

IZA DP No. 7610

Incentives and Information as Driving Forces of Default Effects

Steffen Altmann
Armin Falk
Andreas Grunewald

September 2013

Incentives and Information as Driving Forces of Default Effects

Steffen Altmann

IZA

Armin Falk

*University of Bonn
and IZA*

Andreas Grunewald

University of Bonn

Discussion Paper No. 7610
September 2013

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Incentives and Information as Driving Forces of Default Effects^{*}

The behavioral relevance of non-binding default options is well established. While most research has focused on decision makers' responses to a given default, we argue that this individual decision making perspective is incomplete. Instead, a comprehensive understanding of the foundation of default effects requires taking account of the strategic interaction between default setters and decision makers. We provide a theoretical framework to analyze which default options arise in such interactions, and which defaults are more likely to affect behavior. The key drivers are the relative level of information of default setters and decision makers, and their alignment of interests. We show that default effects are more pronounced if interests of the default setter and decision makers are more closely aligned. Moreover, decision makers are more likely to follow default options the less they are privately informed about the relevant decision environment. In the second part of the paper we experimentally test the main predictions of the model. We report evidence that both the alignment of interests as well as the relative level of information are key determinants of default effects. An important policy relevant conclusion is that potential distortions arising from default options are unlikely if decision makers are either well-informed or reflect on the interests of default setters.

JEL Classification: D03, D18, D83, C92

Keywords: default options, libertarian paternalism, behavioral economics, incentives, laboratory experiment

Corresponding author:

Steffen Altmann
IZA
P.O. Box 7240
53072 Bonn
Germany
E-mail: altmann@iza.org

^{*} Financial Support from Volkswagen Foundation and the Deutsche Forschungsgesellschaft (SFB/TR 15) is gratefully acknowledged. We thank Paul Heidhues, Eric Johnson, Daniel Martin, Andrei Shleifer, and Elke Weber, as well as seminar and conference participants at Columbia University, DIW Berlin, the 2012 EEA / ESEM meetings, ESA Tucson, Heidelberg University, the M-BEES Symposium, NYU, Pittsburgh University, and the University of Oxford for valuable comments. Nicolas Kaufung provided excellent research assistance.

1 Introduction

A substantial body of empirical research has shown that non-binding default options strongly affect consumption and savings decisions. Default effects have been documented, for instance, in employees’ enrollment and contribution rates to 401(k) plans (Madrian and Shea 2001, Choi et al. 2004), decisions on workplace training courses (Borghans and Golsteyn 2013), choices of insurance contracts (Johnson et al. 1993), car purchases (Levav et al. 2010), or consent to postmortem organ donations (Johnson and Goldstein 2003).

In this paper, we argue that such default effects commonly arise in situations where default setters—firms, organizations, or governmental agencies—deliberately specify the default option. Defaults can thus be seen as the outcome of an inherently strategic interaction between the default setter and a population of decision makers (“agents”). As a consequence, the characteristics of the default setter and the agents might influence both the specification of default options, and the strength of default effects. We analyze how two key dimensions—the alignment of interests between default setters and decision makers, and their relative knowledge of the decision environment—affect these outcomes.

To formalize our argument, we apply the well-established theoretical framework of strategic communication to the question how default effects arise. In our model, both the default setter and a population of agents have incomplete private information about which of the available choices is optimal for them. In particular, both parties may be uncertain about the preferences of the agents, the exact consequences of the decision, or an underlying state of the world. We further assume that all agents are rational, but might differ in their knowledge of the decision environment. The default setter specifies the default option that agents face when making their decisions. She might thus use the default as a means to communicate her private information on the optimal choice for the agents. Agents, in turn, may follow the default or actively choose a different alternative.

Our model shows how the alignment of interests as well as the relative level of information affect default specifications and the strength of default effects in equilibrium. First, the behavior of default setters and decision makers depends on how closely their interests are aligned. We show that, in any Pareto-efficient equilibrium, defaults are more informative if the interests between the default setter and decision makers are more closely aligned, i.e., if

the default setter is more benevolent towards the population of agents. Defaults specified by a fully benevolent default setter, for instance, always truthfully reveal her beliefs about the best decision for the agents. In contrast, defaults specified by a fully selfish default setter convey no information (“babbling”). At the same time, we show that the reaction of rational agents to a default option depends on its informative value, and is therefore itself related to the default setter’s level of benevolence. Second, our model predicts that default effects differ systematically across subgroups of the population: when defaults are at odds with the decision makers’ own information, agents are more prone to follow defaults, the lower is their private information quality.

In the second part of the paper, we test the key predictions of our model in a laboratory experiment. As our workhorse, we use a simple binary-choice paradigm in which agents have to decide whether a set of nine cards, which can be either red or black, contains more red cards or more black cards. The ex-ante likelihood that a given set of cards contains more cards of a given color is 0.5. Before making choices, the default setter and the agents receive independent and informative signals on the agents’ payoffs from the two available choice options. The default setter selects one of the options as the default, and the agents can accept the default option or opt out. To identify the causal effects of default setters’ strategic incentives, we exogenously vary whether preferences of the default setters are (i) fully aligned, (ii) partially aligned, or (iii) misaligned with those of the agents (FUL, PAR, and MIS treatment, respectively). Within each treatment, we additionally vary the relative level of information of default setters and agents. In particular, there are always some agents who are better informed than the default setter, and some agents whose information quality is below that of default setters.

Our empirical results support the notion that strategic incentives as well as the quality of information are key determinants of behavior for default setters and decision makers. First, defaults truthfully reveal default setters’ information in 98% of cases in FUL, but only in 75% of cases in PAR and in 56% in MIS. Default setters with interests aligned to those of the agents, thus, select informative defaults, while defaults specified by default setters with misaligned preferences barely convey any information. Second, agents are substantially more likely to accept defaults that are chosen by more benevolent default setters, with 90%,

74%, and 58% of agents accepting defaults in FUL, PAR, and MIS, respectively. Third, agents' reaction to defaults strongly depends on the quality of their personal information. In FUL and PAR, agents with low and intermediate levels of information strongly rely on the information transmitted through default options. At the same time, decisions of agents with superior information are barely affected by defaults in either of the treatments: in particular, well-informed agents almost always opt out when the default is in conflict to their own information. Finally, our results suggest that defaults set by fully or partially benevolent default setters can in fact enhance the aggregate quality of individual decisions, lending support to a common theme in the literature on "libertarian paternalism" (Thaler and Sunstein 2003, Camerer et al. 2003, Sunstein 2012). In particular agents with lower levels of information benefit substantially from accepting default options in the FUL and PAR treatment. At the same time, default options do not distort choices of well-informed agents in these treatments. However, we also find that the effects on agents' welfare are more mixed if the default setter is fully selfish. While well-informed agents are still not distorted in their decisions, those with intermediate levels of information tend to follow defaults somewhat too frequently, with detrimental consequences for their payoffs.

From an individual choice perspective, default options should be irrelevant for the behavior of rational agents, as long as the costs of opting out of the default are trivial. This perspective is based upon the assumption that defaults do not convey any information. In contrast, we suggest that defaults often arise from a strategic interaction of a default setter and a population of agents. Rational agents should therefore follow defaults whenever the informational quality of the default is sufficiently high. The idea that defaults are behaviorally relevant since they are perceived as implicit recommendation of the default setter has informally been discussed in the psychology and marketing literature on default effects (Johnson and Goldstein 2003, McKenzie et al. 2006). McKenzie et al. (2006) as well as Tannenbaum (2011) provide evidence from vignette surveys and (non-incentivized) questionnaires, supporting the notion that consumers stick to defaults in order to follow the default setter's recommendation. On the other hand, Brown and Krishna (2004) have argued that "marketplace metacognition" might instead make consumers skeptical about defaults set by profit maximizing firms; as a consequence, they should more heavily rely on active choices.

We provide a formal framework that addresses both points of view. It shows that the informational content of default options, and the extent to which rational consumers account for this information both depend on the degree to which the interests of default setters and decision makers are aligned.

Several other important motives for why defaults can influence behavior have been discussed in the literature (see Sunstein 2013 for a comprehensive review). These include status-quo effects and loss aversion (Samuelson and Zeckhauser 1988, Kahneman et al. 1991), pecuniary and non-pecuniary costs of opting out (Schwartz and Scott 2003, Thaler and Sunstein 2003), quasi-hyperbolic discounting and procrastination of active decisions (Madrian and Shea 2001, Carroll et al. 2009), or inattention and other perceptual limitations of decision makers (Caplin and Martin 2013a, Caplin and Martin 2013b). These studies have made important contributions, by demonstrating how various psychological and economic motives can explain individuals' adherence to a given default. All of them, however, largely abstract away from modeling the default setter and the question how default options are specified in a given decision environment. Complementing this approach, we show that strategic motives and informational asymmetries are important for understanding which default option decision makers face in a given context, and how they react to it. The factors that we study should thus also influence the overall strength of default effects, independently of the psychological motives at play. A loss-averse consumer, for example, should also take his experience and knowledge of the decision environment into account when deciding on whether or not to stick to a default. Even if consumers generally tend to postpone active decisions, they should consider the intentions of the default setter in the corresponding choice setting. By focusing on an individual-choice perspective, such aspects have generally been neglected in the literature on defaults effects. While our baseline model also abstracts from analyzing such interactions with different psychological motives, our theoretical framework is flexible enough to be subsequently enriched by these factors.

Our results also inform the discussion on libertarian paternalistic policy interventions. Since defaults do not restrict choices, a number of scholars have argued that they could be a powerful instrument for helping ill-informed individuals make better decisions, without distorting choices of others (Thaler and Sunstein 2003, Camerer et al. 2003, Sunstein 2012).

One of the most important concerns against libertarian paternalism, however, is precisely the fear that consumers do not react optimally to the default setter’s private information and personal motives, and that they might thus also follow defaults that are against their best interest (Glaeser 2005). Participants in our experiment generally exhibit a pretty accurate understanding of the differences in the informational quality of defaults, and they tailor their reactions accordingly. In particular, decision makers with low information quality benefit from defaults if they are specified by a default setter with at least partially aligned interests, while the decisions of well-informed participants are not distorted. However, our empirical results also point out some important deviations from this general pattern. In particular, agents with intermediate-level information seem to follow defaults from default setters with misaligned preferences too frequently. This indicates that decision makers might sometimes be naive about the full consequences of conflicting interests. One means to help decision makers attain a deep-enough understanding of the default setter’s information and strategic incentives might be consumer protection policies, such as mandatory disclosure laws or the regulation of contract cancelation rights. The potential benefits and limitations of such interventions to protect naive consumer have been extensively studied in the literature of financial advice (for an overview see Inderst and Ottaviani 2012a). Our results suggest that broadening the scope of this discussion to the regulation of markets where defaults are prevalent can be important to ensure that consumers reap the benefits of informative defaults, without bearing the detrimental consequences of following ill-specified ones.

