Can Incentives Cause Harm?

Tests of Undue Inducement

Sandro Ambuehl*

August 10, 2021

Abstract

Around the world, laws limit incentives for transactions such as human research participation, egg donation, or gestational surrogacy. A key reason is the notion of undue inducement—the conceptually vague and empirically largely untested idea that incentives cause harm by distorting individual decision making. Two experiments, including one based on a highly visceral transaction, show that incentives bias information search. Yet, such behavior is also consistent with Bayes-rational behavior. I develop a criterion that indicates whether choices admit welfare weights on benefit and harm that justify permitting the transaction but capping incentives. In my experimental data, no such weights exist.

Keywords: Incentives, Undue Inducement, Behavioral Welfare Economics, Experiment

JEL Codes: C91, D6, D9

*Department of Economics and UBS Center for Economics in Society, University of Zurich, Blüemlisalpstrasse 10, 8006 Zürich, Switzerland, sandro.ambuehl@econ.uzh.ch. This paper supersedes a previous paper entitled “An Offer You Can’t Refuse? Incentives Change What We Believe.” I am deeply indebted to my PhD advisors Muriel Niederle, B. Douglas Bernheim, and Alvin E. Roth. I am grateful to Yan Chen, P.J. Healy, and Axel Ockenfels for providing experimental laboratory resources. Experiment 1 was funded by the Department of Economics at the University of Zurich. Experiment 2 was funded by the Department of Economics at Stanford University. It was approved by the Stanford IRB in protocol 34001. Moreover, this project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 741409); the results reflect the authors’ view; the ERC is not responsible for any use that may be made of the information it contains.

1 Introduction

Around the world, laws and guidelines limit the incentives that can be paid for transactions such as human research participation, surrogate motherhood, human egg donation, and organ donation. These transactions are often legal if no or limited amounts of money change hands. An important reason for the limits on incentives is the influential, albeit conceptually vague and empirically largely untested notion of *Undue Inducement* (Belmont Report, 1978). It describes the belief that participation incentives can harm participants, because “something is being offered that is alluring to the point that it clouds rational judgment ... Attention is fixated on the benefit, disallowing proper consideration of the risks” (Stanford Encyclopedia of Philosophy, Eyal, 2012). The American Society for Reproductive Medicine (2007), for instance, urges that “Payments to women providing oocytes should be ... not so substantial that they ... lead donors to discount risks,” and speculates that “the higher the payment, the greater the possibility that women will discount risks.” The National Bioethics Advisory Commission (2001) limits incentives that can be offered for participation in clinical trials based on the assertion that “[o]ffers of large sums of money ... could lead some prospective participants to enroll ... when it might be against their better judgment.” Observe that this *Undue Inducement Hypothesis* consists of two distinct claims. The *behavioral part* ascertains that incentives will cause participants to engage in biased information processing and motivated reasoning about the transaction. Indeed, as Largent et al. (2013) explain, “undue influence is a cognitive distortion relating to assessment of risks and benefits.” The *normative part* claims that these changes cause harm, and that incentives should therefore be capped.¹

This paper formalizes and experimentally tests the Undue Inducement Hypothesis to answer two questions: 1. *How do participation incentives change the way in which individuals learn about the nature of the transaction?* 2. *Do participation incentives cause harm?* These questions are important. Medical trials, for instance, are routinely hampered by recruitment difficulties (McDonald et al., 2006), and demand for donor kidneys greatly outstrips supply (Hippen et al., 2009). The answer is not obvious. Existing literature provides neither a clear conceptualization of the Undue Inducement Hypothesis nor a rigorous analysis of the welfare effects of participation incentives. Direct empirical evidence on the topic is limited to a small number of case studies and unincentivized surveys (Bentley and Thacker, 2004; Halpern et al., 2004; Singer and Couper, 2008; Stunkel and Grady, 2011; Devine et al., 2013; Mngadi et al., 2017).² Indirect evidence from the vast literature on motivated reasoning is no proof for the Undue Inducement Hypothesis as it typically considers settings in which the bias benefits rather than harms the decision maker.³

¹Regarding organ donation, for instance, Satz (2010) mentions “a paternalistic concern that sellers would actually be harmed by the sale of their organs, but that ... they would sell their organs if it were legal.”

²This small amount of empirical research contrasts with hundreds of publications in the medical ethics literature that speculate and opine about the effects of monetary incentives on research subjects without any empirical substantiation.

³In experiments on motivated reasoning, the bias benefits decision makers by allowing them, for instance, to feel better about unchangeable facts (Festinger, 1962; Engelmann et al., 2019), to dodge costly social expectations (Dana, Weber and Kuang, 2007; Exley, 2015; Exley and Kessler, 2018), and to manage self-image without direct material consequences (Gneezy, Saccardo, Serra-Garcia and van Veldhuizen, 2015). A related literature on ‘positivity bias’ shows that people generally tend to view the world in overly rosy terms (Weinstein, 1989; Sharot et al., 2011) including in the case of medical risks (Weinstein, 1980; Oster et al., 2013). While individuals affected by positivity bias will sometimes mistakenly engage in transactions because they underestimate the risks and downsides, positivity bias does not by itself imply that higher incentives would increase the frequency and severity of bad decisions.

To formalize the behavioral part of the Undue Inducement Hypothesis, I develop a model that shows how participation incentives affect the acquisition of costly information by rational Bayesians. A form of motivated reasoning emerges as its main feature: As the incentive for participation increases, rational agents will consult information sources that are more likely to increase their optimism about the transaction. This result reflects a simple cost-benefit calculation. If the incentive for participating in a transaction is low, a prospective participant has little to gain from participating, but potentially much to lose. Accordingly, he seeks information that leads to a low probability of participating by mistake. With a high incentive, by contrast, the agent has more to gain from the transaction. Therefore, the agent is more cautious about mistaken rejection, and thus has a higher demand for encouraging information.

This model implies that upon observing that incentives skew information choice in a direction that appears designed to justify a particular decision, one cannot leap to the conclusion that undue inducement is present, or that the finding justifies a cap on incentives; it might just be a reflection of rational behavior. Instead, addressing the normative part of the Undue Inducement Hypothesis requires direct empirical measurements of welfare consequences (using the methods of behavioral welfare economics). Crucially, the formal welfare framework must allow the possibility that individuals make mistakes. Following the Undue Inducement literature, I focus on individuals who are incentivized to participate in the transaction. Because that literature is especially concerned about individuals who may be harmed by a transaction, I consider welfare functions that vary by the weight α they place on these individuals and weight $(1 - \alpha)$ on those who benefit. Taking the effects of incentives on individual subjects' well-being as given, I ask whether there exists a welfare weight α that prescribes allowing the transaction but capping incentives. I show that the answer to this question boils down to an inequality on observable parameters that I call the *Undue Inducement Condition*. It consists of two terms. First, at low incentives, the transaction is admissible if the benefits to those who gain from the transaction outweigh the losses of those who are harmed. This requirement places an upper bound on α —if potential harm was all that counted, the transaction should be prevented altogether, no matter the incentive. Second, the transaction becomes inadmissible at high incentives if a rising incentive magnifies harms more quickly than it increases the gains of those who benefit (e.g. due inducing worse decision making). This requirement places a lower bound on α —unless sufficient weight is attached to harm, the increasing gains of those who benefit will outweigh the losses of those who are harmed. The Undue Inducement Condition simply requires that the upper bound just defined exceeds the lower bound. If it is satisfied, there exist welfare weights α such that the transaction should be permitted at capped incentives. If it is violated, then the transaction should be allowed for sufficiently generous remuneration (if α is small) or should be prevented altogether (if α is large), but it is suboptimal to permit the transaction at capped incentives. These results apply both in an ex-ante welfare framework (in which the welfare function is defined on expected utilities) and in an ex-post welfare framework (in which the welfare function is defined on realized outcomes).

Relying on this conceptual framework, I test the Undue Inducement Hypothesis empirically. I conduct two experiments that complement each other. Experiment 1 permits a one-to-one connection to the theoretical model, and it allows me to test the normative claims both within the ex-ante and ex-post welfare frameworks. By relying on the induced preferences paradigm (Smith, 1976), however, it excludes a potentially important mechanism through which incentives might cause harm. Participants in real-world transactions subject to

undue inducement concerns must typically evaluate non-monetary consequences such as potential health risks in terms of dollars and cents. It is conceivable that incentives distort this process. Hence, Experiment 2 offers incentives for participating in a highly visceral transaction. I employ laboratory experiments rather than field data because of the tight control and precise measurement they afford, and because they offer the possibility to explore the robustness of the findings across a range of treatment conditions. Moreover, legal constraints on incentives and a lack of record keeping (Bodri, 2013) both make suitable field data hard to identify. Significantly, regulations based on undue inducement concerns apply to college students (who are a frequent target demographic for the recruitment of both human egg donors (Beeson and Lippman, 2006) and clinical trial participants) and to the magnitudes of the incentives used in the current experiment (a common concern of U.S. research ethics boards with social science studies).

In Experiment 1, subjects decide whether to accept a known, fixed payment in exchange for participating in a transaction. The transaction consists of a gamble in which subjects may incur a net loss with a known prior probability (the bad state) or not (the good state). Before committing to a decision, subjects choose between a “bold advisor” and a “cautious advisor” to obtain stochastic information about whether the loss will materialize. The bold advisor more likely recommends participation than the cautious advisor. Subjects know that the bold advisor always recommends participation in the good state, but half the time, he also recommends participation if the state is bad. Symmetrically, the cautious advisor always recommends abstention in the bad state, but half the time, he also recommends abstention if the state is good. After observing the chosen advisor’s recommendation, subjects decide whether to participate in gamble in exchange for the promised incentive. I find that higher incentives substantially increase subjects’ preference for the bold advisor, and that this information biasing causally affects subjects’ participation decisions.⁴ These results resonate with the idea that “payments lead donors to discount risks” (American Society for Reproductive Medicine, 2007); they are consistent with the behavioral part of the Undue Inducement Hypothesis. Yet, they are also consistent with Bayes-rational information acquisition and thus do not constitute proof for harmful incentives.

To test the normative part of the Undue Inducement Hypothesis, I elicit certainty equivalents for various gambles in a second part of the experiment. Unbeknownst to the subjects, each elicitation corresponds to a decision they faced when deciding whether to participate in the transaction after receiving a recommendation from the chosen advisor. Crucially, while subjects’ original participation decisions are based on their own, potentially mistaken beliefs about the success probability, the second part explicitly states success probabilities. They correspond to the posteriors a Bayesian would have held after receiving the chosen advisor’s recommendation. If a subject reveals a negative certainty equivalent, her participation decision in part one was mistaken (in the ex ante sense). The magnitude of the certainty equivalent reveals the welfare cost of the mistake, measured according to her own preferences. As such, this design is an example of the ‘reframed decisions paradigm’ common in behavioral public economics (based on Bernheim and Rangel (2009), and applied in Chetty et al. (2009); Allcott and Taubinsky (2015); Taubinsky and Rees-Jones (2018); Ambuehl et al. (2021); see Bernheim and Taubinsky (2018) for a review). In the language of that literature, the

⁴Following the literature on media bias (e.g. Gentzkow et al., 2015), I use the term *bias* to indicate divergences between the frequency of informative signals and the frequency of underlying states of the world. This use differs from the behavioral economics literature which reserves the term *bias* for irrational or suboptimal behaviors.

participation decisions in part 1, which may be afflicted by information processing errors, are subject to ‘characterization failure’ whereas the reframed decisions in part 2 exclude that possibility. Hence, part 2 decisions constitute the ‘welfare relevant domain’ that serves to benchmark the quality of choice in part 1. Turning to results, I find that the Undue Inducement Condition is robustly violated, which means that there are no welfare weights on benefit and harm that justify capping incentives. It is violated in treatments in which immediate rewards are paid for accepting delayed losses, in treatments that vary the prior success probability, and in treatments in which the incentive amount is lost if the state turns out to be bad. The Undue Inducement Condition is robustly violated even if I offer the planner more scope for disagreeing with individual choices by using ex-post realized outcomes rather than ex-ante certainty equivalents as measures of individual welfare.⁵ While this finding would be a foregone conclusion in a sample of Bayes-rational individuals (who never incur ex-ante losses), some of my treatments induce frequent ex-ante welfare losses in actual subjects’ participation decisions (but incentives do not make them worse). Overall, this experiment shows that undue inducement concerns are misplaced—at least if there is no need to weigh monetary benefits against non-monetary costs.

Experiment 2 tests the Undue Inducement Hypothesis when prospective participants do need to evaluate non-monetary consequences in terms of dollars and cents. To give the Undue Inducement Hypothesis the best chance, I seek a transaction that is visceral and aversive, that is unfamiliar to subjects, and that allows for biased information acquisition. Hence, I incentivize subjects for eating whole insects such as mealworms, silkworm pupae, or mole crickets in exchange for money. Specifically, subjects learn whether they will receive \$3 or \$30 in exchange for eating bugs. Like in the first experiment, they acquire information before making a participation decision. Subjects between two videos titled, respectively, “Why you may want to eat insects” and “Why you may not want to eat insects.” A subject who would like to receive the promised \$30 incentive and thus aims to persuade herself that eating insects is a less miserable experience than she would otherwise have thought, for instance, can do so by selecting the encouraging video. Upon watching the chosen video, subjects decide whether or not to participate in the transaction in exchange for the promised incentive. Consistent with the behavioral part of the Undue Inducement Hypothesis, I find that higher incentives increase the demand for the encouraging video and decrease the demand for the discouraging video. Moreover, the information provided causally affects subjects’ participation decisions; subjects do not just use it to make it easier to follow through with the choice they would have made anyway. As in Experiment 1, the consistency of these results with the behavioral part of the Undue Inducement Hypothesis does not prove harm. Testing the normative part of the hypothesis requires direct welfare measurement.

In this experiment, I measure welfare using the other dominant paradigm of behavioral public economics, namely the ‘informed consumer benchmark’ (Bronnenberg et al. (2015); Allcott et al. (2019); Allcott and Kessler (2019); see Bernheim and Taubinsky (2018) for a review). In this experiment, subjects make the incentivized participation decisions based only on the selected video and a brief verbal description of the insects. After making these choices, each subject receives several containers containing insects of each species. I force subjects to visually inspect, touch, and (inadvertently) smell each of the insects. Based on

⁵A supplementary experiment, presented in Appendix D, shows that the welfare conclusions of Experiment 1 replicate in a dynamic information acquisition context that allows for a large set of possible information structures.

this substantial additional information, subjects reveal reservation prices for ingesting the specimens in front of them. By subtracting these reservation prices from the incentive amount a subject was promised in the first part of the experiment, I obtain a measure of the Marshallian surplus associated with her incentivized participation decision. (If she rejected the transaction, her surplus is zero.) A negative surplus reveals that the subject’s decision to participate was mistaken (in the ex-post sense); its magnitude equals the ex-post welfare cost. In this experiment, too, I find that the Undue Inducement Condition is robustly violated. No welfare weights exist that justify permitting the transaction but capping incentives. This result is robust to the following two concerns. First, the reservation prices used for welfare evaluation might be biased by the incentive amount a subject was promised (for instance through anchoring, or by making it hard for subjects to think clearly about monetary equivalents). I address this possibility by assigning the distribution of reservation prices in the low incentive condition to subjects in the high incentive condition, which yields qualitatively similar results. Second, I account for the possibility of bias caused by noisy elicitation, which does not alter the results either. Overall, in this experiment, undue inducement concerns are misplaced even though subjects need to weigh monetary benefits against non-monetary costs.

As is the case with any empirical study, these results require careful interpretation. The Undue Inducement Hypothesis claims that, *as a general matter*, higher incentives can lead to worse decision making in a way that justifies capping incentives. My experiments show that this assertion is not generally true: even though they generate the behavioral effects the Undue Inducement Hypothesis predicts, in my experimental data, there are no welfare weights that justify permitting the transaction but capping incentives. A WAVE1 study in the sense of [List \(2020\)](#), this paper rests on laboratory experiments with small stakes. Hence, it implies directly that the concerns that U.S. ethical review boards frequently express about incentives in typical social science experiments are misplaced. It cannot, however, rule out the possibility that incentives cause harm in contexts with much larger stakes, possibly through mechanisms other than those studied here.⁶ It is also important to contextualize the normative analysis performed in this study. I do not consider non-welfarist considerations such as concerns with autonomy. Moreover, following the approach in the undue inducement literature, I only consider the welfare of individuals who are being recruited into the transaction. I abstract from beneficiaries on the demand side in real-world transactions such as kidney recipients or societal benefits from medical progress (the theoretical framework briefly addresses this point).

At the same time, it is important to distinguish the Undue Inducement Hypothesis from two related intuitions that arise in the context of incentive payments. First, some commenters take the stance that incentives are problematic because they would disproportionately attract the poor. Concerns about inequality are logically distinct from the Undue Inducement Hypothesis. The latter deems high incentives problematic *per se*; it applies even to a hypothetical world *without* any inequality. This paper addresses the Undue Inducement Hypothesis; it has nothing to say about distributional issues (but see [Ambuehl et al., 2020b](#)). Second, people sometimes intuit that monetary incentives would decrease the supply of scarce goods through motivational crowding out ([Titmuss, 1971](#); [Frey and Jegen, 2001](#)). Concerns about supplied quantities, how-

⁶Candidate mechanisms include choking under pressure, as studied in [Ariely et al. \(2009\)](#), or social comparisons, as in [Ager et al. \(2019\)](#).

ever, are distinct from concerns about the welfare of the incentivized. The Undue Inducement Hypothesis only addresses the latter.⁷

Next to informing the debate on undue inducement, this paper contributes to two strands of academic literature. First, it adds to the burgeoning literature on *repugnant transactions* (Kahneman, Knetsch and Thaler, 1986; Basu, 2003, 2007; Roth, 2007; Leider and Roth, 2010; Niederle and Roth, 2014; Ambuehl, Niederle and Roth, 2015; Elias, Lacetera and Macis, 2015a,b; Clemens, 2018; Ambuehl and Ockenfels, 2017; Elías, Lacetera and Macis, 2019; Stueber, 2020; Sullivan, 2020; Erkut and Kübler, 2021). While that literature typically studies third parties’ judgments, the current paper focuses on the effect of incentive payments on the incentivized. Second, by conducting welfare analysis of incentive-induced information biasing, this paper contributes to a literature on the positive test strategy (Slowiaczek et al., 1992) and on motivated reasoning (see Klayman and Ha, 1987; Kunda, 1990; Klayman, 1995; Bénabou, 2015; Epley and Gilovich, 2016; Gino, Norton and Weber, 2016; Bénabou and Tirole, 2016, for reviews).⁸

The remainder of this paper proceeds as follows. Section 2 presents the conceptual foundations. Sections 3 and 4 present Experiments 1 and 2, respectively. Finally, Section 5 concludes.

2 Conceptual foundations

In this section, I first present a simple model that addresses the behavioral part of the Undue Inducement Hypothesis. It asks how a rational Bayesian will acquire information about an incentivized transaction when information acquisition is costly (subsection 2.1).⁹ It shows that incentive-induced information biasing is consistent with Bayesian rationality, and hence, by itself, is no indication of harmful effects of incentives. Subsection 2.2 addresses the normative part of the Undue Inducement Hypothesis. It presents the behavioral welfare economic framework my experiments use to detect potential harm from incentives. All proofs are in Appendix A.

