3.2. Commitment choices

In the Late-Early treatment, it turned out that the participants were indeed willing to use measures against the early-morning planner. As shown in Table 3, more than half of the participants (63%) preponed the decisive decision one week.[16] This provides empirical evidence that external commitment devices are not only used against the early-morning doer, but also against the early-morning planner. In our experiment, this is a form of self-commitment that goes beyond the usual notion in terms of intertemporal preferences, because there is no intertemporal trade-off when stating the reservation wage in advance regarding early-morning sleep on the day of payment.

In the Early-Late treatment, too, the majority of the participants (57%) preponed the decisive decision. In fact, there was no significant difference between the two treatments in this respect ($X^2 = 0.34$; $p = 0.56$). If an external commitment device is available to the early-morning self, it seems to make use of it to a similar extent as the late-morning self. This finding remains the same if we look only at those participants who actually stated different reservation wages in the two online sessions (see Table 4). Among these participants, the proportion of those who chose the commitment option was almost identical, with 54% in the Late-Early treatment and 52% in the Early-Late treatment.

Finally, Table 5 shows average reservation wages and average revisions of reservation wages in each treatment, separated by participants' commitment choices. It appears that participants who did not commit themselves tended to revise their reservation wage from the first online session more substantially in both treatments (Late-Early: $p = 0.24$; Early-Late: $p = 0.19$). Moreover, participants who did not commit

Table 2
Occurrences, directions, and average amounts of revisions of reservation wages in the second online session.

	Total revisions		Reductions		Increases	
	Obs.	Mean	Obs.	Mean	Obs.	Mean
		(SD)		(SD)		(SD)
Late-Early	37	4.38	5	−5.00	32	5.84
		(6.13)		(3.08)		(5.11)
Early-Late	27	−1.41	21	−2.71	6	3.17
		(3.23)		(2.26)		(1.17)

Notes: The table shows the number of cases, the means (standard deviations), and directions of changes in reservation wages (in euros) in the second compared to the first online session of each treatment. A "Reduction" is a case in which a participant lowered his or her reservation wage in the second compared to the first online session. Analogously, an "Increase" occurred if a participant stated a higher wage in the second than in the first online session.

Table 3
Commitment choices among all participants.

Late-Early		Early-Late	
Obs.	#Commit (Prop.)	Obs.	#Commit (Prop.)
89	56 (0.63)	87	50 (0.57)

Notes: The table shows the number (proportions) of participants who committed themselves on their first reservation wage in each treatment.

---

[15] At the same time, the morning self does not necessarily have to destroy the alarm clock to enforce its own preferences. One way to feel rested in the morning, for example, would be to set the timer on the TV to remind (or force) the evening self to go to bed.

[16] As mentioned in footnote 3, the flexibility option was intentionally always presented above the commitment option on participants' computer screens. Thus, if one of the two options stuck out, it should have been the flexibility option.

Table 4

Commitment choices among participants who revised their reservation wage in the second online session.

Late-Early		Early-Late	
Obs.	#Commit (Prop.)	Obs.	#Commit (Prop.)
37	20 (0.54)	27	14 (0.52)

Notes: The table shows the number (proportions) of commitment choices among all participants who stated a different reservation wage in the second compared to the first online session.

Table 5

Average reservation wages and revisions between participants who committed themselves and those who did not.

	Late-Early			Early-Late		
	Obs.	Wage: 1st session	Revision: 2nd session	Obs.	Wage: 1st session	Revision: 2nd session
Committed	56	15.95 (7.33)	1.07 (2.54)	50	16.54 (7.19)	−0.18 (1.27)
Did not commit	33	15.15 (5.04)	3.09 (6.45)	37	20.00 (8.66)	−0.78 (2.47)

Notes: The table shows means (standard deviations) of reservation wages in the first online session and revisions of reservation wages in the second online session (in euros) of each treatment for all participants who committed themselves versus those who did not commit themselves. A positive revision indicates an average increase of reservation wages in the second compared to the first online session; a negative revision indicates a reduction of reservation wages, on average.

themselves stated higher reservation wages, on average, in the Early-Late but not in the Late-Early treatment (Early-Late: $p = 0.07$; Late-Early: $p = 0.85$).