The remainder of the paper is organized as follows. In the next section we present our model and derive testable implications for the empirical analysis. Section 3 outlines the design and procedures of the experiment. In Section 4 we present the empirical results, and Section 5 concludes.

2 The Model

In what follows, we apply a theoretical framework of strategic communication to the question how default effects arise. Our approach builds upon the classic work of Crawford and Sobel (1982) and subsequent contributions that have studied the influence of strategic communication in various other applications. We focus on two dimensions that are inherent in any

interaction between a default setter and decision makers (“agents”). First, we analyze how default-setting and decision making are affected by the strategic incentives of the interacting parties. That is, we aim at providing comparative statics on how alignment of incentives between default setters and agents influence default specifications, and agents’ tendency to stick to the default. Second, we analyze how asymmetric information between default setters and decision makers influences decision making. In particular, we assume heterogeneous information quality across the population according to which decision alternative maximizes the decision makers’ welfare. This heterogeneity enables us to study how the relative level of information between a decision maker and the default setter shapes default effects. All proofs are found in the appendix.

2.1 Model setup

In the model one default-setter interacts with a population of decision makers. These are assumed to be rational and choose an action z in order to maximize their utility $U_A(z, \theta)$, which is strictly concave in z for all θ . The state of the world, θ , can take value θ_h or θ_l , which determine the preferences of the agents over the choice variable. More precisely, we assume that the partial derivative with respect to z of $U_A(z, \theta_l)$ is smaller than the corresponding derivative of $U_A(z, \theta_h)$ for all z , i.e, higher z s are optimal if the agent puts more probability weight on θ_h . The action z is a discrete choice variable $z \in \{z_1, \dots, z_m\}$. Both states of the world are equally likely and ex ante unobservable for the players. However, all players receive incomplete information about the true state of the world. Decision makers dispose of different levels of information quality. In particular, we assume that the population of agents is distributed according to $f(x)$, with full support over $[\frac{1}{2}, 1]$, where x is the signal strength of agent x . Agents’ signals are denoted by $\sigma \in \{\sigma_l, \sigma_h\}$ and have conditional distributions $p(\sigma_l|\theta_l) = p(\sigma_h|\theta_h) = x$.

The default setter also has private information about the optimal decision for the agents. Her signal is drawn independently from the ones of the agents and denoted by $\rho \in \{\rho_l, \rho_h\}$ with quality $q = p(\rho_l|\theta_l) = p(\rho_h|\theta_h) \in (\frac{1}{2}, 1)$. Consequently, the default setter is always better informed about the state of the world than some agents, and worse than others. While the default setter cannot influence the agents’ decision directly, she chooses a default option $d \in \{d_1, \dots, d_n\}$ prior to the agents’ decision. For the default setter, the utility from choice

z of a particular agent is a weighted sum of the utility of the agent and a term, $b(z)$, that captures a potential conflict of interests:¹

$$U_P(\theta, z) = \mu U_A(\theta, z) + b(z).$$

The parameter μ describes the alignment of interests between the default setter and the agents. For small μ , the default setter cares only little about the agents' well being and focuses more strongly on her private interests. For large μ , however, preferences of the default setter and the agents are more strongly aligned. A firm that anticipates a repeated long run interaction with the consumer may for example weigh consumer satisfaction more strongly than a firm that interacts only once with each consumer. Similarly, the alignment of interests between a government and agents may be stronger, than it may be in the case of a customer and a profit maximizing company. Note that we assume for simplicity that the strategic incentives of the default setter μ are common knowledge. To put structure on the preferences of the default setter, we assume an upward bias in the sense that $b(z)$ is strictly increasing in z . Hence, the utility maximizing choice z for the default setter is always weakly higher than the one for the agents.

The game is divided into four stages. On stage 1 nature draws the state of the world, θ , and all private signals. Moreover, all agents and the default setter observe their corresponding signal. On stage 2 the default setter decides on the default option d . On stage 3, the default is transmitted to the agents, and they decide individually on which choice z they implement. Dependent on the choice and the true state of the world, payoffs for agents and the default setter are realized on stage 4. In this setup, a strategy of the default setter is a mapping

$$s_P : \rho \rightarrow (p(d_1), \dots, p(d_n)) \in [0, 1]^n$$

which specifies the probability of every default to be chosen for all possible signals ρ . A strategy of the default setter, thus, determines the correlation between any default and her signal. For any combination of a private signal σ and default d , agents choose z . A pure

¹We assume for simplicity, that the default setter puts equal weight on the utility received from any decision of a given agent. It is straightforward to extend the model to incorporate a weighting function with heterogeneous weights for different subgroups of the population.

strategy² of an agent with signal strength x is thus a mapping

$$s_A^x : (\sigma, d) \rightarrow z \in \{z_1, \dots, z_m\}.$$

Define s to be a strategy profile, consisting of a strategy of the default setter and the strategy of every agent. As a solution concept we apply Bayesian equilibrium. Hence, all players maximize their utility at any information set occurring with positive probability given their type and the strategies of all opponents.

2.2 Equilibrium analysis

In the following analysis we concentrate on equilibria in which all defaults are played with positive probability. This is possible without loss of generality, since every equilibrium can be replicated by an output equivalent equilibrium without out-of-equilibrium defaults. If there is an off equilibrium path default, it is always possible to construct a default setter's strategy that mixes over this default and an equilibrium default such that both convey the same information. Consequently, the best response of the agent is to treat them identically. Since the default setter is, thus, indifferent between both defaults, the new strategy profile constitutes an output equivalent equilibrium. More precisely, the ex ante probability that a particular utility outcome for the agents and the default setter is realized is identical.

Lemma 1. *For any Bayesian equilibrium s there exists an output equivalent Bayesian equilibrium in which all defaults are played with positive probability.*

Note that any Bayesian equilibrium in which all defaults are played also constitutes a perfect Bayesian equilibrium. Hence, agents choose action z to maximize the expected utility $E(U_A|d, \rho, x)$ at any information set. Once an agent received his signal and the default setter has decided upon a default, agent x solves

$$\max_z E_\theta(U_A|d, \sigma, x) = p(\theta_l|d, \sigma, x)U_A(z, \theta_l) + p(\theta_h|d, \sigma, x)U_A(z, \theta_h).$$

²In general, an agent may be indifferent between different actions z and mix over the actions in equilibrium. Nevertheless, the set of agents that are indifferent between at least two messages has mass zero. Hence, extending the analysis to mixed strategies does not provide any additional insights.

Before the agents decide on z , the default setter can influence choice by specifying a default option to ensure a more favorable outcome to herself. Consider a perfect Bayesian equilibrium with strategy profile s . The default setter's expected payoff from playing default k in information set i associated with signal ρ_i , is given by

$$\int_{\frac{1}{2}}^1 p(\theta_l, \sigma_l | \rho_i, x) U_P(s_A^x(\sigma_l, d_k), \theta_l) + p(\theta_h, \sigma_l | \rho_i, x) U_P(s_A^x(\sigma_l, d_k), \theta_h) \\ + p(\theta_l, \sigma_h | \rho_i, x) U_P(s_A^x(\sigma_h, d_k), \theta_l) + p(\theta_h, \sigma_h | \rho_i, x) U_P(s_A^x(\sigma_h, d_k), \theta_h) dF(x).$$

Whenever the default setter receives a high signal, her preferences are aligned with those of the agents. In this case, a higher z maximizes both, her selfish interests $b(z)$ and the utility of the agents. Hence, the default setter chooses the default which induces the highest z . As a consequence, all defaults that are played with positive probability after a high signal must induce the same z and are qualitatively identical. The remaining defaults reveal that the default setter received a low signal. According to this distinction we split the set of defaults $\{d_0, \dots, d_n\}$ into two parts. Denote by $\{d_0, \dots, d_{k-1}\} = D_l$ the set of defaults that are exclusively played after a low signal and by $\{d_k, \dots, d_n\} = D_h$ the set of defaults that are played with positive probability after a high signal.

Lemma 2. *In every equilibrium the default setter will send at most two qualitatively different messages.*

Using the lemma we are able to describe the information transmission by the default setter by a single parameter $c = p(d_0 | \rho_l) + \dots + p(d_{k-1} | \rho_l)$. The higher c , the more often the default setter truthfully reveals her signal after a low signal, leading to more information transmission. Note that the meaning of every default arises endogenously in equilibrium and is not exogenously given. Clearly, all permutations of messages also support an equilibrium. We name all defaults played with positive probability after a high signal a high default (d_h) and those that are exclusively played after a low signal a low default (d_l).

2.3 Alignment of interests

The decision of rational agents to follow defaults specified by a default setter is likely to depend on the default setter's level of benevolence. A default setter with opposing interests will use his information to extract a higher rent for herself. In contrast, a more benevolent

default setter will specify a default that is more informative for the agents and should thus more likely be followed. In the extreme case of full alignment of preferences, the default setter has incentives to truthfully transmit all of his information.

Proposition 1. *There exists a $\bar{\mu}$ such that for all $\mu > \bar{\mu}$ the Pareto-dominant equilibrium exhibits complete information transmission. Moreover, there exists a $\underline{\mu}$ such that for all $\mu < \underline{\mu}$ defaults convey no information in equilibrium.*

If preferences are not fully aligned, the default setter may choose to conceal some of her information. For this case there may emerge multiple equilibria exhibiting different levels of information transmission. To rank those, we employ the Pareto-efficiency criterion. Agents always prefer an equilibrium corresponding to a larger c to one with less information transmission, since more information yields a better decision and thus a higher expected payoff. The same is true for the default setter, if she receives a high signal, because preferences are aligned in this case. If the default setter receives a low signal she is indifferent between all mixed equilibria. To see this, suppose the default setter chooses a message from the set D_l in a mixed equilibrium. Since this action reveals her signal, the inherent information and the expected payoff associated with a low default are equal in all mixed equilibria. Furthermore, the default setter is indifferent between a low and a high default. Hence, the expected payoff after a high default must also be identical across mixed equilibria. Overall, the default setter prefers the mixed equilibrium with the highest information transmission rate. If the expected payoff following a high default is larger than the one corresponding to a low default for full information transmission, the default setter clearly prefers the full information equilibrium to all mixed equilibria. We conclude that the equilibria featuring the highest information transmission rate are Pareto-efficient.