2.1 Behavior

2.1.1 Setting

A Bayesian expected utility maximizer decides whether or not to participate in a transaction in exchange for a material incentive m . The consequences of participation depend on a state of the world $s \in \{G, B\}$, which is good ($s = G$) with prior probability μ , and bad ($s = B$) otherwise. If the agent participates and the state is good, he experiences utility $u_G(m) > 0$. If the agent participates and the state is bad, he experiences disutility $u_B(m) < 0$. If the agent abstains, he does not receive the payment m and experiences utility 0. Hence, the agent would like to participate in the good state, and abstain in the bad state. I assume that $u_G(m)$ is strictly increasing in m , and that $u_B(m)$ is weakly increasing in m , reflecting the fact that higher

⁷Recent research does not find motivational crowding out in the case of transactions subject to undue-inducement based regulations (see, e.g., Tishler and Bartholomae, 2002; Siddique et al., 2020).

⁸Recent experimental tests of closely related mechanisms in economics include Charness et al. (forthcoming) and Montanari and Nunnari (2019). These papers focus on confirmatory bias and thus examine the effects of variation in prior beliefs rather than of variation in payouts.

⁹The model is closely related to Suen (2004).

compensation increases a participants' utility in each stage. The functions $u_G(m)$ and $u_B(m)$ incorporate the agent's risk preferences.

Before deciding whether to participate in the transaction, the agent chooses an information structure which produces a stochastic signal about the state. Critically, I assume that the agent's information processing capacity is limited, so he needs to be selective about the kind of information to consult. Formally, the agent chooses an information structure from a finite set $\mathcal{I} = \{I_1, \dots, I_n\}$ with $n \geq 2$ which may represent, for instance, advisors with various biases. Each information structure I is a function $I : \{G, B\} \rightarrow \Delta\mathcal{S}$ that maps states of the world into probability distributions over a space of signals $\mathcal{S}$. For simplicity, I assume that each information structures' space of signals has two elements, $\mathcal{S} = \{\mathcal{G}, \mathcal{B}\}$ which I interpret as a recommendation to participate and a recommendation to abstain, respectively. I let p_I^s denote the probability that information structure I produces signal $\mathcal{G}$ if the state is s . Moreover, I let $\gamma_{\sigma, I}$ denote the agent's posterior belief that $s = G$ upon observing signal σ from information structure I . Without loss of generality, I assume $p_I^G \geq p_I^B$ for each information structure $I \in \mathcal{I}$. This assumption implies that $\gamma_{\mathcal{G}, I} > \gamma_{\mathcal{B}, I}$ for every information structure $I \in \mathcal{I}$.¹⁰

I will show that participation incentives change the kind of information a rational agent demands, specifically, the optimal information structure's *bias*, as defined in [Gentzkow, Shapiro and Stone \(2015\)](#). In my setting, that definition specializes as follows:¹¹

Definition 1. Consider two information structures I and I' and a given prior μ . Then, I' is *biased towards* $\mathcal{G}$ relative to I if $P(I' = \mathcal{G}) \geq P(I = \mathcal{G})$. Moreover, I' is *statewise biased towards* $\mathcal{G}$ relative to I if $p_{I'}^s \geq p_I^s$ for all states $s \in \{G, B\}$.

In words, I' is *biased towards* $\mathcal{G}$ relative to I if it has a higher ex ante probability of producing signal realization $\mathcal{G}$. The definition of statewise bias requires that probability to be higher in each state. That definition (not contained in [Gentzkow et al., 2015](#)) helps me derive stronger results.

2.1.2 Analysis

The agent's optimization problem involves two stages. First, the agent selects an information structure. Second, he decides whether to participate based on the signal realization he has observed. The selected information structure is *instrumental* if, in the second stage, the agent chooses to participate after observing signal realization $\mathcal{G}$ and chooses to abstain after observing signal realization $\mathcal{B}$. The agent also has the opportunity to participate regardless of the signal realization, or to abstain regardless of the signal realization.

¹⁰The assumption is without loss of generality because one can always assign label $\mathcal{G}$ to the signal realization that is associated with the higher posterior belief.

¹¹The following argument shows the equivalence to the definition stated in [Gentzkow et al. \(2015\)](#). Consider two information structures I and I' . Let $\lambda(I|I')$ be the distribution of an observers' posterior probability that $s = G$ if he believes that the report is generated by information structure I even though the report is actually generated by information structure I' . [Gentzkow et al. \(2015\)](#) define I' as *biased towards* G relative to I if $\lambda(I|I')$ first-order stochastically dominates $\lambda(I|I)$. An agent who believes the signal is generated by information structure I will have a posterior belief that $s = G$ of $\gamma_{\mathcal{G}, I} = \frac{\mu p_G}{\mu p_G + (1-\mu)p_B}$ if he observes $\mathcal{G}$ and $\gamma_{\mathcal{B}, I} = \frac{\mu(1-p_G)}{\mu(1-p_G) + (1-\mu)(1-p_B)}$ if he observes $\mathcal{B}$. If information structure I actually generates these signals, then the agent's posterior will equal $\gamma_{\mathcal{G}}$ with probability $p = \mu p_G + (1-\mu)p_B$ and $\gamma_{\mathcal{B}}$ otherwise. If the signal is actually generated by information structure I' , then the agent's posterior will equal $\gamma_{\mathcal{G}}$ with probability $p' = \mu p'_G + (1-\mu)p'_B$ and $\gamma_{\mathcal{B}}$ otherwise. Because the support of the distribution of the agent's posterior probability is the same across these two cases, the required first order stochastic dominance relationship obtains if and only if p' exceeds p . Accordingly, I' is biased towards $\mathcal{G}$ relative to I if and only if $p' \geq p$.

To easily account for this possibility in the first stage, I define the information structure $I^G = (1, 1)$ which produces signal $\mathcal{G}$ regardless of the state, and information structure $I^B = (0, 0)$ which always produces signal $\mathcal{B}$. I let $\bar{\mathcal{J}} = \mathcal{J} \cup \{I^G, I^B\}$ and I study the agent's choice from $\bar{\mathcal{J}}$. Further, I assume, without loss of generality, that the agent follows the signal produced by the chosen information structure. Hence, the agent's probability of participating in state s equals the probability of a good signal in that state, p_I^s . Because the agent's utility upon non-participation is 0, the agent's utility maximization problem in the first stage takes the following form:

$$\max_{I \in \bar{\mathcal{J}}} \mu p_I^G u_G(m) + (1 - \mu) p_I^B u_B(m) \quad (1)$$

Optimal information choice. The main result of this section concerns the comparative statics of the optimal solution to (1) with respect to the incentive m .¹² It shows that Bayes-optimal behavior appears as if the agent were engaged in *motivated reasoning*: as the participation incentive m increases, the agent's optimal information structure is more strongly biased towards $\mathcal{G}$.

Proposition 1. *Consider incentive amounts m and m' with $m > m'$. Suppose at least one information structure is instrumental at m , and at least one information structure is instrumental at m' . Let I^* and I'^* denote the solutions to (1) at m and m' , respectively, and suppose that $I^* \neq I'^*$. Then I^* is statewise more biased towards $\mathcal{G}$ than I'^* .*

Intuitively, proposition 1 arises from a simple cost-benefit calculation. Information helps the agent limit choices that are *ex-post mistaken* in the sense that the chosen action is suboptimal conditional on the state of the world (even though the choice may have been optimal conditional on the information available at the time of decision). There are two possible ex-post mistakes. A *false positive* occurs if the agent participates in the bad state. A *false negative* occurs if the agent abstains in the good state. The agent seeks to minimize the cost of these mistakes, which depend on the incentive amount m . A higher incentive decreases the cost of false positives (because the additional money acts as insurance against ex-post undesirable outcomes) and increases the cost of false negatives (because abstention means foregoing a larger amount of money). Accordingly, as the incentive increases, the agent becomes more interested in preventing false negatives and less interested in preventing false positives. The agent achieves this goal by selecting an information structure that is more strongly biased towards $\mathcal{G}$.¹³

Figure 1 illustrates Proposition 1 graphically. It plots some set of information structures $\mathcal{J}$ in $(p_G, 1 - p_B)$ -space. These information structures, together with unconditional participation and abstention information structures I^G and I^B , constitute the agent's choice set $\bar{\mathcal{J}}$. The agent's utility increases to the northeast. By

¹²The solution to (1) is not necessarily unique. Generally, the solution to (1) is increasing in m in the strong set order.

¹³A related intuition views the problem through the lens of posterior probabilities. If the incentive is small, the agent has little to gain from participation, but potentially much to lose. Hence, he will participate in the transaction only if the posterior belief that the state is good is close to 1. An agent's posterior can be high only if the information structure rarely ever produces signal $\mathcal{G}$ in the bad state. Such an information structure's unconditional probability of producing signal $\mathcal{G}$ must be low. By contrast, if the incentive is high, such that the agent has much gain from the transaction and little to lose, he will abstain only if his posterior that the state is good is sufficiently low. Such a low posterior can obtain only if the information structure rarely produces signal $\mathcal{B}$ in the good state. Such an information structure's unconditional probability of producing signal $\mathcal{G}$ must thus be high. Overall, therefore, a higher incentive amount causes the agent to opt for information structures that are more likely to produce signal $\mathcal{G}$.

Figure 1: Bayesian information choice

expected utility, the agent's indifference curves are straight lines in $(p_G, 1 - p_B)$ -space. Therefore, it suffices to consider the convex hull of the agent's opportunity set, shaded in gray in the figure.¹⁴ A higher incentive amount m increases the benefit of participation in the good state, p_G , and (weakly) decreases the benefit of abstention in the bad state, $1 - p_B$. Accordingly, an increase in m tilts indifference curves counterclockwise. Therefore, the optimal information structure associated with a higher incentive lies further to the northwest. In the figure, the optimal information structure for a given low incentive amount is labeled A , and the optimal information structure for a given high incentive amount is labeled B . The remaining information structures lie in the interior of the convex hull of the agent's opportunity set and hence will not be chosen.

Supply of participants Under the conditions of Proposition 1, the change in information acquisition induced by higher incentives affects the probability the agent will participate in the transaction. Under additional conditions, endogenous information acquisition amplifies the effect of incentives in the following sense. If the agent can freely choose an information structure, a given increase in the incentive amount will cause a greater change in the agent's participation probability than if the most strongly biased information structures are removed from the agents' choice set.

¹⁴Strictly speaking, the grey area is the *free-disposal* convex hull as it also includes any point to the lower left of the convex hull.

Formally, consider two incentive amounts m and m' with $m > m'$. Construct a restricted set $\mathcal{J}_r$ of one or more information structures by removing the k_1 most participation-biased information structures $\mathcal{J}$, as well as the k_2 least biased information structures.¹⁵ Let $p(m)$ denote the unconditional participation probability at incentive m , and let $p_r(m)$ denote the corresponding variable if the subject's opportunity set is restricted to $\mathcal{J}_r$. Finally, let $I^*(m)$ and $I^*(m')$ the agent's optimal choice from the unrestricted set, $\mathcal{J}$. With this notation, we can formally state the result that information biasing can amplify the effect of incentives.¹⁶

Proposition 2. *Consider incentive amounts m and m' with $m > m'$. Suppose that for both m and m' there is at least one information structure in $\mathcal{J}_r$ that is instrumental at the respective incentive. Then,*

$$p(m) - p(m') \geq p_r(m) - p_r(m').$$

The inequality is strict if at least one of the unrestricted optimal information structures, $I^(m)$ or $I^*(m')$, is not contained in $\mathcal{J}_r$.*

The experiment in Section 3 will use only two information structures, due to design considerations explained there. In order to test whether endogenous information acquisition amplifies the effect of incentives in that setting, subjects do not choose from a restricted set of information structures, but instead subjects select one of the two available information structure before knowing the incentive amount. To state the expected amplification effect formally, let $q(m)$ denote the unconditional participation probability in that treatment if the incentive amount m realizes.

Corollary 1. *Suppose $|\mathcal{J}| = 2$. Consider incentive amounts m and m' with $m > m'$. Let I^* and I'^* denote the information structures that solve (1) at m and m' , respectively, and suppose that $I^* \neq I'^*$. Further, suppose that both I^* and I'^* are instrumental at both m and m' . Then*

$$p(m) - p(m') > q(m) - q(m').$$

Overall, this model demonstrates that even a perfectly rational Bayesian will exhibit the types of behaviors that the undue inducement literature has interpreted as evidence that incentives cause defective decision making. Accordingly, evidence of the effect of incentives on information search and participation alone cannot establish whether incentives cause harm. This task requires direct welfare measurement.

2.2 Welfare

The above analysis focusses on rational individuals. Clearly, offering incentives to a Bayes-rational individual cannot make that person worse off: a Bayes-rational individual will correctly assess the expected utility

¹⁵Because the bias of an information structure in this setting is simply the unconditional probability of producing signal $\mathcal{G}$, all information structures can be ordered. No similar statement holds for the stronger notion of statewise bias.

¹⁶Observe that proposition 2 uses stronger conditions than proposition 1. In the case of proposition 2, any information structure that is optimal for one incentive must be instrumental for the other incentive. If this condition is violated, the conclusion of the proposition may reverse.

from participation, and will simply refuse to participate if it would make her worse off.¹⁷ If individuals are not Bayesian, however, it is possible, at least in principle, that incentives distort beliefs about the transaction or the quality of choice to such an extent that they harm prospective participants. For instance, subjects may bias information choice akin to Bayesians, but may forget to sufficiently discount positive signals from participation-biased information structures, thus becoming more overconfident with higher incentives. Therefore, I now derive a criterion that indicates whether subjects’ empirical choices admit a welfare function under which low but not high incentives are admissible.

The fact that the transactions I consider involve uncertainty gives rise to two welfare concepts, *ex-ante* and *ex-post* welfare (see [Hammond, 1981](#); [Sandmo, 1983](#)). *Ex-ante* welfare functions are defined on subjects’ expected utilities (or, more generally, on the value of the subject’s objective function at the point of making the decision). *Ex-post* welfare functions, by contrast, take subjects’ outcomes (after the realization of uncertainty) as arguments. *Ex-ante* welfare analysis may deem an agents’ decision as mistaken only if the agent’s decision is based on a faulty understanding of her opportunity set. It is a common approach in choice-based behavioral welfare economics (see [Bernheim and Taubinsky, 2018](#)). By contrast, *ex-post* welfare analysis also allows the planner to take issue with agents’ decisions because her attitudes to inequality may differ from the agents’ attitudes to risk, or because she disagrees with the agents’ risk attitudes for other reasons. The undue inducement literature is typically reluctant to defer to individuals’ risk preferences, and hence more closely resembles *ex-post* analysis than *ex-ante* analysis. It relies, “for example, [on] rules expressing how ordinary or average or prudent persons would view the offer” ([Faden and Beauchamp, 1986](#)).¹⁸

I will conduct both *ex-post* and *ex-ante* analysis (wherever possible). In either case, I consider a planner who aggregates welfare according to

$$W = \sum_i f(w_i).$$

In the case of *ex-post* welfare, w_i denotes the subject’s realized outcome. In the case of *ex-ante* welfare, w_i denotes the subjects’ certainty equivalent from her chosen action. The certainty equivalent is measured conditionally on the information the subject has observed before making a participation decision, based on Bayesian posteriors. It will be negative if a subject participates by mistake, for instance because her participation decision was based on biased posterior beliefs.

The undue inducement literature places much weight on subjects who might be harmed by participating in a transaction. Hence, I consider social welfare functions that place weight α on individuals who are harmed from being offered the opportunity to participate, and weight $(1 - \alpha)$ on those who benefit, for $0 \leq \alpha \leq 1$.

¹⁷Notably, preventing a transaction cannot improve the welfare of a Bayes-rational individual even if she seeks information in an irrational or haphazard manner, as long as she updates her beliefs rationally conditional on the information obtained, and chooses the (expected) utility maximizing action based on those beliefs.

¹⁸For instance, transactions are frequently deemed unacceptably risky even if large numbers of individuals would like to participate. The regulation of human challenge trials for the development of a Covid19-vaccine is a recent example ([Ambuehl et al., 2020a](#)). Similar examples exist in other regulatory environments. The ‘Prudent Person Rule’ in pension finance, for example, requires financial investments to be made in the way “men of prudence, discretion and intelligence manage their own affairs” ([Harvard College v Amory \(1830\)](#); see [Goldman \(2000\)](#) for a recent account).

Formally, f is given by

$$f(w) = \begin{cases} \alpha w & \text{if } w < 0 \\ (1 - \alpha)w & \text{if } w \geq 0. \end{cases} \quad (2)$$

If $\alpha = \frac{1}{2}$, the welfare function places equal weight on each individual, and hence is utilitarian. Throughout, I will call a transaction *admissible* at incentive m if the welfare from permitting the transaction at that incentive, $W(m)$, exceeds the welfare from outlawing the transaction.¹⁹

In order to model the effect of participation payments on admissibility in a way that permits parsimonious empirical estimation, I assume that individual welfare gains and losses relate linearly to the incentive amount, m .²⁰ While linearity is a restrictive condition, it greatly aids the transparency of the analysis, and it is approximately satisfied in my experiments (see Figure 5). Specifically, letting $g_i = \max(0, w_i)$ and $l_i = \min(0, w_i)$, I consider the following functional form:²¹

$$\begin{aligned} g_i(m) &= \beta_0^G + \beta_1^G m + \epsilon_i^G(m) \\ l_i(m) &= \beta_0^L + \beta_1^L m + \epsilon_i^L(m) \end{aligned} \quad (3)$$

Here, $\epsilon_i^G(m)$ and $\epsilon_i^L(m)$ are independent individual effects with expectation zero such that $g_i(m) \geq 0 \geq l_i(m)$ for all i . Observe that $g_i(m)$ and $l_i(m)$ are sufficient statistics for the class of welfare functions I consider (Chetty, 2009).

If the transaction is offered at incentive m , then a planner who uses weight α evaluates societal welfare (2) as

$$\begin{aligned} E(W) &= (1 - \alpha)E(g_i(m)) + \alpha E(l_i(m)) \\ &= \beta_0(\alpha) + \beta_1(\alpha)m, \end{aligned}$$

where $\beta_0(\alpha) = (1 - \alpha)\beta_0^G + \alpha\beta_0^L$, and $\beta_1(\alpha) = (1 - \alpha)\beta_1^G + \alpha\beta_1^L$.

We can now express the normative part of the Undue Inducement Hypothesis by stating two conditions that must hold simultaneously. First, the transaction is admissible at a low incentive if $\beta_0(\alpha) > 0$. Intuitively, the planner must not place excessive weight on those who lose from the transaction, for otherwise the transaction is inadmissible even at low incentives. Second, the transaction eventually becomes inadmissible as incentives rise if $\beta_1(\alpha) < 0$. Intuitively, if higher incentives increase both the gains of those who benefit and the losses of those who are harmed, the planner must place sufficiently much weight on the latter for the transaction to eventually become inadmissible. Defining $\bar{\alpha} = \frac{\beta_0^G}{\beta_0^G - \beta_0^L}$ and $\underline{\alpha} := \frac{\beta_1^G}{\beta_1^G - \beta_1^L}$, we can write the

¹⁹This analysis only considers welfare on the supply side. I discuss an extension to account for welfare on the demand side at the end of this section.

²⁰The linearity assumption excludes the possibility of a discontinuity or kink that might justify placing restrictions at incentives of that magnitude, as suggested by pertinent regulation (e.g. [US Department of Health, Education, and Welfare](#), DHEW). The vast majority of ethics professionals in the survey by [Largent et al. \(2012\)](#), however, do not feel they can judge whether an incentive amount is ‘excessive,’ calling the existence of such a discontinuity or kink into question.