3.2.1. Discussion of results

Of course, one may discuss whether Elster's argument about authenticity of different selves is sufficiently plausible from a normative perspective (see, e.g., Read, 2006; Sugden, 2018). From an empirical perspective, we must reject Elster's assertion that in any given intra-personal conflict only one self is capable of forward-looking behavior. Neither self in our experiment is superior to the other in terms of strategic abilities and both selves exploit the self-commitment opportunity to the same extent. Thus, based on Elster's criterion of forward-looking behavior, the early-morning self appears to be no less authentic than the late-morning self. Please note that projection bias cannot account for participants' propensity to choose the self-commitment option in our experiment either. Projection bias conceptualizes people's (alleged) inability to correctly predict future preferences in a different state, not to counter these preferences. Self-commitment is not part of the theory.

In the tradition of experimental economics, we described the two

options (i.e., flexibility vs. commitment) neutrally to the participants to avoid their possible associations with various labels. Participants were given the choice of whether their today's valuation of early-morning sleep (i.e., on the day of the first online session) or their upcoming valuation (i.e., on the day of the second online session) would be relevant for them with 80% probability. Preponing the relevant valuation to the first online session fixed their current valuation with high probability and greatly reduced the permissibility of a possible revision in the second online session. In this respect, the stochastic commitment device that we used in our experiment is meaningfully binding in our view. Augenblick et al. (2015) studied the demand for commitment in the context of present-biased preferences, and they also used a stochastic commitment device which was described neutrally to the participants.[17]

Regarding the observation that participants who did not commit themselves tended to revise their reservation wage in the second online session more substantially compared to participants who committed themselves, two explanations seem plausible. On one hand, it would be conceivable that more volatile preferences are correlated with a more pronounced liking for spontaneity. People whose preferences regarding consumption fluctuate more strongly might therefore have an equally strong preference for flexibility. Thus, in this case, participants would have been aware of their substantially fluctuating valuations of early-morning sleep, but their preference for flexibility was even stronger. On the other hand, these more substantial revisions of reservation wages could have been associated with a particular form of naiveté, akin to a phenomenon in self-control problems in the context of intertemporal preferences (O'Donoghue and Rabin 1999, 2001). Some people do not foresee their alternating preferences and the potential conflict that might be associated with them. Our data cannot distinguish between the two explanations.

3.3. Follow-up study

Although we relied on an established experimental feature for our stochastic commitment device, we acknowledge that we cannot be certain whether participants perceived this device as self-commitment. That they did presupposes that the participants during the first online session anticipated the possible change of their own valuation of early-morning sleep in the upcoming second online session. We therefore addressed people's anticipation of changing preferences in the context of early-morning sleep with the help of an additional survey.[18] Our follow-up study was conducted several years after our experiment with a sample from a different country. However, because the purpose of the survey was to gain insights on people's general ability to anticipate changing preferences in the context of early-morning sleep, the difference in time and place should not have mattered much.

The survey was conducted on CloudResearch in October 2022. CloudResearch is an online platform to recruit U.S. residents and it was found to provide reliable and valid results across a wide range of tasks and measures in the social sciences (Litman et al., 2017; Chandler et al., 2019). As is common in such studies, participants first had to give their consent to participate in the survey. The survey itself was preregistered

---

[17] Their longitudinal experiment took place over seven consecutive weeks and was much more complex than ours. In weeks 4 and 5 of their experiment, the participants made several allocation decisions for different tasks they had to fulfill. Only one of these allocations was the "decision-that-counts" for the participants. However, in week 4, the participants were offered a stochastic commitment device. That is, the participants could choose between the following two options: (i) "a 10 percent chance that the decision-that-counts will come from your Week 4 [today's] allocations (and 90 percent chance that it comes from Week 5)", and (ii) "a 90 percent chance that the decision-that-counts will come from your Week 4 [today's] allocations (and 10 percent chance that it comes from Week 5)" (see, Online Appendix F.4 of Augenblick et al., 2015).

[18] We thank the editor for suggesting such a survey.

at AsPredicted (https://www.aspredicted.org), where we specified the objective of the survey, the number of observations as well as the variables of interest.[19] The survey contained a control question to check the participants' attention and their understanding of the task. Out of a total of 105 participants in the survey, 87 answered the control question correctly. As specified in the preregistration, the following analysis is based on these 87 participants.