Lemma 3. *Consider two equilibria with corresponding $c_1 < c_2$. Then the equilibrium corresponding to c_2 Pareto-dominates the one belonging to c_1 .*

In the following, we focus on the question how a Pareto-efficient equilibrium depends on the model's main parameters. Define $c^{pd}(\mu)$ to be the information transmission rate in a Pareto-dominant equilibrium. As argued above, μ represents the degree to which default setter's and agents' preferences are aligned. For increasing μ , any Pareto-efficient equilibrium becomes more favorable for the agents.

Proposition 2. $c^{pd}(\mu)$ is weakly increasing in μ . Agents internalize this effect and exhibit more pronounced reactions to defaults, i.e., $s_A^x(\sigma_i, d_h)$ is weakly increasing and $s_A^x(\sigma_i, d_l)$ is weakly decreasing in μ .

In any equilibrium with full information transmission this is trivially the case. Consider the case of a Pareto-dominant equilibrium with a mixed strategy of the default setter. Then she is indifferent between specifying a high or a low default in response to a low signal:

$$\int_{\frac{1}{2}}^1 E_{\sigma, \theta} [U_P(\theta, s_A^x(\sigma, d_h)) - U_P(\theta, s_A^x(\sigma, d_l)) | x, \rho_l] dF(x) = 0.$$

With increasing μ the default setter's preferences become more strongly aligned to those of the agents. As a consequence, her utility of a truthful report increases relative to the utility of concealing information, hence, the difference above is decreasing in μ for all x . This effect is offset along the equilibrium path by a higher transmission rate c . The second part of the difference does not change with c . Playing a default from the set D_l always conveys the same information and thus induces the same action z . In contrast, the first part must be increasing in c . In the proof of Proposition 2 we show that if this was not the case there would always exist an equilibrium with a higher transmission rate which is a contradiction to the assumption of playing a Pareto-efficient equilibrium. In general, the proposition implies that a higher level of benevolence leads to a higher informational content of the defaults, and thus ceteris paribus to a higher incentive for agents to adapt their choice towards the default. In particular, in the case of a conflicting private signal and default, agents weigh the default more strongly. Hence, default effects are stronger in a given population of agents if interests of the population and the default setter are more aligned.

2.4 The quality of information

Differences in agents' knowledge about the decision environment are a potential explanation why the stickiness of defaults might differ across subgroups of the population. Intuitively, agents that are less familiar with the choice environment rely more heavily on the information entailed in the default option. In our model, such differences in agents' information are captured by the agents' signal strength, $x = p(\sigma_l | \theta_l) = p(\sigma_h | \theta_h)$.

Proposition 3. *In equilibrium, the strategies of any two agents with $x_1 < x_2$ exhibit the following properties:*

$$s_A^{x_1}(\sigma_l, d_i) \geq s_A^{x_2}(\sigma_l, d_i) \quad s_A^{x_1}(\sigma_h, d_i) \leq s_A^{x_2}(\sigma_h, d_i) \quad \text{for } i \in \{l, h\}$$

Agents are less susceptible to adapt their decision towards a given default if the quality of their personal information increases. The relationship between agents’ knowledge and the degree of default adherence is most clearly seen if the default and the agent’s private signal are in conflict. For instance, the strategy $s_A^x(\sigma_l, d_h)$, which describes agents’ behavior after a high default and a low personal signal, is decreasing in the information quality x . Consequently, agents with more informative low signals shift “further” away from the high default and rely more strongly on their own information about the optimal decision.

3 Design of the experiment

3.1 The game

In the experiment, we aim at testing the key comparative statics of the model empirically. Firstly, we endogenously vary the alignment of interests between default setters and decision makers. We consider three different situations in which the preferences of the default setter and the agents are (i) fully aligned, (ii) partially aligned, or (iii) misaligned (in what follows, we refer to the three conditions as the “FUL”, “PAR”, and “MIS” treatment, respectively). Secondly, for a given level of benevolence, our empirical approach ensures controlled variation in the relative level of information between the default setter and the agents. Some agents are better informed than the default setter about which decision maximizes their payoff, while others are less informed.

As a workhorse for implementing these treatment conditions, we use a simple paradigm in which one default setter interacts with one decision maker. In each period, the agent has to decide whether a set of nine cards, which can either be red or black, contains more red cards or more black cards. Each card in each period is drawn independently with probabilities $p(\text{Red}) = p(\text{Black}) = 0.5$, i.e., the ex-ante likelihood that a given set of cards contains more cards of a given color is 0.5. Before making choices, the default setter and the agent receive independent signals about the composition of the current set of cards.

The default setter receives the signal via a message on her screen. The message can either indicate that the current set of cards contains “more black cards”, or “more red cards”. The signal is private information of the default setter. Her signal strength, however, is common knowledge and held constant at $q = 0.8$. Whether a default setter receives a correct or wrong signal in a given period is determined randomly and independently between default setters, periods, and sessions.

Agents receive information about the number of red and black cards in the current set, by a subset of cards that is privately revealed to them. In each period, a coin flip determines whether the first *two* or the first *five* cards in the set are uncovered for a given agent.³ The signal-generating mechanism for agents ensures that we obtain controlled variation of the information quality for different types of agents. Since each of the revealed cards for the agent is black or red with probability 0.5, we obtain five different levels of signal strengths. The resulting distribution of agents’ signal qualities is reported in Table 1. Since the default setter’s signal strength is always 0.8, the agent is informed worse than the default setter in about 56% of cases (Columns 1 and 2 in Table 1); in about 25% of cases he has about the same signal strength (Column 3 in Table 1), and in 19% of cases (Columns 4 and 5), the agent is better informed than the default setter. As a within-subject dimension, we can test whether agent types 1–5 systematically differ in their behavior, holding default setters’ information quality and benevolence constant (Proposition 2).

Agent Type	A50	A69	A77	A94	A100
Signal Quality	0.5	0.69	0.77	0.94	1.00
Occurrence Probability	0.25	0.313	0.25	0.156	0.031
Example	1 black 1 red	3 black 2 red	2 black 0 red	4 black 1 red	5 black 0 red

Table 1: *Distribution of agents’ signals.*

Participants in the experiment do not see Table 1, but the signal-generating mechanism for default setters and agents is common knowledge. In particular, the default setter knows

³Figure A.1 in the supplementary material depicts an example of an agent’s information screen.

the procedure how the agent is informed, and consequently the distribution of signal qualities for the agent. However, she is not informed about the number or colors of the cards that are revealed to the agents in a specific period. Hence, the default setter essentially plays against a population of agents as depicted in Table 1, although default setters and agents are matched one-to-one in a given period.

After having received her signal, the default setter selects the default option for the current period. She can either specify “more red cards” or “more black cards” as the default. In a next step, the agent is informed about the default option. He can then accept the default or opt out and take an active decision. To confirm or change the default option in the experiment, the agent has to press the respective button displayed on his screen. In case the agent accepts the default option, the selected default is implemented as final decision for both the default setter and the agent. If the agent presses “opt out”, a new screen pops up on which the agent can (actively) choose between “more red cards” and “more black cards” as her ultimate decision for this period.⁴ After the agent has taken his decision, he is asked to state his perceived certainty that his choice was correct on an 8-points Likert scale. In a final stage of the game, default setter and agent are provided with a feedback screen, on which the entire set of cards for the current period is revealed. Furthermore, players are informed about the agent’s final decision and the resulting payoffs for both players.

3.2 Treatments and payoffs

Agents earn points if their decision in a given period is correct (i.e., the chosen color matches the color that occurs more frequently in their current set of cards). In each of the three treatments, payoffs for agents were calculated as follows:

$$\pi_A = \begin{cases} 50 \text{ points} & \text{if decision correct} \\ 0 & \text{if decision wrong} \end{cases}$$

Payoff functions for default setters differed across treatment conditions. In treatment FUL, default setters’ payoffs are perfectly aligned with those of the agents. Thus, a default

⁴We neither impose a time limit for default setters nor for agents making their final decision. This procedure ensures that the cost of opting out of the default is minimal, while at the same time avoiding mistakes due to time pressure or accidental clicks.

setter receives 50 points if the agent’s decision in a given period is correct, and 0 otherwise:

$$\pi_P^{FUL} = \begin{cases} 50 \text{ points} & \text{if A's decision correct} \\ 0 & \text{if A's decision wrong} \end{cases}$$

In contrast, default setters in the treatment with misaligned preferences (MIS), receive 50 points if and only if the agent’s decision in a given period is “more red cards”. They receive 0 points if the agent’s decision is “more black cards”:

$$\pi_P^{MIS} = \begin{cases} 50 \text{ points} & \text{if A's decision "red"} \\ 0 & \text{if A's decision "black"} \end{cases}$$

This payoff function resembles a default setter intending to direct the decision of the agent towards one particular alternative, without taking the welfare consequences for the agent into account, i.e., a default setter with selfish interests. Another theoretically equivalent⁵ way to implement misaligned preferences is to assume that default setters receive points if and only if agents do not receive any points. We opted for the first version to keep the setting as understandable for the subjects as possible. To induce partial benevolence, each matching group in our third treatment, PAR, consisted half of benevolent and half of selfish default setters with payoff functions π_P^{FUL} and π_P^{MIS} , respectively. Agents are not informed about which type of default setter they are matched with in a given period. From agents’ (ex ante) perspective, this treatment is thus equivalent to interacting with a partially benevolent default setter.

3.3 Parameters and procedures

Overall, we conducted 12 sessions of the experiment. In each of the sessions, we had 12 default setters and 12 agents that interacted for 50 periods. Subjects within a session were divided into two matching groups with 12 participants each. Default setters and agents within a given matching group were randomly rematched between periods, yielding 8 independent observations per treatment for the non-parametric tests reported below. Points earned throughout the experiment were converted at an exchange rate of 100 points = 1 Euro. Overall, sessions

⁵In both cases the unique Bayesian equilibrium is the babbling equilibrium with no information transmission.

lasted about 120 minutes, and subjects earned on average 24.32 euros (about 32 USD at the time of the experiment), including a showup fee of 4 euros.