²¹All the following results generalize to the case in which m is replaced by some monotonic and unbounded function $\varphi(m)$.

first condition as $\alpha < \bar{\alpha}$, and the second as $\alpha > \underline{\alpha}$.²² Combining the two conditions, we see that the undue inducement concern is justified only if the upper bound $\bar{\alpha}$ exceeds the lower bound, $\underline{\alpha}$. Otherwise, there exists no welfare weight $\alpha \in [0, 1]$ such that low but not high incentives are admissible. I call the condition $\bar{\alpha} > \underline{\alpha}$ the *Undue Inducement Condition* and state it formally in the following proposition.

Proposition 3. *There exists a welfare weight $\alpha \in [0, 1]$ under which low but not high incentives are admissible if and only if*

$$\frac{\beta_0^G}{\beta_0^G - \beta_0^L} - \frac{\beta_1^G}{\beta_1^G - \beta_1^L} > 0. \quad (4)$$

By estimating the parameters $\beta_0^G, \beta_0^B, \beta_1^G, \beta_1^B$ from experimental data I will test whether the effects of incentives on subjects' behavior are consistent with condition (4). To aid intuition, observe that given such data, the Undue Inducement Condition (4) can easily be checked visually. First, consider Panel A of Figure 2. To check the condition, extend both functions $g(m)$ and $l(m)$ to the left until they intersect. The condition is satisfied if the intersection point, denoted A , lies above the horizontal axis; otherwise it is violated.²³ To see why, consider the case of $\alpha = \frac{1}{2}$ (utilitarian welfare) indicated by the dashed line. It hits the horizontal axis from below, meaning that the transaction is inadmissible for incentives below the intersection point and admissible at incentives higher than that. Graphically, a change in the welfare parameter α pivots the dashed line around point A . Therefore, for any α , the social welfare curve will hit the horizontal axis from below, or not at all. It is not possible to make the social welfare function hit the horizontal from above—there is no welfare parameter α such that the transaction is admissible at low but not high incentives.

Panel B illustrates the case in which the Undue Inducement Condition is satisfied. Because the intersection of $g(m)$ and $l(m)$ lies above the horizontal axis, there are values of α such that the social welfare function hits the horizontal axis from above, so one should permit the transaction but cap incentives.

Connecting the behavioral and normative parts of the Undue Inducement Hypothesis How does information biasing (which formalizes the behavioral part of the Undue Inducement Hypothesis) relate to the Undue Inducement Condition (which formalizes the normative part of the Undue Inducement Hypothesis)? The following example shows how information biasing may facilitate undue inducement.

Using the setting and notation of subsection 2.1, consider $\mathcal{I} = \{I^G, I^B\}$ with $I^G = (1, 0.5)$ and $I^B = (0.5, 0)$ (the parameters I will use in Experiment 1). Hence, the participation-biased information structure I^G only recommends abstention if the state is bad, and the abstention-biased information structure I^B only recommends participation if the state is good. Consider two incentive levels, m_0 and m_1 with $0 < m_0 < m_1$. Further, suppose the agent chooses I^B in case of incentive m_0 , chooses I^G in case of incentive m_1 , and always follows the signal produced by the chosen information structure.

In the case of the low incentive, $m = m_0$, the agent never participates in the bad state, but sometimes participates in the good state. Accordingly, mean gains are positive, $E(g_i(m_0)) > 0$, and mean welfare losses

²²This transformation makes use of $\beta_0^G > \beta_0^L$ which is true by definition, and of $\beta_1^G > \beta_1^L$ which I will test empirically.

²³To prove this statement formally, define x by $\beta_0^G + \beta_1^G x = \beta_0^L + \beta_1^L x$ as the intersection of $g(m)$ and $l(m)$. Solving for x , inserting the resulting expression into $\beta_0^G + \beta_1^G x > 0$, and rearranging yields condition (4).

Figure 2: Welfare weights and the possibility of harmful incentives

A. Condition (4) is not satisfied: the intersection of $E(g_m)$ and $E(l_m)$ (labeled A) occurs at a negative value of W . Hence, there are no welfare weights α such that low but not high incentives are admissible. Graphically, changing the weight α pivots the dashed line around point A. Therefore, societal welfare $W(m)$ will always intersect the horizontal axis from below, if at all.

B. Condition (4) is satisfied: the intersection of $E(g_m)$ and $E(l_m)$ (labeled A) occurs at a positive value of W . Hence, there exists an interval $[\underline{\alpha}, \bar{\alpha}]$ such that for all $\alpha \in [\underline{\alpha}, \bar{\alpha}]$ low but not high incentives are admissible. Graphically, changing the weight α pivots the dashed line around point A. Therefore, societal welfare $W(m)$ will always intersect the horizontal axis from above, if at all.

are zero, $E(l_i(m_0)) = 0$, both according to ex-ante and ex-post welfare interpretations. Accordingly, the first term of the Undue Inducement Condition (4) is given by $\frac{\tilde{\beta}_0^G}{\tilde{\beta}_0^G - \tilde{\beta}_0^L} = 1$. Here, $\tilde{\beta}_0^G$ and $\tilde{\beta}_0^L$ denote the level of gains and losses at incentive m_0 , respectively.²⁴

In the case of the high incentive, $m = m_1$, by contrast, there is a positive probability that the agent participates in the bad state and suffers an ex-post loss. In the case of ex-post welfare, this fact directly implies $E(l_i(m_1)) < 0$.²⁵ In the case of ex-ante welfare $E(l_i(m_1)) < 0$ obtains if the subject is non-Bayesian (making it possible to observe the assumed participation strategy) and sufficiently risk-averse (putting sufficient weight on the potential loss).²⁶ Accordingly, for the second term of condition (4), we have $\frac{\beta_1^G}{\beta_1^G - \beta_1^L} < 1$.

Together these expressions imply that the Undue Inducement Condition (4) is satisfied.

Demand-side benefits The undue inducement literature typically focuses on the supply side of transactions (participants who receive money for providing goods and services). Addressing the undue-inducement literature heads-on, this paper takes the same approach. How would welfare benefits on the demand side changes the analysis? For simplicity, assume that each participant on the supply side generates welfare benefits c for the demand side. Here, c incorporates the welfare weight placed on the demand side. Further, assume that incentives m cause participation by $S = \delta_0 + \delta_1 m$ individuals on the supply side, where $\delta_0 > 0$ and $\delta_1 > 0$. Total benefits to the demand side are thus given by $B(m) = c\delta_0 + c\delta_1 m$. A transaction is now admissible at incentive m if $W(m) + B(m) \geq 0$. The Undue Inducement Condition takes the following form:²⁷

$$\frac{\beta_0^G - c\delta_0}{\beta_0^G - \beta_0^L} - \frac{\beta_1^G - c\delta_1}{\beta_1^G - \beta_1^L} > 0. \quad (5)$$

In terms of Figure 2, $-B(m)$ is a decreasing linear function that intersects the vertical axis in the non-positive range. If the intersection of $G(m)$ and $L(m)$ lies above $-B(m)$, the augmented Undue Inducement Condition (5) is satisfied; otherwise it is violated. Trivially, accounting for demand side benefits tips the scales in favor of permitting the transaction at any given incentive and welfare weight α . Without imposing further assumptions on the parameters, however, it is not possible to draw general conclusions about whether (5) is easier or harder to satisfy than (4), that is, whether undue inducement is a more or less of a possibility once one accounts for demand side benefits.

²⁴Here, I evaluate levels at the incentive $m = m_0$ rather than at the intercept $m = 0$ for convenience. The evaluation of condition (4) can be performed using the level of the welfare gain and welfare loss functions at any incentive amount. To see this formally, notice that for any $x > 0$, $\frac{\beta_0^G + x\beta_1^G}{\beta_0^G + x\beta_1^G - \beta_0^L - x\beta_1^L} - \frac{\beta_1^G}{\beta_1^G - \beta_1^L} > 0$ if and only if $\frac{\beta_0^G}{\beta_0^G - \beta_0^L} - \frac{\beta_1^G}{\beta_1^G - \beta_1^L} > 0$, as straightforward algebraic manipulation shows.

²⁵Notably, this is true even if the agent is Bayesian and chooses information rationally, because ex-post welfare analysis permits the planner's weight on losses to differ from the agent's risk attitudes.

²⁶A Bayes-rational subject will never participate at an ex-ante loss, and unless the subject is sufficiently risk averse, participation following signal $\mathcal{G}$ at m_1 is not necessarily a mistake even for non-Bayesian subjects.

²⁷The transaction is admissible at low incentives if $(1 - \alpha)\beta_0^G + \alpha\beta_0^L \geq -\delta_0$. The transaction eventually becomes inadmissible at sufficiently high incentives if $(1 - \alpha)\beta_1^G + \alpha\beta_1^L < -\delta_1$. Rearranging and combining these inequalities in the same way as for the derivation of condition (4) yields condition (5).

3 Experiment 1: Choosing Between Advisors

In this section, I experimentally test both parts of the Undue Inducement Hypothesis using the induced preferences paradigm (Smith, 1976). Throughout, I make design decisions with the goal of maximizing the chance of finding harmful effects of incentives if they exist. I deploy a range of treatment conditions such as delayed consequences and asymmetric prior probabilities to examine the robustness of my findings.

3.1 Design

Overview The experiment consists of two parts, each of which proceeds in multiple rounds.

In each round of the first part, the subject decides whether to participate in a gamble that depends on a state of the world s . The state is good ($s = G$) with prior probability μ or bad ($s = B$) otherwise. If the subject participates in the gamble and the state is bad, she suffers a monetary loss L . If the state is good, she does not suffer a loss. The subject receives the incentive payment m if she participates; otherwise she receives nothing. Before deciding whether to participate, the subject chooses one of two information structures $\mathcal{I} = \{I^G, I^B\}$ from which she observes a stochastic signal about the state. Information structure I^G is statewise biased towards G relative to I^B . Decisions in this part provide all information required to test the behavioral part of the Undue Inducement Hypothesis (Proposition 1 and Corollary 1).

In the second part of the experiment I elicit the welfare benchmarks needed to test the normative part of the Undue Inducement Hypothesis (proposition 3) using ex-ante welfare analysis. In each round, the subject reveals her certainty equivalent for a lottery that leads to a gain $g \geq 0$ with probability p or to a loss $l < 0$ with probability $(1 - p)$. Unbeknownst to the subject, the parameters g, l , and p correspond to the participation decisions she faced in the first part of the experiment after having observed a signal from the information structure she had chosen. Crucially, the success probability p is equal to the posterior a Bayesian would have held at that stage. For instance, if, in the first part, the subject faced the incentive m and had observed a signal σ from information structure I , then the corresponding lottery in part two is given by $g = m$, $l = -L + m$, and $p = \gamma_{\sigma, I}$. I test the Undue Inducement Condition (4) by estimating system (3) on these certainty equivalents.

The ex-ante welfare analysis of this experiment relies on the assumption that choices in part 2 reveal subjects' welfare-relevant preferences while those in part 1 do not. As Bernheim and Taubinsky (2018) elucidate, all behavioral welfare economic analyses involve a separation of choice domains into those that are afflicted by 'characterization failure' and those that merit deference to subjects' choices (the 'welfare-relevant domain'). In the context of this experiment, a large literature on fallacies in belief updating (see Benjamin, 2019, for a review) suggests that decisions made in part 1 of the experiment may be afflicted by characterization failure. By contrast, the literature typically interprets choices over lotteries with transparently described outcome distributions (part 2 of the experiment) as normatively valid, thus justifying the assumptions underlying my ex-ante welfare analysis. For readers who question the normative validity of subjects' choices in both parts of the experiment, ex-post welfare analysis will provide the relevant test of the Undue Inducement Condition.

Figure 3: Information choice (screenshot).

	Ask the bold advisor.	Ask the cautious advisor.
IF...	Advice you'll get	Advice you'll get
If venture succeeds Keep \$0.5	100% chance: "Take it!" (good advice)	50% chance: "Take it!" (good advice) 50% chance: "Don't take it!" (too pessimistic)
If venture fails Cannot keep \$0.5 Pay damages \$5 Net loss \$5	50% chance: "Don't take it!" (good advice) 50% chance: "Take it!" (too optimistic)	100% chance: "Don't take it!" (good advice)

Framing Part 1 of the experiment is framed as follows. The subject decides whether to accept a ‘venture’ that may either ‘succeed’ or ‘fail,’ in exchange for a ‘venture participation payment.’ If the venture succeeds, she can keep the venture participation payment, and no further consequences occur. If the venture fails, she must ‘pay damages’ and may or may not keep the venture participation payment (depending on treatment; see below). Before deciding whether to participate, the subject chooses between a ‘bold advisor’ (the participation-biased information structure) and a ‘cautious advisor’ (the abstention-biased information structure), as in Figure 3. She then observes a stochastic signal which reads either “The [type] advisor recommends: Participate in the venture!” (accompanied with a thumbs-up symbol on green background) or “The [type] advisor recommends: Don’t participate in the venture!” (accompanied with a thumbs-down symbol on red background). In order to alert the subject to the fact that a new state is drawn in each round, each round begins with a display of 20 red and green symbols signifying ventures that are successes and failures. She clicks a first time to hide the colors, and clicks three more times to shuffle the ventures (animation). A final click on a button randomly selects one of the ventures. The subject learns that the venture thus selected is hers for the round.²⁸

Part 2 of the experiment is framed neutrally. It simply describes the amount of money the subject can gain or lose from participating in each gamble, the corresponding success probability, and the certain amount of money the subject receives or loses if she refuses the gamble.

²⁸The shuffling animation is coded such that no individual will be under the impression that she has more information than the prior probability about whether the selected venture is a success or a failure.

Treatments Treatments vary along five dimensions. The first two dimensions serve to test the theoretical predictions from Section 2; I administer these within subjects.

1. *Incentive amount.* The effect of the incentive amount m is the main variable of interest. All my tests will make use of this variation. Specific amounts are given below.

2. *Information choice.* In order to test whether endogenous information choice amplifies the effect of incentives on participation (corollary 1) I vary whether subjects know the incentive amount when choosing an advisor. The *Incentive First* condition proceeds as described above; subjects learn the incentive before selecting an information structure. The *Advisor Choice First* condition ask subjects to choose an information structure *before* learning the incentive amount; that information is revealed simultaneously with the signal provided by the information structure. While subjects know the distribution of incentive amounts they may be offered in this condition, they cannot tailor information choice to the specific incentive they will face.²⁹

The remaining three treatment dimensions let me explore the context dependence of my findings. Variation along the *Compensability* dimension occurs within subject, the remaining variation occurs across subjects.

3. *Prior failure probability.* I vary the prior probability of the state because deviations from Bayesian belief updating may be more pronounced for some priors than for others.³⁰ If such difficulties interact with the incentive amount (for instance through selective information acquisition), harmful effects of incentives may be more likely for some prior probabilities than for others.

4. *Timing of consequences.* I vary the point time at which potential negative consequences of the transaction occur. For subjects in the *Contemporaneous* condition, all payments accrue at once, shortly after completion of the study. It is conceivable, however, that immediately disbursed incentives more likely cause harm if potential negative consequences of the transaction occur with a delay, for instance because a delay makes them easier to ignore.³¹ Therefore, subjects in the *Delayed* condition receive incentive payments immediately, but potential losses take effect only with a three-months delay. To practically implement this possibility, subjects in the Delayed condition receive a part of their study completion payment with a 3 months delay. Potential losses will be discounted from that payment. In reality, transactions with immediate consequences (such as discomfort during stage 1 clinical trials) and transactions with delayed consequences (such as the potential long-term medical effects of human egg donation) both exist.

5. *Compensability.* In the *Compensable* condition, a subject who participates in the bad state loses L but nonetheless receives the participation payment m , leading to a net loss of $L - m$. By contrast, a subject in the *Noncompensable* condition does not receive the participation payment m if she participates in the bad state, implying a net loss of L . There are two motivations for this variation. First, negative welfare effects of incentives may be more pronounced if subjects have more to lose, as they do in the *Noncompensable*

²⁹Variation in whether subjects learn their incentives before or after information acquisition has been used in many previous studies such as Ellingsen and Johannesson (2005); Gneezy et al. (2015); Saccardo and Serra-Garcia (2020), starting with Babcock et al. (1995).

³⁰For instance, due to base rate neglect (Tversky and Kahneman, 1974).

³¹See, for instance, Burrows (1995): “The law is concerned with the investor’s short-sighted view of his preferences; he may not ex ante fully anticipate the extent of the injury he would suffer, an injury that may be bound up with a significant change in preferences following the completion of the transaction.” Anticipatory utility may affect choices in this condition, but preferences for the intertemporal resolution of uncertainty may not. The reason is that subjects learn whether or not they will incur the loss in three months immediately after completion of the experiment. Crucially, subjects know that this information will be provided regardless whether a decision from part 1 or part 2 determines their payment.

condition. Second, real-world cases of participation incentives fall between two extremes. At one extreme, such as with purely financial losses, participation incentives can be fully used to compensate for negative effects of participation. At the other extreme, negative consequences can be compensated only imperfectly (for instance if they affect health) or not at all (in the case of death).

Throughout, I refer to variation within a single one of these five dimensions as a *condition*. I reserve the term *treatment* to refer to a vector specifying a value on each of the five dimensions.

Parameters Subjects receive a completion payment of €15 to which gains are added and from which losses are discounted. The amount a subject can lose from participating in the gamble in the bad state is $L = €6$. Subjects face incentive amounts $m \in \{1, 2, 4, 5\}$.³² Subjects in the Delayed condition know that they will receive €9 of the completion payment immediately and €6 with a three months delay. While these amounts are small in absolute terms, the incentive of $m = €5$ is high compared to the possible loss (€1 or €6 in the compensable and non-compensable conditions, respectively). More importantly, the experiment produces results consistent with the behavioral part of the Undue Inducement Hypothesis.

I choose the set of information structures based on two considerations. First, I aim to maximize the behavioral and welfare impacts of information biasing by choosing the most strongly skewed information structures subject to the constraint that signal $\mathcal{G}$ is more likely than signal $\mathcal{B}$ if and only if the state is good. If participation incentives are harmful and if information biasing is the underlying reason, then this choice leads to the highest chance of detecting such harmful effects.³³ Second, proposition 1 requires that each incentive amount is associated with at least one instrumental information structure.³⁴ Given that subjects may have heterogeneous and potentially extreme risk preferences, instrumentality can be ensured only through information structures that may produce signals that fully reveal the state. Accordingly, I select $I^{\mathcal{G}} = (1, 0.5)$ and $I^{\mathcal{B}} = (0.5, 0)$. In words, the participation-biased information structure produces signal $\mathcal{G}$ with probability 1 if the state is good and with probability 0.5 if the state is bad, while the abstention-biased information structure produces signal $\mathcal{B}$ with probability 0.5 if the state is good and with probability 0 if the state is bad. I use prior success probabilities $\mu \in \{0.2, 0.5, 0.8\}$. Each subject is randomly assigned to one of these prior probabilities, which remains unchanged for that subject throughout the experiment. As shown in Section 2.2, these parameters permit the Undue Inducement Condition (4) to be satisfied or to be violated in each treatment condition, both according to the ex-ante and ex-post welfare standard.