The survey was sent out at about 1 p.m. (PT) on a Tuesday and the preregistered number of participants was reached within about an hour. The focus of the survey was on the following two questions, which participants had to answer in consecutive order:

1. This is about your estimation of your value of sleep next Sunday morning:
2. What would you be willing to pay to avoid getting up at 6:00 a.m. next Sunday morning?
3. To avoid getting up at 6:00 a.m. next Sunday morning, I would be willing to pay the following amount (in whole dollars): _____
4. Please imagine you would answer the same question tomorrow morning at 6:00 a.m.
5. What would then be your estimation of your value of sleep next Sunday morning?
6. When asked tomorrow morning at 6:00 a.m. how much I would be willing to pay to avoid getting up at 6:00 a.m. next Sunday morning, I would state the following amount (in whole dollars): _____

Participants could report any positive amount (in whole dollars) for their willingness to pay (WTP) to avoid getting up early next Sunday morning. On average, participants stated a WTP of $27.53 on the first and $50.59 on the second question ($p = 0.003$). 21 of the 87 participants increased their WTP from the first to the second question and three participants decreased it. Two participants increased their WTP on Question 2 to a substantially greater extent than the other participants.[20] Even if these two participants are treated as outliers and excluded from the data, the WTP to avoid getting up next Sunday morning increased, on average, from $27.53 to $30.60 ($p = 0.009$) when participants were asked to imagine that they would answer the same question on Wednesday morning at 6 a.m. instead of Tuesday afternoon.

At the end of the survey, we also asked the participants how much they agreed with the statement that they often struggle a bit to get out of bed in the morning. Participants could indicate their agreement with this statement on a Likert scale ranging from 0 to 6, with 0 labeled "not at all", 3 labeled "in between", and 6 labeled "for sure." 26 participants disagreed with the statement that they struggle getting out of bed in the morning (i.e., agreement < 3). Of these participants, 85% reported the same WTP on both questions, and only 15% changed their WTP between the two questions. Overall, the WTP did not differ significantly between the two questions for this group (mean of WTP_1: 29.04, mean of WTP_2: 25.88, $p = 0.855$).[21] In contrast, 35 participants agreed with the statement that they struggle getting out of bed in the morning (i.e., agreement > 3). Of these participants, 46% changed their WTP between the two questions. Overall, the WTP increased significantly from Question 1

to Question 2 in this group (mean of WTP_1: 31.28, mean of WTP_2: 88.34, $p = 0.008$).[22]

3.3.1. Discussion of results

The survey shows that people anticipated their varying valuations of early-morning sleep at different times of the day. The results are not consistent with an unconscious misprediction about the discomfort of getting up early in the morning, as projection bias would imply. People were aware that they value the prospect of early-morning sleep on the upcoming Sunday higher in the morning than in the afternoon. Nonetheless, they were not willing to pay as much for this prospect in the afternoon as they were in the morning. People have different preferences in different states, and they know about it. Again, not everyone has state-dependent preferences regarding early-morning sleep. As it appeared to be the case in our experiment, participants in the survey who often struggle to get up early, were most likely to be subject to state-dependent preferences regarding early-morning sleep. If people have state-dependent preferences in this context, however, it is unclear which state is more authentic. Attributing weak will to the morning self is a normative judgment.

4. Conclusion

To the best of our knowledge, we provide the first experiment that explicitly tests for the empirical content of Schelling's interpretation of a reciprocal intrapersonal conflict. In the context of our study, conceptualizing intrapersonal conflict as a problem of time-inconsistency naturally reads the lazy early-morning self as a mere lack of willpower that spoils the informed preference of the authentic late-morning self, which rationally wants to seize the day. In this view, it seems obvious that the late-morning self's commitment against the early-morning self is a materialization of the only true rationale. In contrast to previous studies, we implemented a symmetric design with a second treatment in which the early-morning self analogously got the chance to bind the late-morning self to the early-morning self's preferences. In our experiment, the early-morning self was as likely as the late-morning self to use this device. Both selves did so for good reasons, because both selves equally sought to revise the choices of the first mover irrespective of whether the late-morning self could revise the early-morning self's choice or whether the early-morning self could revise the late-morning self's choice. A complementary survey further supports the idea that people are fully aware that their advance choices for sleeping at a future date systematically depend on whether one asks them late or early in the morning. These results challenge Elster's general statement that no examples for a bidirectional intrapersonal conflict exist.