All sessions were carried out in the BonnEconLab, the laboratory for economic experiments at the University of Bonn. The experiment was computerized using the software z-Tree (Fischbacher 2007), and subjects were recruited with the online recruitment system by Greiner (2003). A total of 288 subjects (96 in each treatment) took part in the experiment. Subjects were mainly undergraduate university students from all majors, and participated in only one of the treatment conditions. To ensure public knowledge of the rules and structure of the experiment, a summary of the instructions for the respective treatment was read out aloud at the beginning of each session. Participants then received detailed written information about the experiment.⁶ The experiment started only after all participants had answered several control questions correctly.

3.4 Hypotheses

Applying our model to the setup and parameters of the experiment yields the following predictions for differences in behavior between treatments and agent types.

Hypothesis 1. *Default setters' propensity to truthfully reveal their signal through defaults is highest in FUL, intermediate in PAR, and lowest in MIS.*

In particular, default setters in FUL should always truthfully reveal their signal. In contrast, defaults specified by fully selfish default setters (MIS) should convey no information (see Proposition 1). Since 50% of default setters in PAR are benevolent and 50 % are selfish, the truthfulness of defaults in this treatment should lie in between MIS and FUL.

Hypothesis 2. *Agents' aggregate propensity to accept defaults should be highest in FUL, intermediate in PAR, and lowest in MIS.*

This aggregate hypothesis is derived from a more specific sub-hypothesis. Proposition 2 implies that the strength of default effects is weakly increasing in the benevolence of the default setter. Applying Proposition 2 to the parameters of the experiment it predicts that agents' propensity to accept a default that is in conflict with their private information should

⁶A translation of the instructions is available upon request.

be strictly higher in FUL than in PAR, and strictly higher in PAR than in MIS for “low-information types” (type 1, 2, and 3 in Table 1).⁷ In contrast, the model predicts no treatment difference in default adherence after conflicting signals for “high-information types” (type 4 and 5 in Table 1). Similarly the application of Proposition 3 to each of the treatments yields the following hypothesis for behavioral differences in the within-subjects dimension (behavior of low-information vs. high-information types within a given treatment):

Hypothesis 3. *In case of conflicting signals, low-information agents should be strictly more likely to accept defaults than high-information agents in FUL and PAR. There should be no difference in acceptance rates between low-information and high-information types in MIS.*

Finally, the above predictions for differences in behavior yield the following hypothesis for differences in the overall quality of agents’ decisions as measured by the resulting monetary payoffs.

Hypothesis 4. *For low-information types, the quality of decisions in FUL should be strictly higher than in PAR which in turn should be strictly higher than in MIS. The quality of decisions for high-information types should not be affected by default setters’ level of benevolence ($FUL=PAR=MIS$).*

4 Results

In this section, we present the results of the experiment. We first summarize the behavior of default setters and analyze whether the informational content of defaults differs across treatments (Hypothesis 1). We then focus on the agents, and study how agents react to defaults in FUL, PAR, and MIS (Hypothesis 2). In a next step, we analyze differences in agents’ behavior, depending on their relative level of information (Hypothesis 3). Finally, we compare the quality of decisions for the different types of agents across treatments (Hypothesis 4).

⁷Since some of the predictions are equal for individual types of agents who have lower signal quality than the default setter, we jointly denote types A50, A69, and A77 from Table 1 as “low-information types” for ease of exposition. Types A94 and A100 are denoted as “high-information types”, accordingly.

4.1 How do default setters specify defaults?

Figure 1 summarizes default setters' behavior in the different treatments. The figure depicts the average frequency of defaults specified according to the private signal of default setters about the state of the world. In line with Hypothesis 1, we observe strong treatment differences in the likelihood that defaults truthfully reveal default setters' private signal. Fully benevolent default setters (FUL) almost always reveal their signal (in 98% of cases). In contrast, default setters reveal their private signal only in 75% of cases in PAR⁸ values of defaults setters. Selfish default setters set the default according to their private signal in only 56% of cases. Hence, defaults in the MIS treatment convey hardly any information about the default setters' knowledge of the state of the world. The overall difference in the

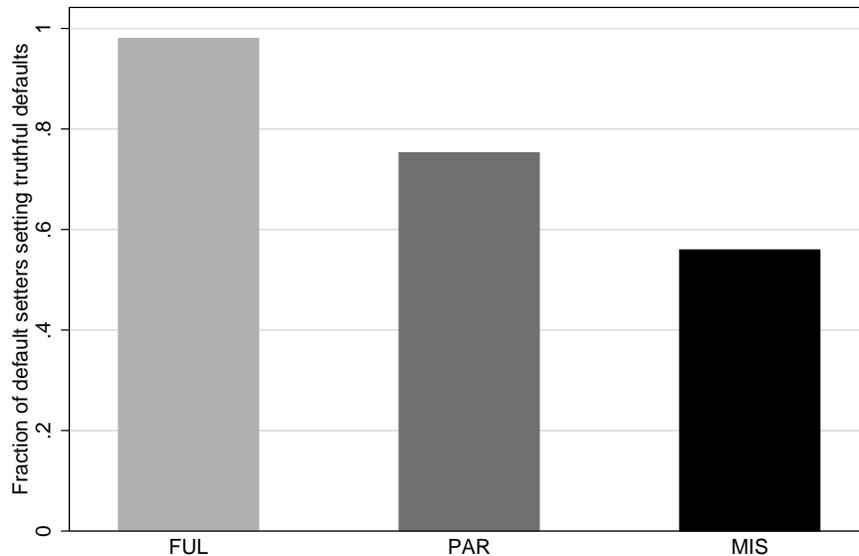


Figure 1: *Frequency of default setters setting default according to their signal. Average values per period in FUL, PAR, and MIS treatment. Predicted levels are (100 %, 75 % and 50%, respectively)*

information quality of defaults is highly significant for all pairwise treatment comparisons

⁸For the PAR treatment, we display averages over default setters with misaligned and aligned preferences, since this information is crucial for the agents.

(Fisher-exact tests, $p < .001$ for FUL vs. PAR, FUL vs. MIS, and PAR vs. MIS).⁹ Figure A.2 in the supplementary material shows that default setters' behavior is relatively stable over time in all treatments. The strong difference between treatments is already observed in the first periods, and we find no significant time trend in either of the treatments.

Overall, default setters' behavior in the experiment matches the predictions accurately. A small, but noteworthy exception is observed in the MIS treatment. The model predicts that defaults set by fully selfish default setters convey no information in equilibrium. That is, the likelihood of a default setters' signal to be red or black, should be exactly 50% independently of the observed default. Default setters' behavior in the experiment comes close to this prediction. However, the actual frequency of informative defaults in MIS is 56%. There could be two potential explanations for this deviation from the model's point predictions. First, the higher informativeness of defaults could be explained by a failure of default setters to completely shade their private information. Second, some default setters in MIS could be not fully selfish in the sense that they deliberately provide agents with information about their private signal. This could, for instance, be due to preferences for honesty (Vanberg 2008, Erat and Gneezy 2012, Fischbacher and Heusi 2013), aversion towards payoff inequalities (Fehr and Schmidt 1999, Falk and Fischbacher 2006), or other forms of social preferences. Put differently, the true preferences of some default setters might not coincide with the monetary incentives induced in the MIS treatment.

To shed light on which of the two potential explanations is likely to drive the deviation, we study behavior of individual default setters in the MIS treatment. As a measure for the truthfulness of reports we calculate the average frequency of a realized black default after a default setter received a black signal.¹⁰ We then estimate an OLS model with the default setters' individual frequency as dependent variable and different potential determinants. As a proxy for the "social preference" explanation, we include the "Honesty-Humility-Scale", HHS—a subscale of the HEXACO personality questionnaire designed to measure an individ-

⁹Unless otherwise noted, all non-parametric tests are based on matching-group averages. Reported p-values are always two-sided.

¹⁰We concentrate on the case of black signals since default setters almost unanimously select red defaults after a red signal (in 98.6% of cases). That is, default setters mostly choose to shade their private information by always selecting "red" as the default (the overall frequency of red defaults in MAL is 93.4%).

ual’s inclination to avoid manipulation of others for personal gain (Ashton and Lee 2009).¹¹

Table A.1, in the supplementary material, shows that default setters who score higher on the Honesty-Humility-Scale have a significantly higher likelihood to truthfully reveal a black signal in the MIS treatment ($p=.026$). In further specifications, we also include default setters’ scores in the Cognitive Reflection Test (Frederick 2005), and default setters’ final math grade in high school as proxies for the “cognitive mistakes” hypothesis. While the measure for honesty stays weakly significant ($p=.064$), the proxies for cognitive reflection ($p=.911$) and math abilities ($p=.507$) are not significantly related to default setters’ behavior. Our evidence thus suggests that the “too informative” defaults observed in the MIS are due to some default setters’ preferences for honest behavior, rather than due to strategic errors by the default setters.

4.2 How do agents react to defaults?

In the next step, we turn to the analysis of agents’ reactions to defaults in the different treatments. Figure 2 depicts the aggregate frequencies of default adherence for agents in FUL, PAR, and MIS. Agents accept the default specified by default setters in 90% of cases in FUL, 74% of cases in PAR, and 58% of cases in MIS. All pairwise treatment differences in agents’ behavior are statistically significant (Fisher-exact tests, $p < .001$ for FUL vs. PAR, FUL vs. MIS, and PAR vs. MIS). Figure A.3 in the supplementary material depicts the frequency of default acceptance in the different treatments over time. Again, we observe a strong difference between treatments already in the first periods and relatively stable behavior over time. The only treatment exhibiting a significant decrease in default acceptance over time is the MIS treatment.¹² Similar to default setters’ behavior, agents match the predictions of the theoretical model well. The aggregate treatment comparisons for agents’ default acceptance support the comparative static predictions of our model (Hypothesis 2).

Thus, agents account for differences in the default setters’ benevolence and the resulting differences in the informativeness of defaults on the aggregate level. This also holds for each

¹¹The HHS ranges from -60 to +40. Higher scores indicate a higher inclination to avoid manipulations.

¹²A linear time trend is significant at the 5% level in a probit estimation where the dependent variable is 1 if an agent accepts the default, and 0 otherwise ($p=.030$, accounting for potential clustering in standard errors on the matching-group level).