In the first part of the experiment, each subject completes 18 rounds in random order. Each round proceeds as explained above, and is followed by an elicitation of the subjects' posterior belief that the state in the current round is good.³⁵ 16 rounds present her with all combinations of (i) the four incentive amounts, (ii) the Incentive First and Advisor Choice First conditions, and (iii) the Compensable and Noncompensable

³²I conducted the experiment at the University of Cologne. According to the most recent PPP data from 2019, USD 1 = EUR 0.737 (<https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm>). The University of Cologne estimates monthly student living expenses of €832 (<https://portal.uni-koeln.de/ar/international/study-in-cologne/costs-financing>).

³³The same reason motivates my choice to present subjects with only two information structures, neither of which is unbiased in the sense that it is equally accurate in both states.

³⁴An information structure is *instrumental* for a decision problem if the optimal action depends on the signal realization; i.e. the agent does not find it optimal to ignore the signal.

³⁵Subjects select one of 12 bins, corresponding to 0%, 1-10%, 10-19%, 20-29%, ..., 80-89%, 90-99%, or 100% certainty that the state in that round was good. Truth-telling is incentivized by the probabilistic quadratic scoring rule as proposed in Karni

conditions. The two remaining rounds offer incentive $m = 0$ in the Incentive First condition, once framed as compensable and once framed as non-compensable (there is no substantive difference).

In the second part, subjects face multiple decision lists consisting of 11 binary choices each. On each line of each list, subjects in the Contemporaneous condition decide between the option “Win ϵm with $\gamma \cdot 100\%$ chance, lose L with $(1 - \gamma) \cdot 100\%$ chance” and the option “Receive ϵc for sure” (or “Lose ϵc for sure” in case c is negative). For subjects in the Delay condition the certain payment is presented as “Your immediate payment rises [falls] by ϵc for sure.” In case of non-compensable delayed consequences, the lottery is described as “With $\gamma \cdot 100\%$ chance: your immediate payment rises by ϵm , and your delayed payment stays unchanged. With $(1 - \gamma) \cdot 100\%$ chance: your immediate payment stays unchanged, and you lose ϵL of your delayed payment.” If the round corresponds to the Compensable condition, the text “your immediate payment stays unchanged” is replaced by “your immediate payment rises by ϵm .” Here, γ is the Bayesian posterior corresponding to a signal the subject may have received from an information structure. In each round, the sure payment amounts are $c \in \{-6, -4.5, -3, -1.5, -0.5, 0, 0.5, 1.5, 3, 4.5, 6\}$. The binary decisions are shown in a random order, drawn anew in each round.³⁶

Table 1: Lotteries for stage 2 of Experiment 1

€ Gain (g)	€ Loss (l)	Success prob. (p)	Corresponds to Comp. cond.	€ Gain (g)	€ Loss (l)	Success prob. (p)	Corresponds to Comp. cond.
4	-2	$\gamma_{\mathcal{B}, I^{\mathcal{S}}}$	No	4	-2	$\gamma_{\mathcal{B}, I^{\mathcal{B}}}$	No
4	-2	$\gamma_{\mathcal{G}, I^{\mathcal{B}}}$	No	4	-6	$\gamma_{\mathcal{G}, I^{\mathcal{S}}}$	Yes
2	-4	$\gamma_{\mathcal{B}, I^{\mathcal{S}}}$	No	4	-6	$\gamma_{\mathcal{B}, I^{\mathcal{B}}}$	Yes
2	-4	$\gamma_{\mathcal{G}, I^{\mathcal{B}}}$	No	2	-4	$\gamma_{\mathcal{G}, I^{\mathcal{S}}}$	No
4	-6	$\gamma_{\mathcal{B}, I^{\mathcal{S}}}$	Yes	2	-4	$\gamma_{\mathcal{B}, I^{\mathcal{B}}}$	No
4	-6	$\gamma_{\mathcal{G}, I^{\mathcal{B}}}$	Yes	2	-6	$\gamma_{\mathcal{G}, I^{\mathcal{S}}}$	Yes
2	-6	$\gamma_{\mathcal{B}, I^{\mathcal{S}}}$	Yes	2	-6	$\gamma_{\mathcal{B}, I^{\mathcal{B}}}$	Yes
2	-6	$\gamma_{\mathcal{G}, I^{\mathcal{B}}}$	Yes	1	-5	$\gamma_{\mathcal{G}, I^{\mathcal{S}}}$	No
5	-1	$\gamma_{\mathcal{G}, I^{\mathcal{S}}}$	No	1	-5	$\gamma_{\mathcal{B}, I^{\mathcal{B}}}$	No
5	-1	$\gamma_{\mathcal{B}, I^{\mathcal{B}}}$	No	1	-6	$\gamma_{\mathcal{G}, I^{\mathcal{S}}}$	Yes
5	-6	$\gamma_{\mathcal{G}, I^{\mathcal{S}}}$	Yes	1	-6	$\gamma_{\mathcal{B}, I^{\mathcal{B}}}$	Yes
5	-6	$\gamma_{\mathcal{B}, I^{\mathcal{S}}}$	Yes	0	-6	$\gamma_{\mathcal{G}, I^{\mathcal{S}}}$	No
4	-2	$\gamma_{\mathcal{G}, I^{\mathcal{S}}}$	No	0	-6	$\gamma_{\mathcal{B}, I^{\mathcal{B}}}$	Yes

Note: Each row in each half of the table displays a separate lottery. The table displays the loss amounts for the Contemporaneous condition. In the Delay condition, the listed loss amount consists of a loss of €6 incurred three months after the experiment and the smaller gain amount paid immediately. Notation: $\gamma_{\sigma, I}$ denotes the Bayesian posterior that $s = G$ upon observing signal $\sigma \in \{\mathcal{G}, \mathcal{B}\}$ from information structure $I \in \{I^{\mathcal{G}}, I^{\mathcal{B}}\}$. Independently of the prior, we have $\gamma_{\mathcal{B}, I^{\mathcal{S}}} = 0$, $\gamma_{\mathcal{G}, I^{\mathcal{B}}} = 0$. Remaining posteriors vary with the prior. For a prior success probability of 0.2, $\gamma_{\mathcal{G}, I^{\mathcal{S}}} = \frac{1}{3}$ and $\gamma_{\mathcal{B}, I^{\mathcal{B}}} = \frac{1}{9}$. For a prior success probability of 0.5, $\gamma_{\mathcal{G}, I^{\mathcal{S}}} = \frac{2}{3}$ and $\gamma_{\mathcal{B}, I^{\mathcal{B}}} = \frac{1}{3}$. For a prior success probability of 0.8, $\gamma_{\mathcal{G}, I^{\mathcal{S}}} = \frac{8}{9}$ and $\gamma_{\mathcal{B}, I^{\mathcal{B}}} = \frac{2}{3}$.

Subjects complete a total of 26 multiple decision lists covering the lotteries listed in Table 1. For each of the four incentive amounts and for each of the Compensable and Noncompensable conditions, the subject sees one decision list in which the success probability is given by the posterior a Bayesian would hold after

(2009) and Holt and Smith (2009). Subjects can return from the belief elicitation stage to the betting stage to change their decision if they like.

³⁶This design feature serves to alleviate biases that may occur in ordered lists of multiple decisions, such as a tendency to switch towards the middle of the page.

observing signal $\mathcal{G}$ from information structure $I^{\mathcal{G}}$ and another one corresponding the posterior a Bayesian would hold after observing signal $\mathcal{B}$ from information structure $I^{\mathcal{B}}$, leading to 16 lists. In addition, subjects proceed through two lists corresponding to incentive zero for these two success probabilities. The posteriors corresponding to these 18 rounds are non-degenerate. Subjects also complete a decision list for the degenerate posteriors corresponding to signal $\mathcal{B}$ from information structure $I^{\mathcal{G}}$ (which is zero) and corresponding to signal $\mathcal{G}$ from information structure $I^{\mathcal{B}}$ (which is one). For each of the two degenerate posteriors, subjects complete one round corresponding to the incentive amounts €2 and €4 both corresponding to the Compensable and Noncompensable condition, totaling 8 rounds corresponding to degenerate posteriors.

Implementation All subjects complete part 1 before completing part 2.³⁷ Within each part of the experiment, subjects proceed through the rounds in individually randomized order. For a random half of subjects, the bold advisor is displayed on the left of the screen and the cautious advisor on the right; for the remaining half of subjects the advisors appear in the opposite order.

The study payment is entirely determined by a single, randomly drawn decision. With a 2/3 chance that decision is from part 1; with the remaining 1/3 chance it is from part 2. In case of the former, it is determined with an 80% chance by the decision whether to participate in the venture and the success of the venture. With the remaining 20% chance it is determined by the belief elicitation.

3.2 Analysis

Data and preliminary analysis A total of 348 subjects from the pool of the Cologne Laboratory for Economic Research at the University of Cologne completed the study online between November 25 and 27, 2020. All materials were presented in English. Subjects were recruited with ORSEE (Greiner, 2015) and were paid through PayPal. All amounts are in Euros, all fees were covered by the experimenter. Subjects took between 45 and 90 minutes to complete the study. Table 2 displays the number of subjects in each across-subject condition.

Because subjects participated online, attrition is a potential concern. The 348 complete responses make up 95.6% of the 364 subjects who started the survey on a machine that satisfied the technical requirements.³⁸ Out of the 16 attriters, 3 did not proceed to the first comprehension check, 10 dropped out at the first comprehension check, 1 dropped out at the second comprehension check, and 2 left the survey after the second comprehension check. At the point of attrition, subjects may have learned the prior failure probability and the delay condition to which they are assigned. Attrition is unrelated to these treatment conditions.³⁹ More importantly, the crucial variations in the incentive amount and information order are varied within subject and hence are immune to differential attrition by design.

³⁷The reason for this design choice is that the decisions in part 1 are more cognitively demanding.

³⁸Subjects could only participate on desktop or laptop computers with sufficient screen width. Participants logging in from mobile devices or from other devices with small screens were automatically terminated.

³⁹Of the 16 attriters, 8 were in the Delay condition, and 4, 7, and 5 were in the 80%, 50%, and 20% prior failure probability condition, respectively. Regressing an indicator for whether a subject attrited on an indicator for the Delay condition, the prior failure probability and the interaction between the two has no explanatory power ($p > 0.3$ on all coefficients except for the constant term, and $p > 0.7$ for the F -test of joint explanatory power of all variables compared to a constant-only model).

Table 2: Number of subjects in Experiment 1.

	Prior failure probability			Total
	0.8	0.5	0.2	
Delay condition				
No	53	64	58	175
Yes	64	50	59	173
Total	117	114	117	348

The certainty equivalents that subjects reveal in part 2 of the experiment are interval-coded. Following standard procedures, I use interval midpoints for analysis. Because I elicit the certainty equivalents in multiple price lists with randomly ordered certain amounts, subjects’ choices are potentially non-monotonic (in the sense that a subject chooses a sure amount c over the lottery but choose the lottery over some larger sure amount c'). Averaging across all subjects, non-monotonicities occur in 2.9 of the 26 rounds (median: 2 rounds).⁴⁰ In case of a multiple switch, I use the lowest certain amount the subject chooses over the lottery. Accordingly, the certainty equivalents I use correspond to a lower bound on the welfare from participating in the transaction.⁴¹

Each information structure in this experiment provides certainty about the state of the world following one of its signal realizations. To keep the duration of the experiment manageable, I elicit certainty equivalents for such degenerate lotteries only for payment amounts €2 and €4. I infer the certainty equivalents for degenerate lotteries corresponding to the €5 incentive by scaling the measured certainty equivalent corresponding to €4 incentive. Likewise, I scale the certainty equivalents for degenerate lotteries corresponding to the €1 incentive by scaling the measured certainty equivalent corresponding to the €2 incentive.⁴²

Information choice and participation decisions I begin the analysis by studying the effect of incentives on information choice. Panel A of Figure 4 displays the fraction of subjects who opt for the bold advisor, I^g , by incentive amount and information choice condition, pooling across the Delay, Compensability, and Prior conditions. If subjects know the incentive amount when choosing an advisor (the Incentive First condition), their preference for the bold advisor is strongly increasing in the incentive amount, consistent with proposition 1. Specifically, if the incentive amount is €1, subjects choose the bold advisor 25% of the time. That fraction more than doubles, to 51%, as the incentive amount m increases to €5. By comparison, if subjects cannot tailor information choice to the incentive amount (the Advisor Choice First condition), they display a preference for cautious advice, opting for the bold advisor just 37% of the time. To confirm these findings econometrically, I run an OLS regression of an indicator for choosing of the Bold Advisor on the incentive amount, an indicator for the Incentive First condition, as well as the interaction between the

⁴⁰There is no substantial difference in the incidence of non-monotonic choices across the Contemporaneous and Delay conditions ($p > 0.5$), even though the lotteries are arguably more difficult to understand in the latter case.

⁴¹My results regarding welfare are qualitatively unchanged if I instead only include observations in which subjects exhibit monotonicity.

⁴²For instance, I infer a subjects’ certainty equivalent for the lottery that pays €5 for certain by scaling her certainty equivalent for the lottery that pays €4 for certain by the factor 4/5. In the Contemporaneous condition, I could instead define the certainty equivalent as equal to the payoff of the degenerate lottery. This is not possible in the Delay condition, however, because certainty equivalents for degenerate lotteries incorporate subjects’ the time preferences.

Figure 4: Information choice and supply in Experiment 1.

Notes Panel A displays the frequency with which subjects choose the participation-biased information structure, by participation payment and information condition, pooling across all remaining conditions. Panel B displays the fraction of subjects participating in the transaction, by participation payment and information condition, pooling across all remaining conditions. Whiskers represent 95% confidence intervals (standard errors clustered by subject).

two. I cluster standard errors by subject. I find that a €1 increase in the incentive causes a 5.8 percentage point increase in the frequency with which subjects opt for the Bold advisor ($p < 0.001$).

The possibility for incentive-induced information-biasing also affects participation decisions, as Panel B of Figure 4 shows. It plots supply curves in the Incentive First and Advisor Choice First conditions, averaging across the Compensability, Delay, and Prior conditions. If the incentive is €1 or €2, subjects participate in the venture with approximately equal frequency across the Information conditions. As the incentive increases, differences in participation rates emerge. With a €5 incentive, subjects in the the Advisor Choice First condition participate in the venture in 53% of cases. This number is 6 percentage points larger (59%) if subjects know the incentive amount when selecting the advisor (the Incentive First condition). Hence, consistent with corollary 1, endogenous information choice amplifies the effect of incentives on the supply of participants.⁴³ To document these effects formally, I regress an indicator of the participation choice on the incentive amount, on an indicator for the Incentive First condition, and on the interaction between the two. I use OLS, clustering standard errors on the subject level. I find that each €1 increase in the participation payment increases the probability of participation by an additional 2.4 percentage points in the Incentive First condition compared to the Advisor Choice first condition. This result is highly statistically significant ($p < 0.01$). It is worth noting that the effect on participation is necessarily smaller than the corresponding

⁴³Corollary 1 holds under the condition that both information structures are instrumental at all incentive levels. Empirically, this condition is satisfied for a large fraction of subjects. In the Advisor Choice First condition, across all incentive amounts, subjects are at least 42 percentage points more likely to participate if the chosen advisor recommends participation rather than abstention (averaged across the Compensability, Delay, and Prior conditions). In fact, unless the subject receives a signal from the Bold Advisor in the €1 incentive condition, the effect size is at least 69 percentage points. While Corollary 1 does not predict an increase in participation in the Advisor Choice First condition, such an increase will occur if some subjects decide to participate unconditionally if the incentive is high and to abstain unconditionally if the incentive is low.

effect on advisor choice. To see why, consider the case of a 50% prior failure probability. In this case, the Bold Advisor will recommend participation in 75% of cases whereas the Cautious Advisor will recommend participation in 25% of the cases, a difference of 50 percentage points. Accordingly, if subjects always follow the recommendation from the chosen advisor, the difference in the slopes of the supply curves across the Incentive First and the Advisor Choice First conditions will be half as large as the effect on advisor choice.⁴⁴

The foregoing analysis pools across treatment conditions. Appendix B.1 shows that the same qualitative results obtain within the vast majority of treatment conditions, though not always to a statistically significant extent.

Overall, these results are consistent with the behavioral part of the Undue Inducement Hypothesis. Next, I test whether the evidence justifies the normative part of that hypothesis.

Welfare I study whether higher incentives may harm participants in the sense of Proposition 3, using ex-ante welfare analysis. I define a subject’s ex ante welfare $w_i(m)$ from the possibility to participate in a transaction at incentive m as a function of the subject’s certainty equivalent c_i for the gamble corresponding to the participation decision based on Bayesian posteriors. Specifically, $w_i(m) = c_i$ if the subject accepts, and $w_i(m) = 0$ if the subject rejects. Panel A of Figure 5 displays cumulative distribution functions of individual welfare, w_i . I plot the CDFs separately for the Compensable condition (Panel A) and the Noncompensable condition (Panel B), pooling across the Timing and Prior conditions. I only use data from the Incentive First condition because only that condition allows for the possibility that biased information choice underlies potential harmful effects of incentives. For easier visibility, I combine the €1 and €2 incentive conditions into a low incentive category, and the €4 and €5 conditions into a high incentive category. In each panel, mass to the right of zero indicates decisions for which the possibility to participate in the transaction increases subjects’ ex ante welfare. Mass to the left of zero indicates ex-ante mistaken decisions. In these trials, subjects participated in transactions for which they exhibit a negative certainty equivalent. Both panels show that an increase in the incentive causes a large increase in the gains of those who benefit, and a small increase in the losses of those who are harmed. Due to the latter effect, directional information alone is not sufficient to draw conclusions about the Undue Inducement Condition.

Panels C and D show how average welfare gains, $g(m) = \frac{1}{n} \sum_{i=1}^n \max(w_i(m), 0)$, and average welfare losses, $l(m) = \frac{1}{n} \sum_{i=1}^n \min(w_i(m), 0)$, depend on the incentive amount, using the same data and pooling as Panels A and B, respectively. Here, n denotes the total number of subjects who have the possibility to participate in the transaction.⁴⁵ Visual inspection of Panel C immediately shows that the Undue Inducement Condition (4) is violated in the case of compensable consequences: the intersection of the linear extensions of $g(m)$ and $l(m)$ lies below the horizontal axis. Two features are responsible for this violation. First, at low incentives, welfare losses substantially outweigh welfare gains. According the estimates of system (3) using a

⁴⁴This result is independent of the prior success probability μ . The probability of receiving a participation signal from the bold advisor is $\mu + (1 - \mu)\frac{1}{2}$ whereas the probability of receiving a participation signal from the cautious advisor is $\frac{\mu}{2}$. The difference between these probabilities is $\frac{1}{2}$ regardless of μ .

⁴⁵The averages $g(m)$ and $l(m)$ are taken across all observations, including those in which the individual abstained from the transaction. The increasing distance between the mean gains and mean losses is partially attributable to the fact that higher incentives cause a larger number of subjects to participate.

seemingly unrelated regression model with standard errors clustered at the subject level,⁴⁶ the mean welfare gain at the €1 incentive is €0.13, compared to a mean welfare loss of €-0.21.⁴⁷ Hence, the transaction is admissible at a low incentive only if the welfare weight on those who are harmed by the transaction is lower than 0.38. Second, while each additional euro of the participation payment increases welfare gains by €0.33 ($p < 0.001$), it causes an insignificant increase in the magnitude of welfare losses of €0.01 ($p > 0.3$). Hence, overall welfare increases with the incentive amount only if the welfare weight on those who are harmed by the transaction is very close to one. These two requirements cannot be satisfied simultaneously. Formally, inserting the estimates into equation (4) we find that its left hand side is given by -0.57 (s.e. 0.10), so the Undue Inducement Condition is violated (statistically significant at the 0.1% level).