It should be noted that our study tested for the reciprocity of intrapersonal conflict in only one of many possible contexts. We chose this setting primarily because it seemed like a prototypical example of an intrapersonal conflict that is usually interpreted in terms of a present bias. Furthermore, it allowed for a natural manipulation of different selves in a mundane and pervasive context. It is of course not inconceivable that the conflict between a sleepy and a rested self is peculiar in evoking a conflict between two strategically emancipated selves and that most, if not all, other possible contexts evoke more hierarchical conflicts of selves. It is, however, important to attempt replication in other domains, because our evidence suggests that intrapersonal conflicts that were previously interpreted as prototypically time-inconsistent could justifiably be reconstructed as state-dependent. While it may seem obvious for some with whom we should side in the former case, it should be much less clear in the latter.

For the time being, our exemplary finding merely demonstrates that

---

[19] To access the preregistration, please use the following link: https://aspredicted.org/blind.php?x=6WB_TQM.

[20] One participant increased the WTP on Question 2 by $745, while another increased it by $1000. The third highest increase on Question 2 was $100.

[21] One participant in this group decreased his or her WTP from Question 1 to Question 2 by $100 (from $300 to $200), which was a substantially greater decrease than any other. Without this participant, the average WTP was virtually identical between Questions 1 and 2 in this group (mean of WTP1: 18.20, mean of WTP2: 18.92).

[22] Even without the two participants who changed their WTP by $745 or $1,000, the WTP increased significantly from Question 1 to Question 2 in this group (mean of WTP1: 31.48, mean of WTP2: 39.15, p = 0.02).

the possibility of a conflict of emancipated selves might not just be the perspective of an "errant economist" (Schelling 1984a). Based on our findings, Schelling's interpretation seems no less plausible than the time-preference interpretation. According to Schelling, a higher or lower authenticity or legitimacy of one of these selves' preferences is not self-evident but lies in the eyes of the beholder. It might also depend on the zeitgeist. While Protestant work ethic and its attempt to overcome the "weaker self" were dominant in the age of industrialization, many modern Western societies consider "workaholism" a threat to what life is fundamentally about. Along these lines, Cowen (1991) argued that the rational choice approach should move in a similar direction as management science and planning theory, which nowadays deemphasize the need for command and embrace self-regulating orders. Successful self-management programs should be no less based on the effective coordination of conflictive desires than firms or economies. In his view, giving up the assumption of a hierarchy between selves would be an important step in this direction.

Our study is not meant to call into question the explanatory power of time-inconsistent preferences. We rather see it as a contribution to a richer debate on intrapersonal conflict. A more impartial view on the question of the "true" self, which the results of our experiment suggest, would call into question the self-evidence with which modern paternalism intervenes in the decision-making autonomy of the individual. If both selves are granted a rationale, paternalism faces similar problems as the beneficial dictator in his or her attempt to foster the welfare of society. Proponents of paternalism might then be forced to frame their interventions as political value judgments rather than unambiguous means of fighting irrationality. Libertarian paternalists argue that a default-free environment does not exist and that we should take the chance to set these defaults for the good of the "true" self of the weak-willed individual (Thaler et al. 2013). Indeed, default-free environments do not exist, and behavioral research has provided convincing evidence that defaults influence people's decisions. If a liberal state sets defaults, it should at least make this explicit and sensitize individuals for their influence on people's decisions. In this sense, libertarian paternalism that uses subtle nudges to align people with what it considers to be people's true interests might be even more problematic than hard paternalism, which fights certain preferences openly and can thus be more easily identified and criticized (Scoccia 2008). Ultimately, the results of our experiment urge caution regarding paternalists' confidence in exploiting "nudges" for the best of the individual. If economists are concerned about the Hayekian pretense of knowledge with regard to social choices, how can they be confident to know what the best interest of an inwardly torn individual is?

Data availability

Data will be made available on request.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.socec.2022.101971.

References

Allais, Maurice (1953). Le Comportement de l'Homme Rationnel Devant le Risque: Critique des Postulats et Axioms de l'École Américaine. *Econometrica*, 503–546.

Arad, Ayala, & Rubinstein, Ariel (2018). The people's perspective on libertarian-paternalistic policies. *The Journal of Law and Economics, 61*(2), 311–333.