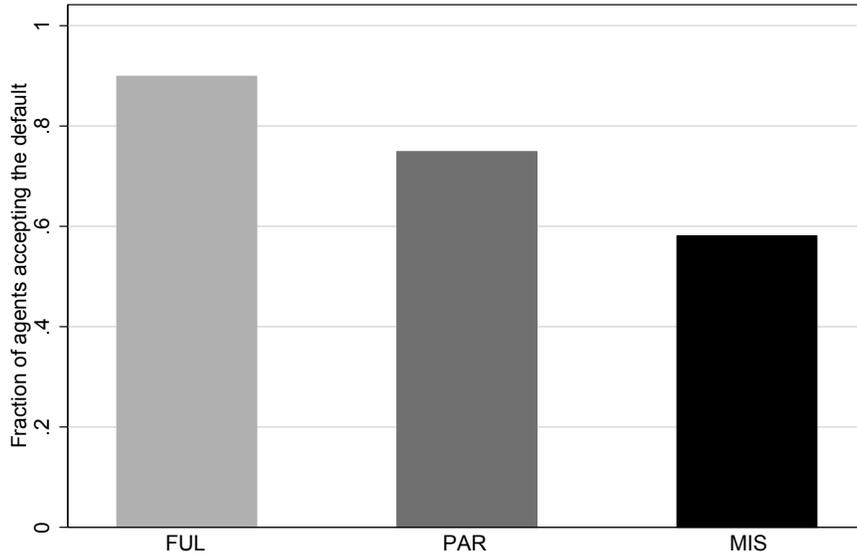


Figure 2: *Fraction of agents accepting the default. Average values in FUL, PAR, and MIS.*

type of agent. The top panel of Figure 3 depicts the average frequency of default acceptance for the five different types of agents in the different treatments.¹³ Substantiating our previous results, we find that the likelihood of accepting defaults increases in the benevolence of the default setter for each individual type of agent.

Next we analyze agents’ reactions to defaults that are at odds with their private signal. The bottom panel of Figure 3 shows the behavior of the different types of agents, if they face a default that is in conflict with their private signal (e.g., the agent observes three black and two red cards, but a red default).¹⁴ The data depicted in the graph allow us to directly test the sub-hypothesis implied by Proposition 2. In particular, we predicted that low-information types are strictly more likely to accept defaults in FUL than in PAR after conflicting signals, and strictly more likely in PAR than in MIS. In contrast, we expect no treatment effects for agents who have superior information than the default setter.

Indeed, we find that the default adherence rate for low-information types if the default is in conflict with their private signal is increasing in the benevolence of the default setter

¹³Agents are ordered according to their signal quality. Type “A69” in Figure 3 are agents who received an a signal with 69 % correlation in a given period (Type 2 in Table 1).

¹⁴Note that type A50 is excluded from the analysis. Since this types has no informative signal (1 red card and 1 black card), there can be no conflict between own signal and the default.

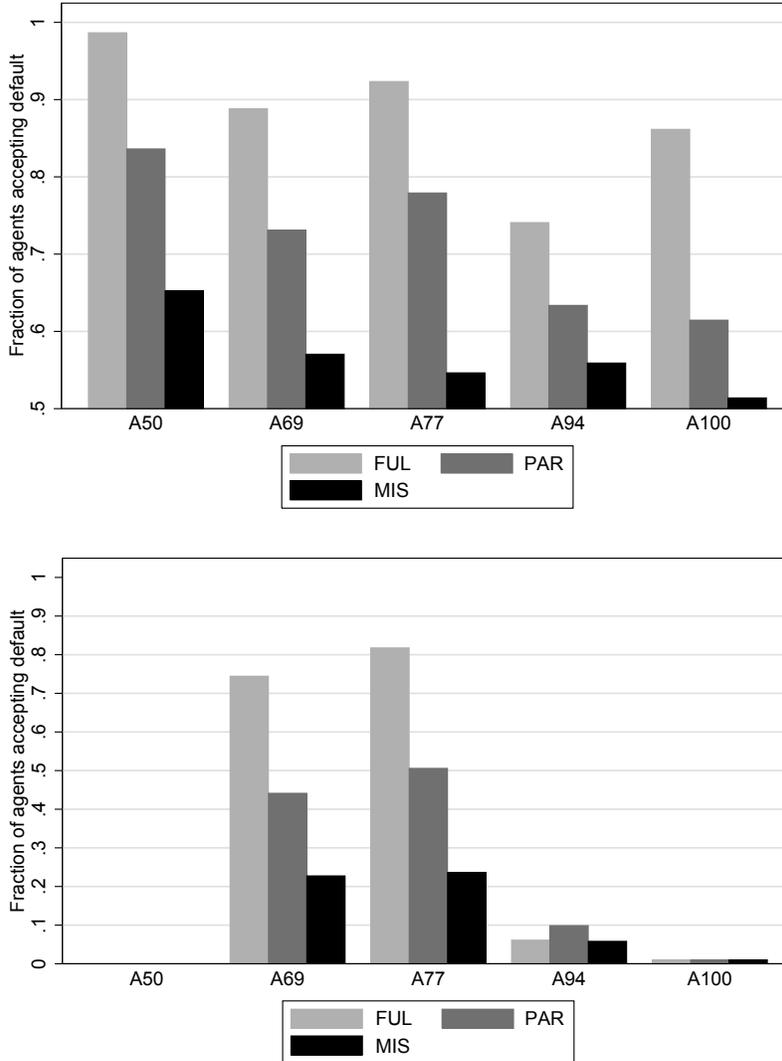


Figure 3: *Fraction of agents accepting the default. Upper panel: average aggregate values for different agent types in FUL, PAR, and MIS. Bottom panel: cases where default is in conflict with agents' private signal; average values for different agent types in FUL, PAR, and MIS.*

(Fisher-exact test for low-information types. FUL vs. PAR: $p < .001$, FUL vs. MIS: $p < .001$, PAR vs MIS: $p = .010$). On the contrary, the reactions of high-information types do not differ significantly across treatments (Fisher-exact test for high-information types. FUL vs. PAR: $p = 1.000$, FUL vs. MIS: $p = .619$, PAR vs MIS: $p = .619$).

The bottom panel of Figure 3 also allows to test Hypothesis 3, which predicts that high-information types are less likely to accept conflicting defaults than agents with lower quality

information when facing a (partially) benevolent default setter. This within treatment effect is born out by the data (Wilcoxon signed-rank tests for high-information types vs. low-information types, $p = .012$ for FUL and PAR). However, the difference remains significant when analyzing the MIS treatment, while we hypothesized that there are no type-specific differences in behavior as a response to selfish defaults (Wilcoxon signed-rank test, $p = .012$). The top panel of Figure 3 shows that the diverging reactions of agents to conflicting defaults induce overall differences in type specific default acceptance rates. Agents who are better informed than the default setter (i.e., Type A94 and A100) are less likely to accept defaults than agents with inferior information quality if the default setter is (partially) benevolent. A Wilcoxon signed-rank test confirms that high-information types behave significantly different than low-information types in both the FUL and PAR treatment ($p = .012$, for both FUL and PAR). When being confronted with selfish default setters, the effect is less pronounced and turns out to be insignificant (Wilcoxon signed-rank test, $p = .161$). This is in line with the prediction of our theoretical model: Since defaults convey no information in a babbling equilibrium, all types of agents should merely rely on their private signal.

Overall, the empirical results support Hypothesis 2 and Hypothesis 3 from our model. However, we again find some deviations from the model’s predictions. In particular, agents with low and intermediate levels of information seem to trust defaults in the MIS treatment “too much”. This becomes most evident when looking at behavior of type A69 and A77 in case of conflicting signals (bottom panel of Figure 3). Our model predicts that all types of agents completely ignore defaults set by a fully selfish default setter. For conflicting signals, we thus expect default acceptance rates close to zero for A69 and A77, whereas the acceptance rates in the experiment are about 20% for both types of agents.¹⁵

We have already seen that some default setters in the MIS treatment do not behave in a fully selfish way. That is, defaults in the MIS treatment are—on the aggregate level—not fully uninformative from the agents’ perspective. This raises the question whether the default acceptance rate by low-information agents that we observe is caused by “too much trust” in selfish defaults (i.e., a mistake by low-information agents) or rather a best response to the

¹⁵A similar effect is observed in the PAR treatment where we observe acceptance rates of 40-50% instead of 20-30% as predicted by our model, when weighing all types with their respective occurrence probability.

informational content of defaults, that agents experience during the experiment. If agents' choices are a best response to the behavior of default setters, the overall quality of agents' decision should be as least as high as in a (hypothetical) situation without defaults in which agents always follow their private signal. The next section explores in more detail whether this is the case.

4.3 Do defaults improve decisions?

In a final step of our empirical analysis, we turn to the question whether defaults improve the overall quality of agents' decisions in our experiment, and how the impact on decision quality interacts with the agents' level of information and the default setters' level of benevolence. Figure 4 depicts the percentage change in the agents' decision quality compared to a hypothetical situation in which agents are assumed to always follow their private signals.¹⁶ This scenario mimics a choice environment in which there are no default options and agents can base their decision only on their private signal. Positive values of the difference between actual and hypothetical decision quality in Figure 4 thus indicate that agents in the experiment benefitted from defaults, while negative values mean that they would have attained higher payoffs in the absence of default options.¹⁷

Figure 4 indicates that the decision quality of high-information types (A94, A100) is hardly affected by the presence of defaults. This holds independently of the alignment of preferences between default setters and decision makers, and it reiterates the observation that these types react appropriately to their high signal quality: whenever the default option is in conflict with their private information, these agents tend to rely on the latter. Thus, there are little or no distortions in the decisions of well-informed agents.

For agents with lower information quality, the extent to which default options are welfare-enhancing depends on the agent's level of information and, more importantly, on the alignment of interests with the default setter. In the FUL treatment where default setter's and

¹⁶To construct a measure of decision quality in this situation, we use the sets of cards and the actually realized private signals in the experiment for each agent.

¹⁷Note that, strictly speaking, this analysis provides a lower bound on the potentially beneficial effects of defaults in our setup, since it assumes that agents make no mistakes in the decision environment without defaults.