Panel D shows similar albeit less pronounced results. Welfare gains increase by €0.22 per additional euro of the participation incentive ($p < 0.001$), whereas losses increase by €0.05 ($p < 0.05$). Focusing on the levels of mean welfare gains and losses at the €1 incentive according to the best linear fit, we find that the losses of €0.24 outweigh the gains of €0.13 by a factor of nearly two. Accordingly, equation (4) evaluates to -0.47 (s.e. 0.11), indicating a violation of the Undue Inducement Condition (statistically significant at the 0.1% level).

The foregoing analysis pools across treatments, which may mask heterogeneity in treatment effects. Hence, I repeat it separately for each treatment. I report intercepts at $m = 1$ rather than at $m = 0$, using the notation $\tilde{\beta}_0^w = \beta_0^w + \beta_1^w$, for $w \in \{G, L\}$.⁴⁸ It is important to note that if certainty equivalents are elicited with noise, their classification into gains and losses may be mistaken for some observations.⁴⁹ This may lead to biased estimates. While the current exposition ignores this issue for simplicity, I address it in Appendix B.2 where I obtain qualitatively similar findings.⁵⁰

Table 3 displays the estimates. Each row corresponds to a separate treatment. Columns 1 and 2 show the levels of mean gains and losses at the incentive $m = 1$, $\tilde{\beta}_0^G$ and $\tilde{\beta}_0^L$, respectively. Columns 3 and 4 show the marginal effect of a €1 increase in mean gains and losses, β_1^G and β_1^L , respectively. Column 5 shows the minimal weight $\underline{\alpha} = \frac{\beta_1^G}{\beta_1^G - \beta_1^L}$ that needs to be placed on those who are harmed by the transaction such that high incentives are deemed worse than low incentives. Column 6 displays the maximal weight $\bar{\alpha} = \frac{\tilde{\beta}_0^G}{\tilde{\beta}_0^G - \tilde{\beta}_0^L}$ that may be placed on those who lose from the transaction such that the transaction is admissible at the low incentive. As explained in Section 2.2, the Undue Inducement Condition (4) simply requires that the upper bound $\bar{\alpha}$ exceeds the lower bound $\underline{\alpha}$. I use the difference between the two bounds, $T = \bar{\alpha} - \underline{\alpha}$, as test statistic, with standard errors calculated by the delta method. Column 7 reports whether the condition is

⁴⁶I use Stata's user-programmed command `mysureg`.

⁴⁷Here, I evaluate levels at the lowest incentive rather than at the intercept because the incentive amount in this experiment is the only possible source of welfare gains. See footnote 24 for a formal justification.

⁴⁸The motivation is explained in footnote 24.

⁴⁹Stochastic choice might also affect participation decisions. In the context of this paper, that type of randomness plays a fundamentally different role. If subjects make participation decisions in a random or thoughtless manner, these decisions are potentially mistaken. Whether incentives affect the frequency of random or otherwise negligent choice is the main focus of this paper.

⁵⁰Appendix B.2 addresses the issue by averaging out elicitation noise. Such averaging is nontrivial because subjects' decisions differ across all rounds of part 2. For each round r , I predict a subject's choices in that round using her choices in all remaining rounds, and perform welfare analysis using the predicted values. In order to circumvent the need to specify a model of decision making under risk, I use ridge regression, a statistical learning tool which has been shown to predict risky choice better than most formal models (Peysakhovich and Naecker, 2017).

Figure 5: Welfare in Experiment 1.

Notes Panels A and B display cumulative distribution functions of certainty equivalents corresponding to the transactions the subjects faced in part 1 of the experiment, for the Compensable and Noncompensable conditions, respectively. Only data from the Incentive First condition are shown, averaged across the Prior and Timing conditions. For trials in which subjects refused the transaction, the certainty equivalent is set to zero. For easier visibility, I group the €1 and €2 incentives into the low incentive category, and I group the €4 and €5 incentives into the high incentive category. Probability mass to the left of zero indicates welfare losses, probability mass to the right of zero indicates welfare gains. Panels C and D display average welfare gains $g(m) = \frac{1}{n} \sum_{i=1}^n \max(w_i, 0)$ and average welfare losses $l(m) = \frac{1}{n} \sum_{i=1}^n \min(w_i, 0)$ by incentive amount, using data from the Incentive First condition, averaged across the Prior and Timing conditions. Whiskers indicate 95% confidence intervals. Standard errors clustered by subject.

satisfied, and column 8 lists the p -value for a two-sided test of the Null hypothesis that $T = 0$. Throughout, I cluster standard errors on the subject level.

The top half of Table 3 shows that the Undue Inducement Condition is never satisfied in the Compensable condition, for any Prior or Timing condition. The reason is that in five out of six cases, total welfare losses decrease in magnitude or stay constant as the incentive increases. Yet, in at least four out of six conditions, significant welfare losses at the incentive $m = 1$ contrast with negligible gains. These two facts preclude the possibility of welfare weights under which low but not high incentives are admissible.

Table 3: Welfare gains, welfare losses, and tests for undue inducement in Experiment 1

Variable	(1) β_0^G	(2) β_0^L	(3) β_1^G	(4) β_1^L	(5) $\underline{\alpha}$	(6) $\bar{\alpha}$	(7) Condition (4)	(8) satisfied p -value
<i>Consequences</i>								
Compensable								
Contemporaneous								
Prior failure probability								
80%	0.155*** (0.059)	-0.224*** (0.083)	0.051 (0.031)	-0.017 (0.027)	0.753** (0.318)	0.409** (0.138)	No	0.416
50%	0.100 (0.070)	-0.267** (0.105)	0.304*** (0.056)	0.025 (0.031)	1.090*** (0.119)	0.272*** (0.165)	No	0.001
20%	0.267*** (0.091)	-0.121** (0.055)	0.550*** (0.063)	0.002 (0.025)	1.003*** (0.045)	0.689 (0.139)	No	0.057
Delayed								
Prior failure probability								
80%	0.000 (0.057)	-0.207* (0.109)	0.208*** (0.050)	0.017 (0.035)	1.091*** (0.200)	0.000*** (0.276)	No	0.002
50%	0.114* (0.069)	-0.362*** (0.123)	0.353*** (0.060)	-0.028 (0.043)	0.927*** (0.108)	0.239*** (0.134)	No	0.000
20%	0.174** (0.074)	-0.080** (0.040)	0.506*** (0.058)	-0.082** (0.041)	0.861*** (0.066)	0.685 (0.154)	No	0.311
Noncompensable								
Contemporaneous								
Prior failure probability								
80%	0.112* (0.064)	-0.361** (0.170)	0.062 (0.041)	-0.035 (0.059)	0.638 (0.440)	0.237* (0.143)	No	0.468
50%	0.175*** (0.063)	-0.113 (0.088)	0.159*** (0.046)	-0.126*** (0.043)	0.558*** (0.114)	0.608 (0.213)	Yes	0.861
20%	0.177** (0.082)	-0.119* (0.071)	0.451*** (0.063)	-0.031 (0.026)	0.936*** (0.053)	0.597 (0.194)	No	0.127
Delayed								
Prior failure probability								
80%	0.069 (0.046)	-0.444*** (0.161)	0.064* (0.033)	0.040 (0.044)	2.712 (5.456)	0.134* (0.091)	No	0.637
50%	0.168** (0.077)	-0.237** (0.119)	0.221*** (0.058)	-0.041 (0.051)	0.844*** (0.174)	0.415*** (0.176)	No	0.153
20%	0.095 (0.065)	-0.159** (0.066)	0.360*** (0.049)	-0.072** (0.033)	0.833*** (0.070)	0.375*** (0.200)	No	0.038

Note: Each row corresponds to a separate treatment. Coefficients in each row are estimated in a separate SUR-regression. Columns 1 and 2 show the level of mean welfare gains and losses, respectively, at the €1 incentive. Columns 3 and 4 show the change in mean welfare gains and losses, respectively as the incentive increases by €1. Column 5 shows the least weight α that needs to be placed on those who lose from the transaction such that an increase in the incentive is deemed welfare-decreasing. Column 6 shows the maximal weight α that may be placed on those who lose from the transaction such that the transaction is admissible at sufficiently low incentives. Reported significance levels in columns 5 and 6 reflect the tests that $\underline{\alpha} = 0$ and $\bar{\alpha} = 1$, respectively. Column 7 indicates whether $T = \bar{\alpha} - \underline{\alpha}$ is nonnegative. Column 8 lists p -values for two sided tests of the Null hypothesis that $T = 0$. Standard errors in parentheses, clustered by subject. Standard errors in columns 5, 6, and 8 are calculated by the Delta method.

The bottom half of the table displays estimates for the Noncompensable condition. Column 4 shows that the magnitude of losses now increases with the incentive amount in five out of six cases. While the Undue Inducement Condition is satisfied one case (contemporaneous, noncompensable consequences with a 50% prior failure probability), the point estimate on the test statistic for the Undue Inducement Condition, T , is far from statistically significant. Indeed, in that condition, low but not high incentives are admissible only if the welfare weight α lies in the narrow range from 0.558 to 0.608. Also note that the estimate of the effect on losses is much larger in this case than in any other case, suggesting the possibility of a statistical outlier. The effect disappears once I account for noisy elicitation of certainty equivalents (see Appendix B.2).

Next, I check how the welfare effects vary across the individual treatments. Column 6 shows that a high prior failure probability leads to substantially tighter upper bounds on the welfare weight that may be placed on those who are harmed by the transaction while keeping the transaction admissible at a low incentive. In all cases with an 80% prior failure probability, a utilitarian ($\alpha = 0.5$) prefers preventing the transaction to allowing it. Prevention of the transaction at the lowest incentive is also the optimal utilitarian policy in all cases of delayed, noncompensable consequences. Moreover, the fact that the parameters in column 3 exceed those in column 4 throughout implies that utilitarian welfare increases with the incentive in each treatment. Hence, the policy espoused the undue inducement literature, namely to allow the transaction at low but not high incentives, *minimizes* utilitarian welfare in these cases.

Table 3 tests the Undue Inducement Condition in an ex-ante welfare framework. Appendix Table B.4 reproduces Table 3 using ex-post welfare, setting w_i equal to the realized monetary gain or loss for each subject in each trial. The Undue Inducement Condition is violated throughout. Since the ex-post benchmark offers more scope for the planner to disagree with the subjects' own choices than the ex-ante benchmark, the violation of the Undue Inducement Condition according to the ex-post benchmark is an arguably stronger result. The violation of the Undue Inducement Condition according to the ex-post benchmark is particularly noteworthy in the Delay condition in which the planner imposes judgment not only on risk preferences (by weighting outcomes in different states) but also on time preferences (by weighting outcomes that obtain at different points in time).⁵¹ Appendix Table B.4 shows that even a social planner who takes the stance that future consequences should not be discounted at all will not want to permit the transaction but cap incentives.

A potential criticism of the current experiment concerns the fact that subjects have access to only one of two advisors, both of which provide highly skewed information. Would the results differ if subjects could choose information more flexibly? Appendix D answers this question in a separate experiment in which subjects acquire information dynamically and thus can implement nearly any conceivable information structure. Using the same treatment conditions as in the present experiment (but slightly different incentive amounts), I find no single case in which the Undue Inducement Condition (4) is satisfied.

Overall, this experiment shows that even though incentives cause the type of behavior that has been interpreted as evidence for deficient decision making, it is virtually never the case that low but not high incentives are admissible, regardless of the welfare weight the social planner places on those who are harmed

⁵¹The average subject in the Delay condition is indifferent between a €1 loss incurred with a three-months delay and a €0.83 immediate loss.

by the transaction. Higher incentives do not cause a deterioration in decision quality that warrants caps on incentives.

4 Experiment 2: Visceral Transaction

Experiment 1 relies on the induced preferences paradigm. Therefore, it is subject to two limitations that the current experiment will address. First, in transactions affected by undue-inducement based regulation, prospective participants face the difficult task evaluating non-monetary consequences such as possible health risks in dollars and cents. High incentives may potentially distort this valuation process. Second, Experiment 1 excludes affectual mechanisms which may lead to different decisions than settings constrained to monetary consequences (Rottenstreich and Hsee, 2001).

Experiment 2 addresses these issues by incentivizing subjects to ingest real insects in exchange for money. This is a highly visceral activity that is novel and unfamiliar to the vast majority of subjects. Hence, subjects may find it difficult to evaluate the disutility of participation in monetary terms. The transaction also induces substantial affect. It is intensely aversive to many subjects, some of whom reported that the experiment was “stressful” or that the “insects were scary,” while others refused to even touch the containers that contained the specimens.⁵² In addition, the experiment retains the possibility for selective information search as it lets subjects consult rich and multifaceted information about bug-eating.

The use of a physical transaction comes at the cost that I can no longer observe a ‘true state of the world’ or the type of Bayesian benchmark that is necessary for conducting ex-ante welfare analysis.⁵³ Therefore, welfare analysis in this experiment relies on the ex-post benchmark.

4.1 Design

Structure. This experiment has five stages. First, subjects receive an offer specifying the magnitude of the incentive: either \$3 or \$30 in exchange for eating “food items.” Subjects learn that they will see a sequence of five items, that they will decide whether to eat the specific item in exchange for the promised incentive amount, and that there is an 80% chance that one of these five decisions will be carried out. Only at this point do they learn that all of the food items are whole insects that are either baked, or cooked and dehydrated, and produced for human consumption. Subjects know of both incentive amounts and that they are randomly assigned to one of them. This feature precludes rational inference from the incentive amount. It works in favor of finding detrimental effects of incentives, because subjects tend to infer greater disutility from higher payment amounts (Cryder et al., 2010; Hoffart and Scheibehenne, 2019).

Second, before they decide whether to eat the insects, subjects select and watch one of two 6-minute videos. The *encouraging* video is titled “Why you may want to eat insects,” the *discouraging* video is called

⁵²Even in countries such as China, Thailand, and Mexico, insect eating is not practiced by a wide majority. Rather, it is concentrated within particular regions and / or communities, and is often limited to a small number of insect species. In my data, Asians and Hispanics are neither more nor less willing to eat insects than Caucasians.

⁵³For the related reasons, the present experiment cannot be regarded as a test of the model in Section 2 even if its conclusions are consistent with that model. For instance, subjects may have systematically biased prior beliefs about the physical transaction. Moreover, the naturalistic setting prevents information being drawn from a well-specified stochastic information structure; instead, a small number of specific pieces of information must be created.

Figure 6: Insects eaten by subjects. A. House cricket (*acheta domesticus*) B. Mole cricket (*gryllotalpae*) C. Field cricket (*gryllus bimaculatus*) D. Mealworm (*tenebrio molitor*) E. Silkworm pupa (*bombyx mori*).

“Why you may not want to eat insects.” These titles and the approximate 6-minute duration are all the information subjects have when selecting a video.⁵⁴ Paralleling Experiment 1, each subject watches exactly one video; subjects are not allowed to watch both. This design choice serves to increase the chance of finding harmful effects of incentives. Moreover, it is motivated by the fact that information acquisition in the real world is typically costly, which constrains the number of information sources that can be consulted.⁵⁵ While a video is not a stochastic information structure of the kind modeled in Section 2, the encouraging video is more participation-biased than the discouraging video as long as subjects believe that the encouraging video is more likely to recommend participation than the discouraging video.⁵⁶ Subjects also select at least four out of a selection of 9 video clips. The clips are grouped in bins of three named “Reasons for eating insects”, “Reasons against eating insects”, and “Other information about eating insects.” Unbeknownst to the subjects, the clips are snippets of the aforementioned 6-minute videos. With an exogenous 3% chance, a subject watches the selected clips; with a 97% chance she watches the selected 6-minute video. Subjects know that they will not receive any further information, and that they will decide whether to eat insects in exchange for money based on a brief verbal description alone. Subjects therefore have an incentive to select the videos carefully and pay attention to their content.

Third, subjects reveal reservation prices for eating each of the following five items: 2 house crickets, 5 large mealworms, 3 silkworm pupae, 2 mole crickets, and 2 field crickets. For each item, subjects fill in a multiple-decision list. On each line of each list they choose between the options “Get \$ p . In exchange,

⁵⁴Each video lists various reasons for or against human insect consumption. Transcriptions and links to the videos are included in Appendix E.

⁵⁵In the experiment in Appendix D subjects do not face constraints on the amount or kind of information they consult.

⁵⁶If a Bayesian subject expects that the encouraging video will more likely recommend eating insects, she should discount that recommendation accordingly. It is possible, and consistent with the formal model, that a subject who watches the encouraging video finds it so unconvincing that she forms even more pessimistic beliefs about insect eating than if she had watched the discouraging video.

eat the food item” and “Do not participate in this transaction,” for values of p ranging from \$0 to \$60 in 21 increasingly large steps.⁵⁷ Subjects click on the line at which they prefer to switch from refusing the transaction to accepting it; the remaining choices are filled in automatically to enforce monotonic choice.

Fourth, each subject makes five binary decisions whether to eat each of the five food items in exchange for the incentive amount initially promised, knowing that one of these five decisions will be carried out with an 80% chance. This large implementation probability ensures that the incentive amount originally promised affects the way subjects inform themselves about the food items.

Fifth, all subjects receive five containers containing the insects they may be about to swallow.⁵⁸ Each is filled with specimens of the respective species and a folded piece of paper with a code. Subjects must enter all codes into the computer, which forces them to open each container, remove the label from within and thus view, touch, and (inadvertently) smell each of the food items they may be about to ingest. Subjects then again reveal reservation prices using the same type of interface as in stage 3. The reservation prices elicited at this stage incorporate substantial additional information about the transaction. They will serve as the benchmark for welfare analysis.

Conditions The design just described outlines the *Video* condition. Some subjects participate in the *No Video* condition. These subjects do not choose or watch any video or video clips. Viewing a video may alter reservation prices in a way one might deem normatively invalid. The reservation prices of subjects in the No Video are unaffected by such effects. In addition, the No Video condition provides a benchmark for the effect of information acquisition on participation decisions.

Implementation. Each subject makes multiple decisions. To incentivize truthful revelation of preferences, exactly one of all decisions is randomly chosen for implementation at the end of the experiment. The selected decision entirely determines a subject’s payment and consumption of insects. The implementation probability varies across decisions. There is an 80% chance that the decision selected for implementation is from stage 4, as explained earlier. The chance that a decision from stage 3 or 5 is implemented is 7% each. With the remaining 6% chance, the subject is paid according to a final stage that comes as a surprise to the subject, in which she predicts other subjects’ behavior.⁵⁹ Subjects know that all insects will be consumed in the privacy of a visually secluded space in the presence of only the experimenter who ensures that the subject consumes the animals in entirety. This design choice minimizes social motives such as trying to impress other participants.

Subjects make some consumption decisions before seeing the actual insects, and thus may be unpleasantly surprised. While I cannot force participants to ingest insects against their will, the design needs to ensure that participants do not accept transactions with the expectation that they can easily back out of their choice. Hence, a subject who reneges not only forfeits the money she would have received for eating the

⁵⁷ The amounts are \$0, 1, 2, 3, 4, 6, 8, 10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5, 30, 33, 36, 39, 44, 50, 60. Resolution is finer at lower levels to increase statistical power in light of the positively skewed distribution of reservation prices. The amount \$3 was not included in the decision lists for the first 79 subjects.

⁵⁸ As a filler task during the handing out of the insects, subjects complete an extended version of the Cognitive Response Scale (Toplak, West and Stanovich, 2014), and sets D and E of Raven’s (1960) standard progressive matrices.