Ariely, Dan;, & Wertenbroch, Klaus (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science, 13*(3), 219–224.

Ashraf, Nava, Karlan, Dan, & Yin, Wesley (2006). Tying odysseus to the mast: Evidence from a commitment savings product in the Philippines. *The Quarterly Journal of Economics, 121*(2), 635–672.

Augenblick, Ned, Niederle, Muriel, & Sprenger, Charles (2015). Working over time: Dynamic inconsistency in real effort tasks. *The Quarterly Journal of Economics, 130*(3), 1067–1115.

Becker, Gordon M., DeGroot, Morris H., & Marschak, Jacob (1964). Measuring utility by a single-response sequential method. *Behavioral Science, 9*(3), 226–232.

Chandler, Jesse, Rosenzweig, Cheskie, Moss, Aaron J., Robinson, Jonathan, & Litman, Leib (2019). Online panels in social science research: Expanding sampling methods beyond mechanical turk. *Behavior Research Methods, 51*(5), 2022–2038.

Cowen, Tyler (1991). Self-constraint versus self-liberation. *Ethics, 101*(2), 360–373.

DellaVigna, Stefano, & Malmendier, Ulrike (2006). Paying not to go to the gym. *The American Economic Review, 96*(3), 694–719.

Elster, Jon (1985). Choice and Consequence: Perspectives of an Errant Economist: Thomas C. Schelling (Harvard University Press, Cambridge, MA, 1984). *Journal of Economic Behavior & Organization* 6(1), 89-92.

Greiner, Ben (2004). An online recruitment system for economic experiments. Forschung und Wissenschaftliches Rechnen 2003. *GWDG Bericht, 63*, 79–93.

Karni, Edi, Schmeidler, David, & Vind, Karl (1983). On state dependent preferences and subjective probabilities. *Econometrica, 51*(4), 1021–1031.

Kivetz, Ran, & Simonson, Itamar (2002). Self-control for the righteous: Toward a theory of precommitment to indulgence. *Journal of Consumer Research, 29*(2), 199–217.

Laibson, David (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics, 112*(2), 443–477.

Litman, Leib, Robinson, Jonathan, & Abberbock, Tzvi. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods, 49*(2), 433–442.

Loewenstein, George, O'Donoghue, Ted, & Rabin, Matthew (2003). Projection bias in predicting future utility. *The Quarterly Journal of Economics, 118*(4), 1209–1248.

Madrian, Brigitte, & Shea, Dennis F. (2001). The power of suggestion: Inertia in 401 (k) participation and savings behavior. *The Quarterly Journal of Economics, 116*(4), 1149–1187.

O'Donoghue, Ted, & Rabin, Matthew (1999). Doing it now or later. *The American Economic Review, 89*(1), 103–124.

O'Donoghue, Ted, & Rabin, Matthew (2001). Choice and procrastination. *The Quarterly Journal of Economics, 116*(1), 121–160.

Read, Daniel (2006). Which side are you on? The ethics of self-command. *Journal of Economic Psychology, 27*, 681–693.

Schelling, Thomas (1984a). *Choice and consequence: Perspectives of an errant economist.* Cambridge, MA: Harvard University Press.

Schelling, Thomas (1984b). Self-command in practice, in policy, and in a theory of rational choice. *The American Economic Review, 74*(2), 1–14.

Scoccia, Danny (2008). Defense of hard paternalism. *Law and Philosophy, 27*(4), 351–381.

Strotz, Robert H. (1955). Myopia and inconsistency in dynamic utility maximization. *The Review of Economic Studies, 23*(3), 165–180.

Sugden, Robert (2018). *The community of advantage: A behavioural economist's defence of the market.* Oxford, UK: Oxford University Press.

Thaler, Richard H., & Sunstein, Cass R. (2003). Libertarian paternalism. *The American Economic Review, 93*(2), 175–179.

Thaler, Richard H., Sunstein, Cass R., & Balz, John P. (2013). Choice architecture. In Eldar Shafir (Ed.), *The behavioral foundations of public policy* (pp. 428–439). Princeton: Princeton University Press.

Uhl, Matthias (2011). Challenging the Intrapersonal Empathy Gap: An Experiment with Self-Commitment Power. *Jena Economic Research Papers 2011*, 019.