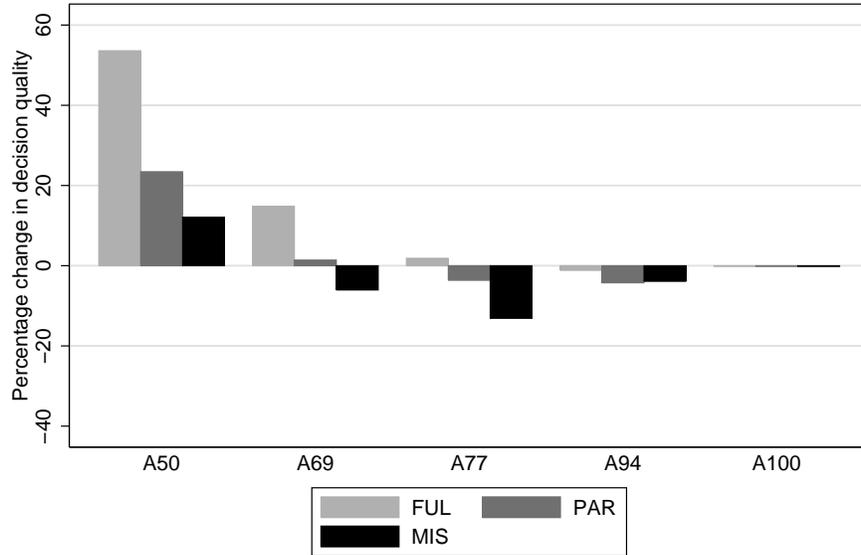


Figure 4: *Percentage change in agents’ decision quality compared to a hypothetical situation in which agents always follow their own signal, for different agent types and treatments.*

agents’ information are fully aligned, we observe generally positive effects of default options, relative to the situation without default specifications. In particular the agents with the lowest quality of private information (A50 and A69) tend to make use of the informativeness of defaults which are set by the better informed default setter, and thereby attain higher payoffs.

For the PAR and MIS treatment, our findings on the effects of defaults for agents with low information quality are somewhat mixed. First, we observe unambiguously positive effects on the decision quality of agents with no information (A50). For type A69 who has intermediate-level quality, defaults set by partially benevolent or selfish default setters have almost no impact on the decision quality, with slightly positive overall consequences in the PAR treatment, and slightly negative ones in the MIS treatment. Finally, the numbers for type A77 suggest that agents with informational levels similar to the one of the default setter do not discount the informational content of the default strong enough to account for the misalignment of preferences. Overall, this finding suggests that a sound understanding of the strategic incentives and the relative quality of one’s own and the default setters’ information are crucial for agents to reap the benefits of informative defaults, without bearing the detrimental consequences of following ill-specified defaults from default setters with misaligned

interests.¹⁸

5 Conclusions

In this paper, we have provided a theoretical model of two key channels through which non-binding default options can affect behavior—the strategic incentives of default setting institutions, and the default setter’s and decision makers’ relative level of information about the decision environment. While both aspects have been discussed informally in the literature, our model provides a unifying framework for analyzing how individual behavior and the overall strength of default effects are shaped by both factors. Our model yields two main testable predictions. First, default options should have a stronger impact on behavior, if interests of the default setter and decision makers are more closely aligned. Second, whenever decision makers experience conflicts between their own information about optimal choices and a default option, those with less information are more prone to accept the default.

In the second part of the paper, we test these predictions in a laboratory experiment that allows us to exogenously vary the key parameters of interest. Our empirical results show that the informational content of defaults and, thus, the overall strength of default effects are increasing in the alignment of interests. In the second dimension, we show that agents with lower information quality are indeed more susceptible to accept default options, if these are in conflict with an agent’s individual information.

On a more general level, our analysis suggests that an integrative perspective on defaults needs to incorporate both, the psychological factors that may strengthen default effects like status quo biases or present-biased preference (Samuelson and Zeckhauser 1988, Carroll et al. 2009), but also the incentives created by the strategic interaction that is inherent in the default setting process. Our model is flexible enough to be extended such that the interaction of these effects can be studied. Our theoretical framework also lends itself to a number of other interesting extensions. These include integrating the possibility that agents can spend cognitive resources for acquiring additional information (Caplin and Martin 2013a), the possibility of heterogeneous preferences of agents for a given state of the world, or allowing

¹⁸Attaining such understanding might, admittedly, be challenging: findings by Cain et al. (2005) indicate that agents sometimes trust the information given by selfish advisers too strongly even if they are informed about potential conflicts of interest.

for imperfect information about the opponent’s interests.

Our findings also provide insights for the discussion on “libertarian paternalistic” policy interventions. In particular, most participants in our experiment do take the default setter’s strategic incentives and information into account, and condition their acceptance of default options on both factors. However, our empirical results also suggest that a high level of comprehension of default setters’ information and strategic incentives are crucial to ensure that consumers react appropriately to default options. If some agents are naive about strategic incentives of the default setter, this suggests a scope for mandatory disclosure of incentives and other instruments of consumer protection legislation. Using our theoretical framework to incorporate naive types and study how consumer protection policies should be designed in choice environments with default options thus seems to be a promising next step for future research—in paralleling the approach that has recently been taken in the literature on financial advice (see Inderst and Ottaviani 2012a, Inderst and Ottaviani 2012b).

References

- Ashton, M. and K. Lee (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment* 91, 340–345.
- Borghans, L. and B. H. H. Golsteyn (2013). Default options and training participation. *Empirical Economics* (forthcoming).
- Brown, C. L. and A. Krishna (2004). The skeptical shopper: A metacognitive account for the effects of default options on choice. *Journal of Consumer Research* 31(3), 529–539.
- Cain, D. M., G. Loewenstein, and D. Moore (2005). The dirt on coming clean: Perverse effects of disclosing conflicts of interest. *The Journal of Legal Studies* 34(1), 1–25.
- Camerer, C., S. Issacharoff, G. Loewenstein, T. O’Donoghue, and M. Rabin (2003). Regulation for conservatives: Behavioral economics and the case for “asymmetric paternalism”. *University of Pennsylvania Law Review* 151, 1211–1254.
- Caplin, A. and D. Martin (2013a). Defaults and attention: The drop out effect. *Working Paper, NYU*.

- Caplin, A. and D. Martin (2013b). A testable theory of imperfect perception. *Working Paper, NYU*.
- Carroll, G. D., J. J. Choi, D. Laibson, B. C. Madrian, and A. Metrick (2009). Optimal defaults and active decisions. *The Quarterly Journal of Economics* 124(4), 1639–1674.
- Choi, J., D. Laibson, and B. C. Madrian (2004). Plan design and 401(k) savings outcomes. *National Tax Journal* 57(2), 275–298.
- Crawford, V. and J. Sobel (1982). Strategic information transmission. *Econometrica* 50(6), 1432–1451.
- Erat, S. and U. Gneezy (2012). White lies. *Management Science* 58(4), 723–733.
- Falk, A. and U. Fischbacher (2006). A theory of reciprocity. *Games and Economic Behavior* 54(2), 293–315.
- Fehr, E. and K. M. Schmidt (1999). A Theory of Fairness, Competition and Cooperation. *The Quarterly Journal of Economics* 114, 817–868.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10, 171–178.
- Fischbacher, U. and F. Heusi (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association* 11(3), 525–547.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives* 19(4), 25–42.
- Glaeser, E. L. (2005). Paternalism and psychology. *Harvard Institute of Economic Research, Discussion Paper 2097*.
- Greiner, B. (2003). An Online Recruitment System for Economic Experiments. In K. Kremer and V. Macho (Eds.), *Forschung und wissenschaftliches Rechnen 2003. GWDG Bericht 63*, pp. 79–93. Göttingen: Ges. für Wiss. Datenverarbeitung.
- Inderst, R. and M. Ottaviani (2012a). Financial advice. *Journal of Economic Literature* 50(2), 494–512.
- Inderst, R. and M. Ottaviani (2012b). How (not) to pay for advice: A framework for consumer protection. *Journal of Financial Economics* 105(2), 393–411.
- Johnson, E. J. and D. Goldstein (2003). Do defaults save lives? *Science* 302, 1338–1339.

- Johnson, E. J., J. Hershey, J. Meszaros, and H. Kunreuther (1993). Framing, probability distortions, and insurance decisions. *Journal of Risk and Uncertainty* 7, 35–51.
- Kahneman, D., J. L. Knetsch, and R. H. Thaler (1991). The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives* 5(1), 193–206.
- Levav, J., M. Heitmann, A. Herrmann, and S. S. Iyengar (2010). Order in product customization decisions: Evidence from field experiments. *Journal of Political Economy* 118(2), 274–299.
- Madrian, B. C. and D. F. Shea (2001). The power of suggestion: Inertia in 401(k) participation and savings behavior. *The Quarterly Journal of Economics* 116, 1149–1187.
- McKenzie, C. R., M. J. Liersch, and S. R. Finkelstein (2006). Recommendations implicit in policy defaults. *Psychological Science* 17(5), 414–420.
- Samuelson, W. and R. Zeckhauser (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty* 1(1), 7–59.
- Schwartz, A. and R. E. Scott (2003). Contract theory and the limits of contract law. *Yale Law Journal* 113, 541–619.
- Sunstein, C. R. (2012). The Storrs lectures: Behavioral economics and paternalism. *Yale Law Journal* (forthcoming).
- Sunstein, C. R. (2013). Impersonal default rules vs. active choices vs. personalized default rules: A triptych. *Working Paper, Harvard University*.
- Tannenbaum, D. (2011). Information asymmetries in policy defaults: Some defaults speak louder than others. *Working Paper, University of California, Irvine*.
- Thaler, R. H. and C. R. Sunstein (2003). Libertarian paternalism. *American Economic Review* 93, 175–179.
- Vanberg, C. (2008). Why do people keep their promises? An experimental test of two explanations. *Econometrica* 76(6), 1467–1480.