⁵⁹ In that stage, subjects predict other subjects’ stage-3 reservation prices.

insect, but also loses an additional \$20 (subjects who rejected the selected transaction cannot renege). This penalty is discounted from a \$35 completion payment the subject would otherwise receive.⁶⁰ Subjects are aware of all of these rules from the outset of the experiment.

At the beginning of the each session I read the initial instructions aloud. All further instructions are displayed on screen.⁶¹ The invitation emails mention that the experiment will involve the consumption of food items on the spot, but do not mention insects.⁶² They ask recipients not to participate if they have food allergies, are vegetarian or vegan, or eat kosher or halal. Subjects know that the experiment is IRB-approved. During the experiment they learn that all the food items are produced in FDA-registered facilities. Accordingly, they have no reason to anticipate health risks.

4.2 Analysis

Data and preliminary analysis A total of 671 subjects participated in one of 39 computerized sessions in May, June, and July 2015 at the Ohio State University (499 subjects), Stanford University (110 subjects), and the University of Michigan (62 subjects),⁶³ recruited using the laboratories' experimental economics participant databases. 271 subjects participated in the No Video condition (136 and 135 with \$3 and \$30 incentives, respectively), and 400 participated in the Video condition (197 and 203 with \$3 and \$30 incentives, respectively).⁶⁴ Each session lasted about 2.5 hours and contained both payment conditions. Either all or none of the subjects in a session were in the Video condition. Five subjects (0.8%) refused to follow through with their decision to eat an insect when that decision was selected for implementation and paid the penalty of \$20.⁶⁵

Randomization into treatments was successful. Of 24 *F*-tests for differences in subjects' predetermined characteristics across the four treatments, only one is significant at the 5% level. Details are in Appendix C.1.

Reservation prices I begin the analysis by studying subjects' reservation prices. Their magnitude helps contextualize the incentive amounts of \$3 and \$30. Moreover, differences in reservation prices between stages 3 and 5 demonstrate that handing out the insects conveys substantial information.

⁶⁰The Stanford IRB required that this penalty be framed as follows: "You will be paid at least \$15 if you complete this experiment, regardless of your choices. In addition ... you automatically receive \$20 if you ... follow through with the decisions you make."

⁶¹The instructions are reproduced in Appendix E. The experiment is coded in Qualtrics.

⁶²An exception exists in the invitation emails in Michigan, and those for the last 31 Stanford subjects, which mentioned that the experiment involves the voluntary consumption of food items, including edible insects. This information had no statistically measurable effect on the fraction of participants who refused to eat an insect for any price offered.

⁶³Power calculations based on the pilots revealed that on the order of 800 subjects would be required. Due to the small size of the Stanford undergraduate population, one can typically obtain only about 150-200 subjects, of which 68 had already been used for an exploratory on this design. I sought universities would accept the Stanford IRB approval, as opposed to requiring a full IRB review of their own. Amongst the universities to which I had access, the Ohio State University was the largest, but could not provide more than roughly 500 subjects, as the experiment was run during the summer quarter. I had planned to obtain the remaining 200 subjects at the University of Michigan, but due to the summer break, only 62 were available.

⁶⁴See Appendix Table C.1 for summary statistics about the subjects. The fraction of subjects in the Video condition exceeds 50% because only that condition reveals information choice. The 79 Stanford students who first participated in this experiment were not given any decisions regarding field crickets. The 48 first Stanford students also did not make any predictions about other participants. In addition, 68 Stanford students participated in an exploratory treatment.

⁶⁵Of these subjects, all were in the \$30-condition, four were in the Video condition, and three had seen the encouraging video.

Table 4: Reservation prices for eating insects

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Species	Res. price All	Res. price 2 house crickets	Res. price 5 meal- worms	Res. price 3 silkworm pupae	Res. price 2 mole crickets	Res. price 2 field crickets
<i>Levels</i>						
Before	23.976*** (0.847)	20.283*** (0.844)	28.084*** (0.927)	23.678*** (0.889)	24.039*** (0.901)	23.769*** (0.953)
After	26.089*** (0.873)	23.957*** (0.919)	27.447*** (0.940)	26.210*** (0.922)	30.719*** (0.954)	21.163*** (0.954)
<i>Difference</i>	2.113*** (0.347)	3.674*** (0.484)	-0.637 (0.494)	2.532*** (0.470)	6.680*** (0.526)	-2.606*** (0.603)
Observations	6,502	1,342	1,342	1,342	1,342	1,134
Subjects	671	671	671	671	671	592

Notes: Mean reservation prices, estimated by interval regression. Standard errors in parentheses, clustered by subject.

Column 1 of Table 4 displays mean reservation prices for consuming insects. Before subjects receive the food items, the mean reservation price is \$23.98 (averaged across species and conditions). This number increases to \$26.09 once subjects receive the insects they are about to eat ($p < 0.01$). I obtain these estimates through interval regression (a generalization of the Tobit model) which accounts for both the interval-coding and the censoring in the reservation price data. Reservation prices are also widely dispersed across individuals. Population standard deviations exceed \$20 for each species both before and after the handout of the insects.

Columns 2 to 6 perform the same analysis separately by species. Mean reservation prices vary substantially across the insect species, from just over \$20 to just over \$30. Differences in reservation prices across stages 3 and 5 also vary across species, as one would expect if handing out the insects conveys substantial information. Specifically, handing out the insects increases mean reservation price for mole crickets by \$6.68 ($p < 0.01$), for house crickets by \$3.67 ($p < 0.01$), and for silkworm pupae by \$2.53 ($p < 0.01$). By contrast, reservation prices for field crickets decrease by \$2.60 ($p < 0.01$), and increase by a statistically insignificant amount of \$0.63 in the case of mealworms ($p > 0.1$). Moreover, the mean absolute change in valuations is \$6.7, again consistent with substantial information provision.

Information choice and participation decisions Next, I study how the incentive amount affects information choice. Panel A in Figure 7 displays the fraction of subjects opting for the encouraging video (left hand side) and the frequency with which the pro and con-clips are selected (right hand side), both as a function of the incentive amount. The increase in the incentive raises the proportion of subjects who select the encouraging video from 81.2% to 88.7%. The 7.5 percentage-points difference is statistically significant at the 5% level. The increase in the incentive changes preferences for the video clips in the same direction, raising the number of pro clips chosen by 8.6% ($p = 0.07$) and lowering the number of con clips by 21.9% ($p < 0.01$). The chosen number of clips labeled “other” remains approximately unchanged (1.08 and 1.07 in

the \$3 and \$30 incentive conditions, respectively). Incentives bias information acquisition in a way consistent with the behavioral part of the Undue Inducement Hypothesis.

Figure 7: Information choice and supply in Experiment 2

Note: Panel A shows information choice. The left half shows the fraction of subjects who choose the encouraging video. The right half shows average number of clips labeled ‘pro’ and ‘con,’ respectively. Panel B shows the fraction of subjects willing to eat the food item for the promised incentive, averaged across species, by incentive and video condition. Whiskers display 95% confidence intervals. Standard errors clustered by subject.

Panel B displays the effect of the increase in the incentive on participation. In the Video condition, participation increases from 38% to 71%, pooled across species. In the No Video condition, participation at the low incentive is a highly similar 36%, but the high incentive raises participation only to 59%. The 10.7 percentage points difference in these effects is statistically significant at the 10% level. The fact that the Video condition has a behavioral effect shows that subjects do not use the videos merely as a way to ex-post rationalize a decision they would have made anyway. While this treatment comparison cannot disentangle the effect of information biasing from the effect of the information provision per se, it does show that the opportunity to select and watch one of the videos makes individuals more eager to accept the transaction at the high but not at the low incentive.⁶⁶

Overall, just as in Experiment 1, these findings are consistent with the behavioral part of the Undue Inducement Hypothesis. Again, they leave open the question whether higher incentives decrease welfare. I address this question next.

Welfare Next, I study whether high incentives may harm participants in the sense of Proposition 3, using ex-post welfare analysis. I measure each subject’s welfare by Marshallian surplus. Specifically, consider a subject who agrees to eat a food item for the promised incentive m_i in Stage 4 of the experiment. After that decision, the subject receives and inspects the food item, and reveals a reservation price p_i . This subject’s

⁶⁶The 10.7 percentage points effect on participation exceeds the 7.5 percentage points effect of incentives on video choice. Hence, video choice alone cannot explain the participation effect. Other explanations include the possibility that the incentive amount also affects how subjects interpret a given video, and the possibility that information provision per se raises participation in a way that depends on the incentive amount.

Figure 8: Welfare in Experiment 2

Note: Panel A displays CDFs of ex post Marshallian surplus by incentive condition, using data from the Video condition. The CDFs start at a positive value on the vertical axis. The height reflects the frequency of left-censored observations. Panel B displays mean welfare gains and losses by incentive using data from the Video condition. Panel C corresponds to Panel A with the distribution of reservation prices imputed from the distribution of reservation prices in the (\$3, No Video)-treatment. Panel D displays the corresponding mean welfare gains and losses.

ex-post Marshallian surplus equals $w_i = m_i - p_i$. If, instead, the subject refused the transaction in Stage 4, her Marshallian surplus is 0 regardless of the reservation price she reveals later. From the ex-post viewpoint, a subject's decision to participate is mistaken if $w_i < 0$, and w_i is the welfare cost of that mistake to the subject.⁶⁷

Panel A of Figure 8 displays the cumulative distribution functions of surplus w_i in the Video condition separately by incentive condition, pooling across insect species. Unsurprisingly, a higher incentive greatly increases the gains of those who benefit from the transaction. Turning to negative surplus, I find that around a fifth of choices yield negative surplus in the \$3 condition. This number drops to just under 10% for the \$30 condition. Moreover, there is no indication that, conditional on experiencing a welfare loss, the magnitude of the loss is increasing in the incentive. This qualitative information alone suggests that the Undue Inducement Condition (4) is violated. Panel B displays mean gains and mean losses, $g(m) = \frac{1}{n} \sum_{i=1}^n \max(w_i(m), 0)$, and $l(m) = \frac{1}{n} \sum_{i=1}^n \min(w_i(m), 0)$, respectively, by incentive amount. I account for the fact that elicited reservation prices are right-censored at \$60 by fitting a lognormal distribution to the distribution of reservation prices that exceed the median, and then replacing the censored observations with the implied expected value.⁶⁸ Visual inspection reveals that the functions $g(m)$ and $l(m)$ intersect below the horizontal, indicating a violation of the Undue Inducement Condition.

Figure 8 pools across species. Panel A of Table 5 formally tests the Undue Inducement Condition (4) separately for each species.⁶⁹ I estimate the parameters of system (3) using seemingly unrelated regression. As in Experiment 1, it is important to note that this procedure produces potentially biased estimates if reservation prices are elicited with noise. While the current exposition ignores this issue for simplicity, I address it in Appendix C.2 where I obtain qualitatively similar findings.⁷⁰

Columns 1 and 2 show mean welfare gains and losses at the \$3 incentive, denoted $\tilde{\beta}_0^G$, and $\tilde{\beta}_0^L$, respectively. Columns 3 and 4 show the estimated slope coefficients, β_1^G , and β_1^L . Column 5 shows the lowest weight, $\underline{\alpha}$ that needs to be placed on those who are harmed by the transaction so an increase in the incentive is deemed to decrease welfare. Mechanically, $\underline{\alpha}$ is smaller than 1 only if $\beta_1^L < 0$ (mole crickets); in all other cases, the Undue Inducement Condition is not satisfied because higher incentives do not increase the magnitude of the losses of those who are harmed. Moreover, at the \$3 incentive, mean losses exceed mean gains by a factor of three or more. This fact greatly limits the highest welfare weight $\bar{\alpha}$ that may be placed on those who lose from the transaction while still keeping it admissible at the \$3 incentive (column 6). In the case of mole crickets, this upper bound is an order of magnitude below the lower bound, thus violating the Undue Inducement Condition (4). It is never the case, for any species in this experiment, and for any welfare weight

⁶⁷Because subjects make participation decisions based on incomplete information, an ex-post error does not imply that the subject's decision was mistaken ex ante.

⁶⁸I use a constant-only Tobit model to obtain estimates of the distributional parameters μ and σ , denoted $\hat{\mu}$ and $\hat{\sigma}$, separately for the low and high incentive condition. The dependent variable is the logarithm of subjects' reservation prices for observations in which the bet was accepted. I only use reservation prices that exceed the median because I aim to predict high rather than low reservation prices. The predicted mean reservation prices for censored observations are \$108 and \$141 in the low and high incentive conditions, respectively.

⁶⁹I impute welfare for censored observations as above, separately for each species.

⁷⁰Appendix C.2 addresses the issue by averaging out elicitation noise. For each species r , I predict a subject's reservation price using her choices for all remaining species using ridge regression. I then perform welfare analysis using the predicted values.

Table 5: Welfare effects by species in Experiment 2

Variable	(1) β_0^G	(2) β_0^L	(3) β_1^G	(4) β_1^L	(5) $\underline{\alpha}$	(6) $\bar{\alpha}$	(7) Condition (4)	(8)
							satisfied	p -value
A. Within-treatment benchmark								
2 house crickets	0.310*** (0.056)	-1.914** (0.821)	0.492*** (0.031)	0.008 (0.042)	1.016*** (0.088)	0.139*** (0.057)	No	0.000
5 superworms	0.254*** (0.050)	-1.011** (0.495)	0.375*** (0.032)	0.025 (0.019)	1.073*** (0.059)	0.201*** (0.086)	No	0.000
3 silkworm pupae	0.269*** (0.051)	-2.691*** (0.903)	0.444*** (0.031)	0.013 (0.047)	1.031*** (0.111)	0.091*** (0.033)	No	0.000
2 mole crickets	0.239*** (0.050)	-4.554*** (1.313)	0.431*** (0.031)	-0.172* (0.090)	0.714*** (0.111)	0.050*** (0.018)	No	0.000
2 field crickets	0.305*** (0.056)	-2.481** (0.999)	0.326*** (0.031)	0.054 (0.042)	1.199*** (0.187)	0.109*** (0.044)	No	0.000
B. No Video, \$3 incentive benchmark								
2 house crickets	0.363*** (0.060)	-2.218*** (0.860)	0.466*** (0.032)	-0.104* (0.060)	0.817*** (0.090)	0.141*** (0.052)	No	0.000
5 superworms	0.312*** (0.058)	-1.571** (0.611)	0.360*** (0.032)	-0.064 (0.046)	0.849*** (0.094)	0.166*** (0.061)	No	0.000
3 silkworm pupae	0.327*** (0.058)	-3.441*** (1.045)	0.398*** (0.032)	-0.249*** (0.091)	0.615*** (0.093)	0.087*** (0.029)	No	0.000
2 mole crickets	0.266*** (0.053)	-5.659*** (1.539)	0.346*** (0.031)	-0.965*** (0.177)	0.264*** (0.045)	0.045*** (0.015)	No	0.000
2 field crickets	0.371*** (0.062)	-0.668*** (0.216)	0.414*** (0.034)	0.025*** (0.008)	1.064*** (0.023)	0.357*** (0.087)	No	0.000

Notes: Each row corresponds to a separate SUR-regression using data from the Video condition. Columns 1 and 2 show the level of mean welfare gains and losses, respectively, at the \$3 incentive. Columns 3 and 4 show the change in mean welfare gains and losses, respectively as the incentive increases by \$1. Column 5 shows the least weight α that needs to be placed on those who lose from the transaction such that an increase in the incentive is deemed welfare-decreasing. Column 6 shows the maximal weight α that may be placed on those who lose from the transaction such that the transaction is admissible at sufficiently low incentives. Reported significance levels in columns 5 and 6 reflect the tests that $\underline{\alpha} = 0$ and $\bar{\alpha} = 1$, respectively. Column 7 indicates whether $T = \bar{\alpha} - \underline{\alpha}$ is nonnegative. Column 8 lists p -values for two sided tests of the Null hypothesis that $T = 0$. Standard errors in parentheses, clustered by subject. Standard errors in columns 5, 6, and 8 are calculated by the Delta method.

α , that high but not low incentives are admissible. The experiment refutes the normative part of the Undue Inducement Hypothesis.

I next address three potential concerns about the analysis just presented. First, the fact that subjects are exposed to an incentive amount before stating their reservation prices potentially causes anchoring effects (see, e.g., Ariely et al., 2003). Indeed, the high incentive increases reservation prices measured after the handout of the insects by \$4.28 in the No Video condition (though it decreases them by \$2.01 in the Video condition, and neither of these estimates is statistically significantly different from zero at the 10%-level). Moreover, subjects' video choice possibly alters subjects' reservation prices in a way one may deem normatively invalid, for instance, because they may be perceived as tendentious. To address these concerns, I

repeat the above welfare analysis under the assumption that true reservation prices in each treatment are distributed in the same way as in the (\$3, No Video)-treatment. I use the \$3 condition as benchmark because the undue inducement literature contends that choices made under low, but not under high incentives are normatively valid.⁷¹ I implement the counterfactual by ranking reservation prices, separately within each treatment condition and for each of the five insect species. For each species and within each treatment, I then replace the reservation price observed at percentile p with the reservation price observed at the same percentile for that species in the (\$3, No Video)-treatment. Finally, within each treatment, I calculate Marshallian surplus by combining these counterfactual reservation prices with subjects' actual participation decisions.

Panel C of Figure 8 displays the resulting CDFs. Welfare losses in the \$30-incentive case are now larger, both in contrast to the \$3-incentive condition in the same figure, and in contrast to the welfare losses in the \$30-incentive condition in Panel A. Moreover, compared to Panel A, the \$30-incentive condition leads to slightly smaller welfare gains for subjects who benefit from the transaction. To test the Undue Inducement Condition for this case, Panel D presents mean gains and mean losses for each incentive amount. As before, I impute the magnitude of welfare losses for observations with censored reservation prices using a lognormal fit. I find that the magnitude of welfare losses significantly increases in the incentive amount, although only by about half the extent to which gains increase. Moreover, because welfare losses far outweigh welfare gains at the \$3-incentive, the Undue Inducement Condition (4) is still violated when pooling across species. Panel B of Table 5 performs the formal econometric tests of the Undue Inducement Condition separately for each species. While the magnitude of welfare losses increases in the incentive for all but one case, the Undue Inducement Condition (4) remains violated for each species.

The second potential concern is that subjects in this experiment have no opportunity to express whether they derive positive utility from eating insects, and if so, by how much. This feature necessarily limits measured welfare gains. Considering non-monetary benefits from the transaction is especially important in light of the position, common in the undue inducement literature, that individuals should be allowed to participate in transactions only if they are intrinsically motivated to participate (see, e.g., [London, 2005](#)).⁷² To address this point, I estimate the smallest utility benefit that subjects would need to derive from eating insects such that the Undue Inducement Condition (4) is satisfied. Such benefits can only apply to the 4.3 percent of observations in which the subject agreed to eat the insects for free. For the remaining observations, reservation prices reveal that eating the insects causes a utility cost to the subject. In terms of equation (3), incorporating utility benefits corresponds to increasing the intercept $\tilde{\beta}_0^G$ while leaving the remaining parameters unchanged.⁷³ As long as the magnitude of welfare losses decreases with the incentive, as is the case in Panel B of Figure 8, no amount of utility benefit from participation can satisfy the Undue Inducement Condition (4). The conclusion differs in case of the counterfactual welfare benchmark used in Panel D of Figure 8. In this case, simple algebra shows that the Undue Inducement Condition is satisfied if $\tilde{\beta}_0^G$ rises

⁷¹I obtain the same results (uniform violation of the Undue Inducement Condition) regardless of the choice of benchmark.

⁷²This normative position contrasts with empirical reality. Financial reward is the primary motivation, for instance, for healthy volunteers to participate in clinical trials ([Stunkel and Grady, 2011](#)).

⁷³This statement assumes that subjects who derive positive utility from bug-eating are not disproportionately more likely to participate at the low rather than high incentive.

by at least $\Delta = \tilde{\beta}_0^L \frac{\beta_1^G}{\beta_1^L} - \tilde{\beta}_0^G$. Based on the estimated parameter values, I find that $\Delta = 4.77$. Accordingly, the 4.3% of subjects who reveal a reservation price of zero would need to derive an average utility from bug-eating of at least $\$4.77/0.043 = \111 for there to exist a welfare weight α under which low but not high incentives are admissible.

The third potential concern relates to the fact that the welfare benchmark used in this experiment is given by the reservation prices subjects reveal after having received and inspected the insects. While these reservation prices embody substantial additional information, they are not measures of the actual experience subjects have when eating the insects. Hence, they measure ex post welfare only imperfectly. This issue is inherent in the ‘informed consumer paradigm’ (Bernheim and Taubinsky, 2018) on which this experiment is based—informed consumers such as pharmacists (Bronnenberg et al., 2015) or nutritionists (Allcott et al., 2019) are better informed than non-experts, but their information about the ex-post welfare from consumption goods related to their expertise is generally still imperfect. In order for such issues to overturn the findings of this experiment, true ex-post welfare would need to deviate from subjects’ reservation prices in ways that differ systematically across the treatment conditions.⁷⁴

Overall, this experiment replicates the conclusions from Experiment 1. Higher incentives cause the type of information biasing that is consistent with the behavioral part of the Undue Inducement Hypothesis, but the normative implications of that hypothesis are soundly refuted. Experiment 2 yields this result even though it incorporates the difficulty of evaluating a visceral experience in terms of dollars and cents, at substantial affect.

5 Conclusion

The vague and empirically largely untested notion of undue inducement is a key reason for laws and regulations around the world that limit the incentives that can be offered for transactions such as human research participation, female egg donation, organ donation, or gestational surrogacy. The Undue Inducement Hypothesis claims, first, that participation incentives lead to biased information processing, and second, that these effects of incentives are harmful in a way that justifies capping incentives for otherwise admissible transactions.

This paper experimentally tests the Undue Inducement Hypothesis, aided by a formal conceptual framework. I show that incentive-induced information biasing is consistent with Bayesian rationality, and as such is not evidence for harm. I conduct formal welfare analysis in a setting that allows for the possibility of mistakes, focusing on the class of welfare functions that vary the weight on subjects who are harmed by the transaction. I show that a transaction should be permitted at capped incentives only if two conditions are satisfied. First, at low incentives, the gains to those who benefit from the transaction must outweigh the losses of those who are harmed. Second, as the incentive increases, the losses of those who are harmed must begin to dominate the gains of those who benefit. I summarize these criteria in a single inequality called the Undue Inducement Condition, which I test in two experiments.

⁷⁴Eliciting true ex-post welfare in an incentive-compatible way is impossible in principle, because any incentivized elicitation will require a subject to decide between various future outcomes.

In both experiments, incentives lead to biased information acquisition which causally affects participation decisions. While this finding resonates with undue inducement concerns, it is also consistent with the Bayes-rational acquisition of costly information. Direct welfare measurement shows that the Undue Inducement Condition robustly violated in both experiments. There is no welfare weight such that the transaction should be permitted but incentives should be capped. This result obtains if subjects face the task of evaluating a highly visceral transaction in terms of dollars and cents, if potential harms from participation occur with a three months delay, and across a variety of prior success probabilities. They arise both in an ex-post welfare framework and in an ex-ante framework (where feasible). They obtain according to both empirical paradigms that dominate behavioral welfare economics, the ‘reframed decisions paradigm’ and the ‘informed consumer paradigm’ (Bernheim and Taubinsky, 2018). Overall, in my experiments, incentives do not cause deterioration in decision quality that would justify capping incentives for otherwise admissible transactions.

There are two classes of issues that I leave to future research. The first consists in an exploration of the robustness of the findings of this study in other contexts and with other subject pools. While university students—the subjects of the present paper—are frequent participants in some transactions subject to undue inducement concerns such as human egg donation, social science experiments, and clinical trials, it is important to examine the Undue Inducement Hypothesis with demographics who are, or would be, participants in other transactions subject to undue inducement concerns such as incentivized organ donation or gestational surrogacy. It is also important to test the Undue Inducement Hypothesis using incentives that are large in absolute terms. While the current experiments generate information-biasing and other behaviors that have been taken as *prima facie* evidence for undue inducement, they cannot exclude the possibility that incentives on the order of several monthly or yearly wages create separate types of cognitive distortions through which they may cause harm.

The second class of issues concerns questions of distribution and justice. A common response to proposals of incentives for clinical trials, kidney donation, and similar transactions is that society should not tolerate levels of inequality so high that anybody would want to engage in these transactions in exchange for money. While this concern is orthogonal to the Undue Inducement Hypothesis (which applies even to hypothetical societies without inequality), it is conceivable that allowing previously prevented transactions such as selling kidneys in exchange for thousands of dollars might affect support for redistribution and social assistance. Whether this is the case and in what direction the effects would point remains an interesting open question.

Most broadly, this paper bridges a gap between economics on the one hand, and the applied ethics and policy literatures on the other. Through standard economic methodology, this paper informs a prominent and highly influential, but vaguely formulated and largely untested hypothesis about the effects of incentives. Generally, using the powerful toolbox of economics to examine moral intuitions and the behavioral assumptions on which they rely is an important direction for future research, for otherwise, laws will be based on intuition and speculation alone.

References

- Ager, Philipp, Leonardo Bursztyn, Lukas Leucht, and Hans-Joachim Voth, “Killer incentives: Status competition and pilot performance during world war II,” *NBER working paper*, 2019.
- Allcott, Hunt and Dmitry Taubinsky, “Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market,” *American Economic Review*, 2015, 105 (8), 2501–38.
- and Judd B Kessler, “The welfare effects of nudges: A case study of energy use social comparisons,” *American Economic Journal: Applied Economics*, 2019, 11 (1), 236–76.
- , Benjamin B Lockwood, and Dmitry Taubinsky, “Regressive sin taxes, with an application to the optimal soda tax,” *The Quarterly Journal of Economics*, 2019, 134 (3), 1557–1626.
- Ambuehl, Sandro and Axel Ockenfels, “The Ethics of Incentivizing the Uninformed. A Vignette Study,” *American Economic Review, Papers & Proceedings*, 2017, 107 (5), 91–95.
- , —, and Alvin E Roth, “Payment in challenge studies from an economics perspective,” *Journal of Medical Ethics*, 2020, 46 (12), 831–832.
- , —, and Colin Stewart, “Who Opts In?,” *unpublished*, 2020.
- , B Douglas Bernheim, and Annamaria Lusardi, “Evaluating Deliberative Competence: A Simple Method with an Application to Financial Choice,” *CEPR Discussion Paper No. DP15863*, 2021.
- , Muriel Niederle, and Alvin E. Roth, “More Money, More Problems? Can High Pay be Coercive and Repugnant?,” *American Economic Review, Papers & Proceedings*, 2015, 105 (5), 357–60.
- American Society for Reproductive Medicine, “Financial compensation of oocyte donors,” *Fertility and Sterility*, 2007, 88 (2), 305–309.
- Ariely, Dan, George Loewenstein, and Drazen Prelec, “‘Coherent Arbitrariness’: Stable Demand Curves Without Stable Preferences,” *The Quarterly Journal of Economics*, 2003, 118 (1), 73–105.
- , Uri Gneezy, George Loewenstein, and Nina Mazar, “Large stakes and big mistakes,” *The Review of Economic Studies*, 2009, 76 (2), 451–469.
- Babcock, Linda, George Loewenstein, Samuel Issacharoff, and Colin Camerer, “Biased judgments of fairness in bargaining,” *The American Economic Review*, 1995, pp. 1337–1343.
- Basu, Kaushik, “The economics and law of sexual harassment in the workplace,” *The Journal of Economic Perspectives*, 2003, 17 (3), 141–157.
- , “Coercion, Contract and the Limits of the Market,” *Social Choice and Welfare*, 2007, 29 (4), 559–579.
- Beeson, Diane and Abby Lippman, “Egg harvesting for stem cell research: medical risks and ethical problems,” *Reproductive BioMedicine Online*, 2006, 13 (4), 573–579.
- Bénabou, Roland, “The Economics of Motivated Beliefs,” *Revue d’économie politique*, 2015, 125 (5), 665–685.
- and Jean Tirole, “Mindful Economics: The Production, Consumption, and Value of Beliefs,” *The Journal of Economic Perspectives*, 2016, 30 (3), 141–164.
- Benjamin, Daniel J, “Errors in probabilistic reasoning and judgment biases,” *Handbook of Behavioral Economics: Applications and Foundations 1*, 2019, 2, 69–186.
- Bentley, John P and Paul G Thacker, “The influence of risk and monetary payment on the research participation decision making process,” *Journal of medical ethics*, 2004, 30 (3), 293–298.
- Bernheim, B. Douglas and Antonio Rangel, “Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics,” *The Quarterly Journal of Economics*, 2009, 124 (1), 51–104.
- Bernheim, B Douglas and Dmitry Taubinsky, *Behavioral public economics*, Vol. 1, Elsevier,
- Bodri, Daniel, “Risk and complications associated with egg donation,” in “Principles of Oocyte and Embryo Donation,” Springer, 2013, pp. 205–219.
- Bronnenberg, Bart J, Jean-Pierre Dubé, Matthew Gentzkow, and Jesse M Shapiro, “Do pharmacists buy Bayer? Informed shoppers and the brand premium,” *The Quarterly Journal of Economics*, 2015, 130 (4), 1669–1726.
- Bundesen, Claus and Lisbeth Harms, “Single-letter recognition as a function of exposure duration,” *Psychological Research*, 1999, 62 (4), 275–279.
- Burrows, Paul, “Analyzing legal paternalism,” *International Review of Law and Economics*, 1995, 15 (4), 489–508.

- Charness, Gary, Ryan Oprea, and Sevgi Yuksel**, “How Do People Choose Between Biased Information Sources? Evidence from a Laboratory Experiment,” *Journal of the European Economic Association*, forthcoming.
- Chetty, Raj**, “Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods,” *Annual Review of Economics*, 2009, 1 (1), 451–488.
- , **Adam Looney, and Kory Kroft**, “Salience and Taxation: Theory and Evidence,” *American Economic Review*, 2009, 99 (4), 1145–177.
- Clemens, Michael A**, “Testing for repugnance in economic transactions: Evidence from guest work in the Gulf,” *The Journal of Legal Studies*, 2018, 47 (1), 5–44.
- Cryder, Cynthia E., Alex John London, Kevin G. Volpp, and George Loewenstein**, “Informative inducement: Study payment as a signal of risk,” *Social Science & Medicine*, 2010, 70, 455–464.
- Dana, Jason, Roberto A. Weber, and Jason Xi Kuang**, “Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness,” *Economic Theory*, 2007, 33, 67–80.
- Dean, Mark and Nathaniel Neligh**, “Experimental Tests of Rational Inattention,” *unpublished*, 2017.
- Devine, Eric G, Megan E Waters, Megan Putnam, Caitlin Surprise, Katie O’Malley, Courtney Richambault, Rachel L Fishman, Clifford M Knapp, Elissa H Patterson, Ofra Sarid-Segal, Chris Streeter, Laurie Colanari, and Domenic A Ciraulo**, “Concealment and fabrication by experienced research subjects,” *Clinical Trials*, 2013, 10, 935–948.
- Elias, Julio J, Nicola Lacetera, and Mario Macis**, “Markets and Morals: An Experimental Survey Study,” *PloS one*, 2015, 10 (6).
- Elias, Julio, Nicola Lacetera, and Mario Macis**, “Sacred Values? The Effect of Information on Attitudes Toward Payments for Human Organs,” *American Economic Review, Papers & Proceedings*, 2015, 105 (5), 361–365.
- Elías, Julio, Nicola Lacetera, and Mario Macis**, “Paying for Kidneys? A Randomized Survey and Choice Experiment,” *American Economic Review*, 2019, 109 (8).
- Ellingsen, Tore and Magnus Johannesson**, “Does impartial deliberation breed fair behavior? An experimental test,” *Rationality and Society*, 2005, 17 (1), 116–136.
- Engelmann, Jan, Maël Lebreton, Peter Schwardmann, Joel J van der Weele, and Li-Ang Chang**, “Anticipatory anxiety and wishful thinking,” 2019.
- Epley, Nicholas and Thomas Gilovich**, “The Mechanics of Motivated Reasoning,” *Journal of Economic Perspectives*, 2016, 30 (3), 133–40.
- Erkut, Hande and Dorothea Kübler**, “Repugnant transactions: The role of agency and extreme consequences,” *Unpublished Manuscript*, 2021.
- Exley, Christine and Judd B Kessler**, “Motivated Errors,” *Unpublished manuscript*, 2018.
- Exley, Christine L**, “Excusing selfishness in charitable giving: The role of risk,” *The Review of Economic Studies*, 2015.
- Eyal, Nir**, *Informed Consent*, Stanford Encyclopedia of Philosophy, 2012.
- Faden, Ruth R and Tom L Beauchamp**, *A history and theory of informed consent*, Oxford University Press, 1986.
- Festinger, Leon**, *A theory of cognitive dissonance*, Vol. 2, Stanford university press, 1962.
- Frey, Bruno S and Reto Jegen**, “Motivation Crowding Theory,” *Journal of Economic Surveys*, 2001, 15 (5), 589–611.
- Gentzkow, Matthew, Jesse M Shapiro, and Daniel F Stone**, “Media bias in the marketplace: Theory,” in “Handbook of media economics,” Vol. 1, Elsevier, 2015, pp. 623–645.
- Gino, Francesca, Michael I Norton, and Roberto A Weber**, “Motivated Bayesians: Feeling Moral While Acting Egoistically,” *Journal of Economic Perspectives*, 2016, 30 (3), 189–212.
- Gneezy, Uri, Silvia Saccardo, Marta Serra-Garcia, and Roel van Veldhuizen**, “Motivated self-deception, identity, and unethical behavior,” *unpublished manuscript*, 2015.
- Goldman, Ruth**, “The development of the ‘prudent man’ concept in relation to pension schemes,” *Pensions: An International Journal*, 2000, 5 (3), 219–224.
- Greiner, Ben**, “Subject pool recruitment procedures: organizing experiments with ORSEE,” *Journal of the Economic Science Association*, 2015, 1 (1), 114–125.

- Halpern, Scott D, Jason HT Karlawish, David Casarett, Jesse A Berlin, and David A Asch**, “Empirical assessment of whether moderate payments are undue or unjust inducements for participation in clinical trials,” *Archives of Internal Medicine*, 2004, 164 (7), 801–803.
- Hammond, Peter J**, “Ex-ante and ex-post welfare optimality under uncertainty,” *Economica*, 1981, 48 (191), 235–250.
- Harvard College v Amory**, “26 Mass (9 Pick) 446,” 1830.
- Hippen, Benjamin, Lainie Friedman Ross, and Robert M Sade**, “Saving lives is more important than abstract moral concerns: financial incentives should be used to increase organ donation,” *The Annals of Thoracic Surgery*, 2009, 88 (4), 1053–1061.
- Hoffart, Janine and Benjamin Scheibehenne**, “Pill or bill? Influence of monetary incentives on the perceived riskiness and the ethical approval of clinical trials,” *Judgment and Decision Making*, 2019, 14 ((2)), 130–134.
- Holt, Charles A. and Angela M. Smith**, “An Update on Bayesian Updating,” *Journal of Economic Behavior and Organization*, 2009, 69, 125–134.
- Kahneman, Daniel, Jack L Knetsch, and Richard Thaler**, “Fairness as a constraint on profit seeking: Entitlements in the market,” *The American Economic Review*, 1986, 76 (4), 728–741.
- Karni, Edi**, “A Mechanism for Eliciting Probabilities,” *Econometrica*, 2009, 77 (2), 603–606.
- Klayman, Joshua**, “Varieties of confirmation bias,” *Psychology of learning and motivation*, 1995, 32, 385–418.
- **and Young-Won Ha**, “Confirmation, disconfirmation, and information in hypothesis testing,” *Psychological review*, 1987, 94 (2), 211.
- Kunda, Ziva**, “The case for motivated reasoning,” *Psychological bulletin*, 1990, 108 (3), 480.
- Largent, Emily A, Christine Grady, Franklin G Miller, and Alan Wertheimer**, “Money, coercion, and undue inducement: A survey of attitudes about payments to research participants,” *IRB*, 2012, 34 (1), 1.
- Largent, Emily, Christine Grady, Franklin G Miller, and Alan Wertheimer**, “Misconceptions about coercion and undue influence: reflections on the views of IRB members,” *Bioethics*, 2013, 27 (9), 500–507.
- Leider, Stephen and Alvin E Roth**, “Kidneys for sale: Who disapproves, and why?,” *American Journal of Transplantation*, 2010, 10 (5), 1221–1227.
- List, John A.**, “Non est disputandum de generalizability? A glimpse into the external validity trial,” *NBER working paper*, 2020.
- London, Alex John**, “Undue inducements and reasonable risks: will the dismal science lead to dismal research ethics?,” *The American Journal of Bioethics*, 2005, 5 (5), 29–32.
- McDonald, Alison M, Rosemary C Knight, Marion K Campbell, Vikki A Entwistle, Adrian M Grant, Jonathan A Cook, Diana R Elbourne, David Francis, Jo Garcia, Ian Roberts et al.**, “What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies,” *Trials*, 2006, 7 (1), 9.
- Mngadi, Kathryn T, Jerome A Singh, Leila E Mansoor, and Douglas R Wassenaar**, “Undue inducement: a case study in CAPRISA 008,” *Journal of medical ethics*, 2017, 43 (12), 824–828.
- Montanari, Giovanni and Salvatore Nunnari**, “Audi alteram partem: An experiment on selective exposure to information,” *Unpublished manuscript*, 2019.
- National Bioethics Advisory Commission**, “Ethical and policy issues in research involving human participants,” Rockville, MD 2001.
- Niederle, Muriel and Alvin E Roth**, “Philanthropically Funded Heroism Awards for Kidney Donors,” *Law & Contemporary Problems*, 2014, 77, 131–144.
- Oster, Emily, Ira Shoulson, and E Dorsey**, “Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease,” *The American Economic Review*, 2013, 103 (2), 804–830.
- Peysakhovich, Alexander and Jeffrey Naecker**, “Using methods from machine learning to evaluate behavioral models of choice under risk and ambiguity,” *Journal of Economic Behavior & Organization*, 2017, 133, 373–384.
- Raven, John C**, *Guide to the standard progressive matrices: sets A, B, C, D and E*, Pearson, 1960.
- Roth, Alvin E**, “Repugnance as a Constraint on Markets,” *Journal of Economic Perspectives*, 2007, 21 (3), 37–58.