Appendix

Proof of Lemma 1

Suppose strategy profile s with strategies s_A^x, s_P constitutes a Bayesian equilibrium. Denote by D_1 the set of all defaults being played in equilibrium and by D_2 the nonempty set of out of equilibrium defaults. Take any $d_i \in D_1$ and $d_j \in D_2$.

$$s_P : \rho_l \rightarrow (\dots, p(d_i) = \alpha_1, \dots, p(d_j) = 0, \dots), \rho_h \rightarrow (\dots, p(d_i) = \alpha_2, \dots, p(d_j) = 0, \dots)$$

Define a new strategy of the default setter:

$$\hat{s}_P : \rho_l \rightarrow (\dots, p(d_i) = \frac{\alpha_1}{2}, \dots, p(d_j) = \frac{\alpha_1}{2}, \dots), \rho_h \rightarrow (\dots, p(d_i) = \frac{\alpha_2}{2}, \dots, p(d_j) = \frac{\alpha_2}{2}, \dots)$$

Moreover, construct new strategies \hat{s}_A^x for the agents such that $\hat{s}_A^x(d_j, \sigma) = \hat{s}_A^x(d_i, \sigma) = s_A^x(d_i, \sigma)$. Leaving everything else fixed \hat{s}_A^x, \hat{s}_P also constitute a Bayesian equilibrium. First note that $p(\theta|d_i, \sigma) = p(\theta|d_j, \sigma)$ for both θ . According to Bayes' rule these are also equivalent to the conditional probabilities following d_i in the original equilibrium. Hence, agents' maximization problem is identical and the best responses do not change. Given their strategy the \hat{s}_P must also be a best response, since messages d_i, d_j induce the same action and the default setter is indifferent between them. By subsequently adding all out of equilibrium defaults to the set of equilibrium defaults we can construct an output equivalent equilibrium without any out of equilibrium defaults.

Proof of Lemma 2

Since we consider informative equilibria, there are at least two different defaults in the sense that $p(\theta_h|d_k) > p(\theta_h|d_f)$. Given the signal ρ_h , we show that it is always more profitable to play d_k for the default setter. The incentive function of the default setter i.e the expected

payoff difference from d_k versus d_f is:

$$\begin{aligned}
& \int_{\frac{1}{2}}^1 E_{\sigma, \theta} [U_P(s_A^x(\sigma, d_k)) | \rho_h, x] df(x) - \int_{\frac{1}{2}}^1 E_{\sigma, \theta} [U_P(s_A^x(\sigma, d_f)) | \rho_h, x] df(x) \\
&= \int_{\frac{1}{2}}^1 p(\sigma_l, \theta_l | \rho_h, x) (\mu(U_A(\theta_l, s_A^x(\sigma_l, d_k)) - U_A(\theta_l, s_A^x(\sigma_l, d_f))) + b(s_A^x(\sigma_l, d_k)) - b(s_A^x(\sigma_l, d_f))) \\
&+ p(\sigma_h, \theta_l | \rho_h, x) (\mu(U_A(\theta_l, s_A^x(\sigma_h, d_k)) - U_A(\theta_l, s_A^x(\sigma_h, d_f))) + b(s_A^x(\sigma_h, d_k)) - b(s_A^x(\sigma_h, d_f))) \\
&+ p(\sigma_l, \theta_h | \rho_h, x) (\mu(U_A(\theta_h, s_A^x(\sigma_l, d_k)) - U_A(\theta_h, s_A^x(\sigma_l, d_f))) + b(s_A^x(\sigma_l, d_k)) - b(s_A^x(\sigma_l, d_f))) \\
&+ p(\sigma_h, \theta_h | \rho_h, x) (\mu(U_A(\theta_h, s_A^x(\sigma_h, d_k)) - U_A(\theta_h, s_A^x(\sigma_h, d_f))) + b(s_A^x(\sigma_h, d_k)) - b(s_A^x(\sigma_h, d_f))) df(x).
\end{aligned}$$

The differences of the selfish parts $b(z)$ of the utility function are clearly larger or equal to zero, since the agents' best responses to d_k are weakly higher z s. We can, thus, concentrate on the utility difference of the agents. The integral of these is definitely larger than zero if this is fulfilled for all x . Consider the difference for an arbitrary x with slightly rewritten conditional probabilities:

$$\begin{aligned}
& p(\sigma_l | \rho_h, x) p(\theta_l | \sigma_l, \rho_h, x) (U_A(\theta_l, s_A^x(\sigma_l, d_k)) - U_A(\theta_l, s_A^x(\sigma_l, d_f))) \\
&+ p(\sigma_h | \rho_h, x) p(\theta_l | \sigma_h, \rho_h, x) (U_A(\theta_l, s_A^x(\sigma_h, d_k)) - U_A(\theta_l, s_A^x(\sigma_h, d_f))) \\
&+ p(\sigma_l | \rho_h, x) p(\theta_h | \sigma_l, \rho_h, x) (U_A(\theta_h, s_A^x(\sigma_l, d_k)) - U_A(\theta_h, s_A^x(\sigma_l, d_f))) \\
&+ p(\sigma_h | \rho_h, x) p(\theta_h | \sigma_h, \rho_h, x) (U_A(\theta_h, s_A^x(\sigma_h, d_k)) - U_A(\theta_h, s_A^x(\sigma_h, d_f))).
\end{aligned}$$

The default setter's expectation about the difference in payoffs for the agent is larger than the agents own prediction, which in turn is given by

$$\begin{aligned}
& p(\sigma_l | \rho_h, x) p(\theta_l | \sigma_l, d_h, x) (U_A(\theta_l, s_A^x(\sigma_l, d_k)) - U_A(\theta_l, s_A^x(\sigma_l, d_f))) \\
&+ p(\sigma_h | \rho_h, x) p(\theta_l | \sigma_h, d_h, x) (U_A(\theta_l, s_A^x(\sigma_h, d_k)) - U_A(\theta_l, s_A^x(\sigma_h, d_f))) \\
&+ p(\sigma_l | \rho_h, x) p(\theta_h | \sigma_l, d_h, x) (U_A(\theta_h, s_A^x(\sigma_l, d_k)) - U_A(\theta_h, s_A^x(\sigma_l, d_f))) \\
&+ p(\sigma_h | \rho_h, x) p(\theta_h | \sigma_h, d_h, x) (U_A(\theta_h, s_A^x(\sigma_h, d_k)) - U_A(\theta_h, s_A^x(\sigma_h, d_f))).
\end{aligned}$$

This is clearly larger or equal to zero, because agents maximize their expected utility. The integral can only attain zero, if the actions following d_k and d_f are equal for all agents. Since there is a continuum of agents with full support over the interval $[\frac{1}{2}, 1]$, this can never be the case. We conclude that default d_f is played with probability zero whenever the default setter receives a high signal. Hence, every default played in response to a high signal inhibits the same information. All other defaults reveal that the default setter received a low signal.

Proof of Proposition 1

For $\mu \rightarrow \infty$, the preferences of the default setter and the agents are fully aligned. Hence, full information is the only possible equilibrium if there is information transferred. Since preferences are continuous in μ , there exists a $\bar{\mu}$ such that for all $\mu \geq \bar{\mu}$ full information transmission is the only informative equilibrium. Furthermore, this equilibrium clearly Pareto-dominates the babbling equilibrium, which always exists. If μ is equal to zero, however, the default setter always chooses the default inducing the highest z . As a consequence, defaults cannot transfer any information. With the same argument as above there exists a $\underline{\mu}$ such that for all $\mu \leq \underline{\mu}$ defaults inhibit no informational content in equilibrium.

Proof of Proposition 2

In the following we show that c^{pd} is increasing in μ along the path of Pareto-efficient equilibria. To simplify notation, we write D_i for any default from the set D_i . Note first that $\int_{\frac{1}{2}}^1 E_{\sigma,\theta}[U_P(\theta, s_A^x(\sigma, D_l))|\rho_h, x]df(x)$ is constant in c^{pd} as explained in Lemma 3. In any mixed equilibrium, the default setter is indifferent between all defaults given she received a low signal

$$\int_{\frac{1}{2}}^1 E_{\sigma,\theta}[U_P(\theta, s_A^x(\sigma, D_h))|\rho_l, x]df(x) - \int_{\frac{1}{2}}^1 E_{\sigma,\theta}[U_P(\theta, s_A^x(\sigma, D_l))|\rho_l, x]df(x) = 0.$$

The change of c^{pd} with increasing μ can be derived by implicit differentiation of this equilibrium condition.

$$\begin{aligned} \frac{dc^{pd}}{d\mu} &= - \frac{\partial \left(\int_{\frac{1}{2}}^1 E_{\sigma,\theta}[U_P(\theta, s_A^x(\sigma, D_h))|\rho_l, x]df(x) - \int_{\frac{1}{2}}^1 E_{\sigma,\theta}[U_P(\theta, s_A^x(\sigma, D_l))|\rho_l, x]df(x) \right) / \partial \mu}{\partial \left(\int_{\frac{1}{2}}^1 E_{\sigma,\theta}[U_P(\theta, s_A^x(\sigma, D_h))|\rho_l, x]df(x) - \int_{\frac{1}{2}}^1 E_{\sigma,\theta}[U_P(\theta, s_A^x(\sigma, D_l))|\rho_l, x]df(x) \right) / \partial c^{pd}} \\ &= - \frac{\partial \left(\int_{\frac{1}{2}}^1 E_{\sigma,\theta}[U_P(\theta, s_A^x(\sigma, D_h))|\rho_l, x]df(x) - \int_{\frac{1}{2}}^1 E_{\sigma,\theta}[U_P(\theta, s_A^x(\sigma, D_l))|\rho_l, x]df(x) \right) / \partial \mu}{\partial \left(\int_{\frac{1}{2}}^1 E_{\sigma,\theta}[U_P(\theta, s_A^x(\sigma, D_h))|\rho_l, x]df(x) \right) / \partial c^{pd}} \end{aligned}$$

The numerator of this fraction is smaller than zero, if

$$\begin{aligned} &\int_{\frac{1}{2}}^1 p(\sigma_l, \theta_l | \rho_l, x) (U_A(\theta_l, s_A^x(\sigma_l, D_h)) - U_A(\theta_l, s_A^x(\sigma_l, D_l))) \\ &\quad + p(\sigma_h, \theta_l | \rho_l, x) (U_A(\theta_l, s_A^x(\sigma_h, D_h)) - U_A(\theta_l, s_A^x(\sigma_h, D_l))) \\ &\quad + p(\sigma_l, \theta_h | \rho_l, x) (U_A(\theta_h, s_A^x(\sigma_l, D_h)) - U_A(\theta_h, s_A^x(\sigma_l, D_l))) \\ &\quad + p(\sigma_h, \theta_h | \rho_l, x) (U_A(\theta_h, s_A^x(\sigma_h, D_h)) - U_A(\theta_h, s_A^x(\sigma_h, D_l))) df(x) \leq 0 \end{aligned}$$

This follows if the inequality holds for all x . Since defaults convey less information than the signal of the default setter and since agents maximize their utility, this holds, because

$$\begin{aligned}
& p(\sigma_l|\rho_l, x)p(\theta_l|\sigma_l, D_l, x)(U_A(\theta_l, s_A^x(\sigma_l, D_h)) - U_A(\theta_l, s_A^x(\sigma_l, D_l))) \\
& + p(\sigma_h|\rho_l, x)p(\theta_l|\sigma_h, D_l, x)(U_A(\theta_l, s_A^x(\sigma_h, D_h)) - U_A(\theta_l, s_A^x(\sigma_h, D_l))) \\
& + p(\sigma_l|\rho_l, x)p(\theta_h|\sigma_l, D_l, x)(U_A(\theta_h, s_A^x(\sigma_l, D_h)) - U_A(\theta_h, s_A^x(\sigma_l, D_l))) \\
& + p(\sigma_h, \rho_l, x)p(\theta_h|\sigma_h, D_l, x)(U_A(\theta_h, s_A^x(\sigma_h, D_h)) - U_A(\theta_h, s_A^x(\sigma_h, D_l))) \leq 0.
\end{aligned}$$

Consequently, the numerator is negative, yielding

$$\frac{dc^{pd}}{d\mu} > 0 \Leftrightarrow \frac{\partial \int_{\frac{1}{2}}^1 E_{\sigma, \theta}[U_P(\theta, s_A^x(\sigma, D_h))|\rho_l, x]df(x)}{\partial c^{pd}} > 0.$$

Lemma 2 states that the equilibrium with the highest information transmission rate is Pareto-efficient. Suppose that the latter derivative is negative. As a consequence, the payoff from a low default is larger than the one from a high default for all transmission rates larger than c^{pd} . In particular, this also holds for $c = 1$, implying that there exists an equilibrium with full information transmission. This is a contradiction to the assumption that we are in a mixed Pareto-efficient equilibrium. There can, thus, only exist mixed Pareto-efficient equilibria with positive derivative and $\frac{dc^{pd}}{d\mu} \geq 0$. Since agents are Bayesian updaters, this leads to weakly stronger responses by the agents, i.e. $s_A^x(\sigma_i, D_h)$ is increasing and $s_A^x(\sigma_i, D_l)$ is decreasing in μ .

Proof of Proposition 3

We assumed that the partial derivative with respect to z of $U_A(z, \theta_l)$ is smaller than the corresponding derivative of $U_A(z, \theta_h)$ for all z . Hence, higher z s are optimal if the agent puts more probability weight on θ_h . Since agents are Bayesian updaters, they weigh their private signal stronger, if it conveys more information. Hence, for any given equilibrium, agents are more susceptible to the default, if they have less private information.

A. Supplementary Figures and Tables

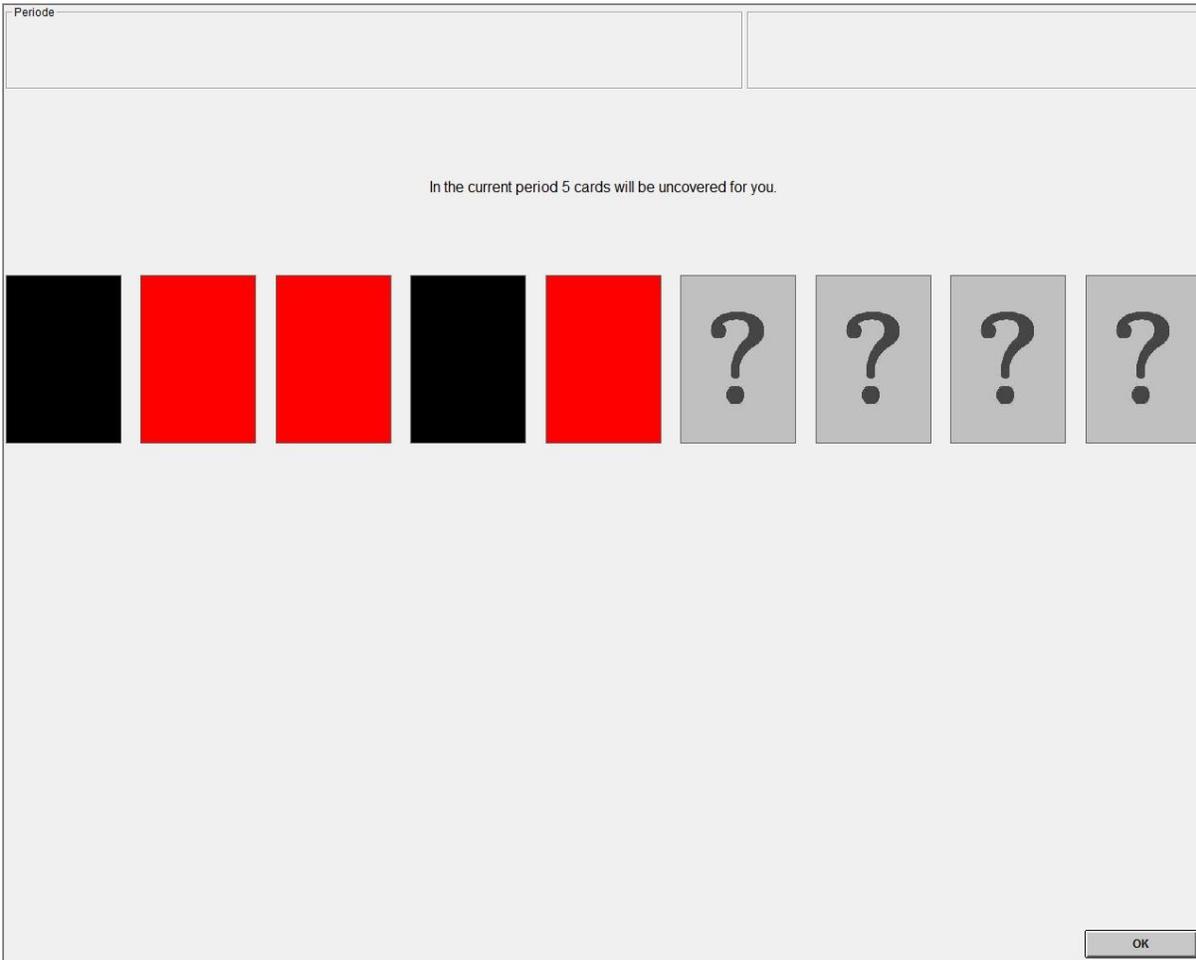


Figure A.1: *Agent's information screen. Example in which two black and three red cards are revealed.*

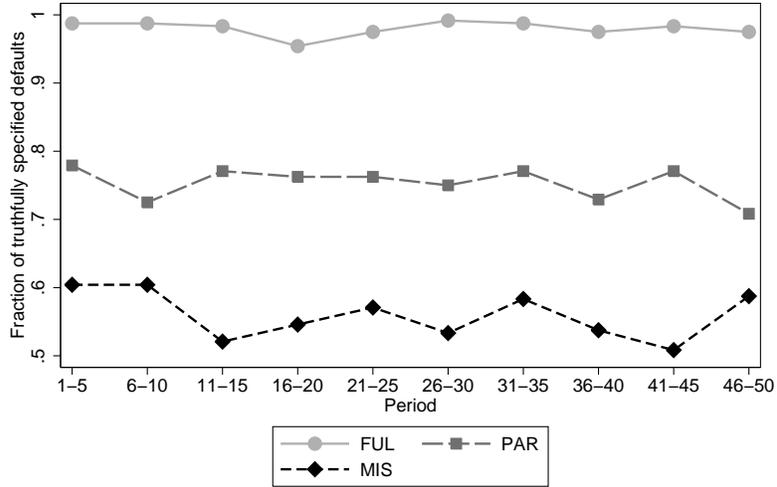


Figure A.2: Frequency with which default setters set default according to their signal. Average values for 5-period intervals in FUL, PAR, and MIS.

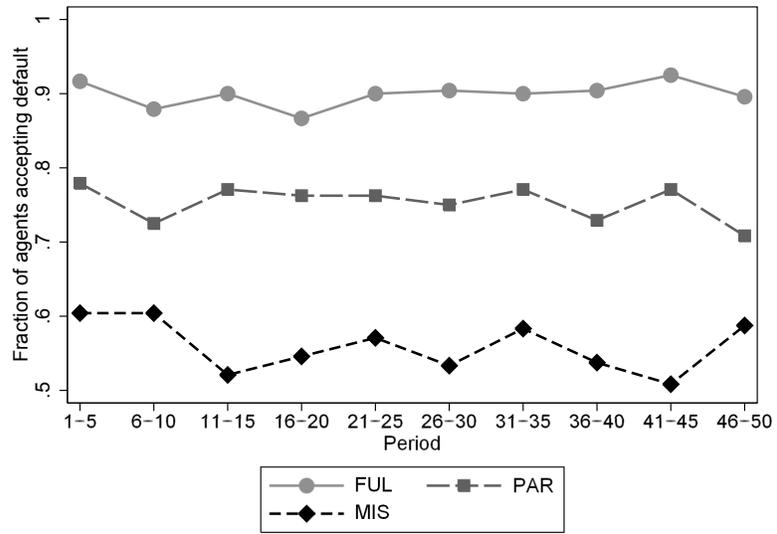


Figure A.3: Frequency with which agents follow default option. Average values for 5-period intervals in FUL, PAR, and MIS.

Dependent variable: Frequency with which default setter chooses black default after black signal (MIS treatment)			
	(1)	(2)	(3)
HHS	.341** (.121)		.314* (.143)
CRT		-.859 (3.221)	-.381 (3.290)
Math grade		2.926 (3.354)	2.621 (3.751)
Const.	12.573*** (1.273)	6.870 (11.294)	6.927 (13.021)
N	48	48	48
R^2	.098	.502	.134

Table A.1: *Determinants of default setters' behavior in MIS. OLS estimations; reported standard errors (in parentheses) account for potential clustering on the matching-group level. The honesty-humility-scale, HHS, measures an individual's inclination to avoid manipulation of others for personal gain (Ashton and Lee 2009). It ranges from -60 to +40, which higher scores indicating a higher inclination to avoid manipulations. The cognitive-reflection test, CRT, is a three-item questionnaire that measures an individual's inclination to suppress a spontaneous, but wrong answer (Frederick 2005). "Math grade" measures participants' final high-school math grade. Grades range from 1-6 with 1 being the best grade.*