- Rottenstreich, Yuval and Christopher K Hsee**, “Money, kisses, and electric shocks: On the affective psychology of risk,” *Psychological science*, 2001, 12 (3), 185–190.
- Saccardo, Silvia and Marta Serra-Garcia**, “Cognitive Flexibility or Moral Commitment? Evidence of Anticipated Belief Distortion,” *unpublished*, 2020.
- Sandmo, Agnar**, “Ex post welfare economics and the theory of merit goods,” *Economica*, 1983, 50 (197), 19–33.
- Satz, Debra**, *Why Some Things Should Not Be For Sale: The Moral Limits of Markets*, Oxford University Press, 2010.
- Sharot, Tali, Christoph W Korn, and Raymond J Dolan**, “How unrealistic optimism is maintained in the face of reality,” *Nature Neuroscience*, 2011, 14 (11), 1475–1481.
- Siddique, Abu Bakkar, Vandana Apte, Sigrid Fry-Revere, Yanhong Jin, and Naoru Koizumi**, “The impact of country reimbursement programmes on living kidney donations,” *BMJ Global Health*, 2020, 5 (8).
- Singer, Eleanor and Mick P Couper**, “Do incentives exert undue influence on survey participation? Experimental evidence,” *Journal of Empirical Research on Human Research Ethics*, 2008, 3 (3), 49–56.
- Slowiaczek, Louisa M, Joshua Klayman, Steven J Sherman, and Richard B Skov**, “Information selection and use in hypothesis testing: What is a good question, and what is a good answer?,” *Memory & Cognition*, 1992, 20 (4), 392–405.
- Smith, Vernon L**, “Experimental economics: Induced value theory,” *The American Economic Review*, 1976, pp. 274–279.
- Stueber, Robert**, “Why High Incentives Cause Repugnance,” *Unpublished Manuscript*, 2020.
- Stunkel, Leanne and Christine Grady**, “More than the money: a review of the literature examining healthy volunteer motivations,” *Contemporary clinical trials*, 2011, 32 (3), 342–352.
- Suen, Wing**, “The Self-Perpetuation of Biased Beliefs,” *The Economic Journal*, 2004, 114 (495), 377–396.
- Sullivan, Colin D.**, “Eliciting Preferences over Life and Death: Experimental Evidence from Organ Transplantation,” *Unpublished Manuscript*, 2020.
- Taubinsky, Dmitry and Alex Rees-Jones**, “Attention variation and welfare: theory and evidence from a tax salience experiment,” *The Review of Economic Studies*, 2018, 85 (4), 2462–2496.
- Tishler, Carl L and Suzanne Bartholomae**, “The recruitment of normal healthy volunteers: a review of the literature on the use of financial incentives,” *The Journal of Clinical Pharmacology*, 2002, 42 (4), 365–375.
- Titmuss, Richard**, *The gift relationship: From human blood to social policy*, London: Allen & Unwin, 1971.
- Toplak, Maggie E, Richard F West, and Keith E Stanovich**, “Assessing miserly information processing: An expansion of the Cognitive Reflection Test,” *Thinking & Reasoning*, 2014, 20 (2), 147–168.
- Tversky, Amos and Daniel Kahneman**, “Judgment under uncertainty: Heuristics and biases,” *Science*, 1974, 185 (4157), 1124–1131.
- US Department of Health, Education, and Welfare (DHEW)**, *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*, National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, Washington DC, 1978.
- Wald, Abraham**, *Sequential Analysis*, New York: Wiley, 1947.
- Weinstein, Neil D**, “Unrealistic optimism about future life events,” *Journal of personality and social psychology*, 1980, 39 (5), 806.
- , “Optimistic biases about personal risks,” *Science*, 1989, 246 (4935), 1232–1234.

ONLINE-APPENDIX

Can Incentives Cause Harm?

Testing Undue Inducement

Sandro Ambuehl

Table of Contents

A	Model: Additional Materials and Proofs	1
A.1	Proof of proposition 1	1
A.2	Proof of proposition 2	2
A.3	Proof of corollary 1	2
B	Experiment 1: Additional Analysis	3
B.1	Information biasing and participation by treatment	3
B.2	Welfare analysis with certainty equivalents predicted by ridge-regression	6
B.3	Welfare analysis, ex-post benchmark	8
C	Experiment 2: Additional Analysis	10
C.1	Randomization check	10
C.2	Welfare analysis accounting for noisy elicitation	11
D	Experiment 3: Sequential Information Search	13
D.1	Design	13
D.2	Analysis	15
E	Experimental Materials	20
E.1	Video transcriptions	20
E.2	Instructions for Experiment 1	22
E.3	Instructions for Experiment 2	43
	References	51

A Model: Additional Materials and Proofs

A.1 Proof of proposition 1

To prove the claim, I parametrize the constraint and insert it into the objective function to characterize the optimal information structure. I then show how a change in m changes the optimal information structure.

The objective function (1) is linear in (p_G, p_B) , with indifference curves that are upward-sloping in (p_G, p_B) -space. Accordingly, the only candidates for optimal information structures are those that lie on the boundary of the free-disposal convex hull of the set of information structures. The latter set, denoted H , is defined as follows. A point (p_G, p_B) is in H if and only if there exists a subset of information structures $(I_1, \dots, I_n) \subseteq \tilde{\mathcal{J}}$ with weights $a_1, \dots, a_n$ such that $\sum_{k=1}^n a_k p_{I_k}^G \geq p_G$ and $\sum_{k=1}^n a_k (1 - p_{I_k}^B) \geq (1 - p_B)$.

Let $\mathcal{J} \subseteq \tilde{\mathcal{J}}$ denote the subset of information structures on the boundary of H ; let n denote the cardinality of $\mathcal{J}$. Enumerate these information structures such that $p_{I_1}^G \leq p_{I_2}^G \leq \dots \leq p_{I_n}^G$. Because H is convex, the boundary of H is convex. Formally, $\frac{p_{I_k}^B - p_{I_{k-1}}^B}{p_{I_k}^G - p_{I_{k-1}}^G}$ is an increasing function of k for $2 \leq k \leq n$.

The boundary of H implicitly defines p_B as a function of p_G . This function, denoted $b(p_G)$, is piecewise linear and continuous. Accordingly, we can write the objective function in (1) as $u(p_G; m) = \mu p_H u_G(m) + (1 - \mu) b(p_G) u_B(m)$. A composition of a finite number of continuous, piecewise linear functions, $u(p_G; m)$ is continuous and piecewise linear. For each $k = 2, \dots, n$ the slope of $u(p_G; m)$ as a function of p_G over the interval $(p_{I_{k-1}}^G, p_{I_k}^G)$ is given by

$$u'(p_G; m) = \mu u_G(m) + (1 - \mu) \frac{p_{I_k}^B - p_{I_{k-1}}^B}{p_{I_k}^G - p_{I_{k-1}}^G} u_B(m)$$

Because $u_B(m) < 0$ and because $\frac{p_{I_k}^B - p_{I_{k-1}}^B}{p_{I_k}^G - p_{I_{k-1}}^G}$ is increasing, $u'(p_G)$ is decreasing.

Let k^* denote the optimal information structure. If $u'(p_G; m) > 0$ for all $p_G \in (p_{I_1}^G, p_{I_n}^G)$, the optimal information structure is I_n , hence $k^* = n$. If $u'(p_G; m) < 0$ for all $p_G \in (p_{I_1}^G, p_{I_n}^G)$, the optimal information structure is I_1 , hence $k^* = 1$. Otherwise, there exists $\hat{k} = 2, \dots, n - 1$ such that $u'(p_G; m)$ is weakly positive to the left of $p_{I_{\hat{k}}}^G$ and weakly negative to the right. In that case, $\hat{k}$ is within the set of optimal information structures. If both the foregoing inequalities are strict, then $k^* = \hat{k}$. If $u'(p_G, m)$ is zero to the left of $p_{I_{\hat{k}}}^G$, then the set of optimal information structures includes $\hat{k} - 1$. If $u'(p_G, m)$ is zero to the right of $p_{I_{\hat{k}}}^G$, then the set of optimal information structures includes $\hat{k} + 1$.

How does an increase in m affect the optimal information structure? Consider m' which, as assumed in the proposition satisfies $m' > m$. Define $k^*(m')$ as the index k such that $u'(p_G; m')$ is positive to the left of $p_{I_{k^*(m')}}^G$ and negative to the right (if such a value exists); otherwise set $k^*(m') = 1$ if $u'(p_G; m')$ is uniformly negative and set $k^*(m') = n$ if $u'(p_G; m')$ is uniformly positive. Then, $k^*(m')$ is the optimal information structure at incentive m' . Because $u_G(m)$ and $u_B(m)$ are both increasing in m , an increase in m to $m' > m$ increases $u'(p_G; m)$. Accordingly, $k^*(m') \geq k^*(m)$.¹

¹Because $k^*(m')$ and $k^*(m)$ may be sets, the appropriate order for this comparison is the strong set order.

Finally, the assumption that both m and m' are associated with at least one information structure that is instrumental implies that neither $k^*(m')$ nor $k^*(m)$ are equal to the unconditional participation or abstention information structures $(0, 0)$ or $(1, 1)$. This completes the proof.

A.2 Proof of proposition 2

This result follows directly from the definition of biased information and the assumption that $\mathcal{J}_r$ contains instrumental information structures. Instrumentality requires that the subject participates if and only if the information structure produces signal $\mathcal{G}$. Because for both m and m' there is at least one information structure in $\mathcal{J}_r$ that is instrumental at the respective incentive, neither unconditional participation nor unconditional abstention are the optimal choice given the restricted set of information structures. Therefore, these strategies are no optimal response for the original set of informations structures i^* either. Next, consider a high incentive. By construction, the information structure I_r^* that is optimal within the restricted set $\mathcal{J}_r$ is (weakly) less biased towards participation than the optimal information structure I^* in the unrestricted set $\mathcal{J}$. Because both I_r^* and I^* are instrumental, and because I_r^* is less biased towards participation than I^* , the agent less likely participates when choosing from the restricted set, $\mathcal{J}_r$. A parallel argument shows that if the incentive is low, the agent (weakly) more likely participates when choosing from the restricted set.

A.3 Proof of corollary 1

The statement $p(m) - p(m') > 0$ follows immediately by Proposition 1. Hence, it suffices to show that $q(m) - q(m') = 0$. By construction, the agent cannot condition the selection of the information structure on the incentive amount. By assumption, both information structures are instrumental at both incentive amounts, so that the agent participates if and only if the selected information structure produces signal $\mathcal{G}$. Accordingly, the participation probability is independent of the realization of the incentive, as needed.

B Experiment 1: Additional Analysis

B.1 Information biasing and participation by treatment

This subsection examines the behavioral effects of incentives in Experiment 1 separately by treatment.

Table B.1 displays the estimated coefficient of OLS regressions of an indicator for the choice of the Bold Advisor on the incentive amount, an indicator for the Incentive First condition, and the interaction between the two, clustering standard errors by subject. Column 1 disaggregates the estimates by prior condition, but keeps pooling across the Compensability and Delay conditions. I find a statistically significantly positive interaction coefficient for each prior condition ($p < 0.01$ in each case), with larger effects for higher prior success probabilities.² Columns 2 - 5 further disaggregate the estimations across the Compensability and Delay conditions. Higher incentives increase the frequency with which the Bold Advisor is chosen in each column for each prior, although not always to a statistically significant extent. Two patterns are notable. First, the effects of incentives on information choice are larger if consequences are compensable in all cases but one, as a comparison of the coefficients in columns 2 and 4 to those in columns 3 and 5 shows.³ Second, perhaps surprisingly, while the Delay condition leads to slightly larger effects if consequences are compensable (compare columns 2 and 4), no similar effect occurs in the case of non-compensable consequences (compare columns 3 and 5).⁴

Table B.2 performs a parallel analysis using the decision to participate in the transaction as dependent variable. In Column 1, which pools across Compensability and Delay conditions, there are significantly positive interaction terms for both the low and middle prior failure probabilities. In both cases, their magnitude is roughly half the magnitude of the corresponding effect on information choice, as expected. For the high prior failure probability I find no effect, possibly due to the fact that the overwhelming majority of subjects decline to participate in the transaction in that condition. Further disaggregation across the Compensability and Consequence Timing conditions suffers from the fact that each cell now contains a comparatively small number of observations, leading to large standard errors. Accordingly, no clear patterns emerge, with the exception that the estimated coefficients are uniformly larger in the 20% and 50% prior failure probability conditions than in the 80% prior failure probability condition.

²Also consistent with the formal model is the finding that higher prior success probabilities are associated with a more frequent choice of the Bold Advisor in the Advisor Choice First condition, as a comparison of the regression constants across the prior conditions shows.

³This effect is consistent with the model in Section 2 for the following reason. In the compensable condition, a higher incentive both increases the costs of false negatives and decreases the cost of false positives. In the non-compensable condition, the latter mechanism is shut off. Accordingly, in the language used in Figure 1, a given increase in the incentive causes indifference curves to tilt to a greater extent in the Compensable condition than in the Non-compensable condition, and therefore causes a greater change in the optimal information structure.

⁴The absence of an effect of the Delay treatment is unexpected to the extent that delayed consequences will be discounted by the subjects' discount factor. Accordingly, the Delay treatment corresponds to a slight increase in the incentive amount. This effect is small, however, because the average discount factor of 0.83 is close to 1.

Table B.1: Information choice in Experiment 1 by treatment condition

VARIABLES	(1)	(2)	(3)	(4)	(5)
	Bold Advisor chosen				
<i>Consequences</i>					
Compensable	Both	Yes	No	Yes	No
Delayed	Both	No	No	Yes	Yes
Prior failure probability: 20%					
Incentive in €	-0.007 (0.008)	0.006 (0.015)	-0.006 (0.016)	-0.028** (0.013)	0.003 (0.017)
Incentive First condition	-0.219*** (0.039)	-0.277*** (0.068)	-0.209*** (0.075)	-0.323*** (0.067)	-0.074 (0.070)
Incentive × Incentive First condition	0.074*** (0.012)	0.095*** (0.020)	0.067*** (0.023)	0.109*** (0.021)	0.028 (0.021)
Constant	0.454*** (0.040)	0.474*** (0.068)	0.413*** (0.073)	0.557*** (0.062)	0.370*** (0.068)
Observations	2,106	477	477	576	576
Subjects	117	53	53	64	64
Prior failure probability: 50%					
Incentive in €	-0.004 (0.008)	0.019 (0.013)	-0.022 (0.015)	-0.032 (0.019)	0.020 (0.014)
Incentive First condition	-0.137*** (0.037)	-0.095 (0.070)	-0.194*** (0.071)	-0.256*** (0.078)	-0.000 (0.060)
Incentive × Incentive First condition	0.063*** (0.011)	0.050** (0.019)	0.079*** (0.022)	0.093*** (0.024)	0.031 (0.021)
Constant	0.356*** (0.040)	0.304*** (0.062)	0.372*** (0.067)	0.506*** (0.079)	0.250*** (0.066)
Observations	2,034	567	567	450	450
Subjects	113	63	63	50	50
Prior failure probability: 80%					
Incentive in €	-0.007 (0.008)	-0.010 (0.017)	0.000 (0.014)	-0.010 (0.015)	-0.008 (0.015)
Incentive First condition	-0.107** (0.041)	-0.224*** (0.078)	-0.028 (0.061)	-0.139* (0.072)	-0.036 (0.071)
Incentive × Incentive First condition	0.037*** (0.013)	0.077*** (0.027)	0.017 (0.019)	0.037* (0.022)	0.016 (0.024)
Constant	0.340*** (0.039)	0.410*** (0.072)	0.224*** (0.061)	0.382*** (0.064)	0.343*** (0.061)
Observations	2,106	522	522	531	531
Subjects	117	58	58	59	59

Note The table displays the estimated coefficients of OLS regressions of an indicator that the subject chooses the Bold Advisor on the listed predictor variables. Standard errors in parentheses, clustered by subject.

Table B.2: Participation decisions in Experiment 1 by treatment condition

VARIABLES	(1)	(2)	(3)	(4)	(5)
	Participates in venture				
<i>Consequences</i>					
Compensable	Both	Yes	No	Yes	No
Delayed	Both	No	No	Yes	Yes
	Prior failure probability: 20%				
Incentive in €	0.079*** (0.012)	0.081*** (0.022)	0.079*** (0.022)	0.081*** (0.020)	0.077*** (0.020)
Incentive First condition	-0.130*** (0.042)	-0.124 (0.085)	-0.030 (0.079)	-0.167** (0.078)	-0.181** (0.070)
Incentive × Incentive First condition	0.035*** (0.013)	0.031 (0.028)	0.039 (0.026)	0.045* (0.024)	0.026 (0.019)
Constant	0.324*** (0.041)	0.360*** (0.074)	0.210*** (0.074)	0.370*** (0.069)	0.341*** (0.069)
Observations	2,106	477	477	576	576
Subjects	117	53	53	64	64
	Prior failure probability: 50%				
Incentive in €	0.051*** (0.010)	0.090*** (0.018)	0.030* (0.018)	0.040* (0.020)	0.038* (0.022)
Incentive First condition	-0.076** (0.037)	0.025 (0.066)	-0.117 (0.072)	-0.161** (0.078)	-0.067 (0.088)
Incentive × Incentive First condition	0.037*** (0.013)	0.007 (0.024)	0.038 (0.023)	0.066*** (0.024)	0.046 (0.028)
Constant	0.252*** (0.033)	0.153*** (0.056)	0.290*** (0.062)	0.340*** (0.071)	0.241*** (0.076)
Observations	2,034	567	567	450	450
Subjects	113	63	63	50	50
	Prior failure probability: 80%				
Incentive in €	0.061*** (0.009)	0.093*** (0.021)	0.050*** (0.015)	0.054*** (0.017)	0.046*** (0.017)
Incentive First condition	0.015 (0.026)	-0.020 (0.057)	0.098* (0.050)	-0.064 (0.057)	0.046 (0.052)
Incentive × Incentive First condition	-0.001 (0.010)	-0.007 (0.019)	-0.016 (0.018)	0.023 (0.017)	-0.002 (0.018)
Constant	0.073** (0.028)	0.104* (0.059)	0.027 (0.043)	0.113* (0.057)	0.049 (0.054)
Observations	2,106	522	522	531	531
Subjects	117	58	58	59	59

Note The table displays the estimated coefficients of OLS regressions of an indicator that the subject participates in the transaction on the listed predictor variables. Standard errors in parentheses, clustered by subject.

B.2 Welfare analysis with certainty equivalents predicted by ridge-regression

Here, I address the fact that noisy elicitation of certainty equivalents might bias the estimates in Table 3, as follows. I run ridge regressions with 10-fold cross-validation to predict individuals' certainty equivalents for each of the 26 lotteries they are presented with in part 2 of the experiment.⁵ I run a separate regression for each round of part 2. As predictors, I include the gain and loss amounts, the success probability, and a treatment indicator for the Delay condition of the current round. I include all two- and three-way interactions between these variables. Additionally, I include the certainty equivalents of all other rounds, as well as gain and loss amounts, success probabilities, and Delay condition indicators. In addition, I include all two-, three-, and four-way interactions between these predictor variables.

Table B.3 displays the results. The Undue Inducement Condition is violated in all but one cases. The exception occurs in the case of non-compensable, contemporaneous consequences with a prior failure probability of 80%. The range of welfare parameters α for which the capping incentives is welfare-optimal is given by $[0.122, 0.177]$, which requires the planner to place much less weight on individuals who are harmed by the transaction than on those who gain. By contrast, the undue inducement literature places more weight on individuals who are harmed. According to such welfare weights, the transaction in the affected row should be prevented at all incentives.

⁵I use ridge regression rather than Lasso or elastic nets because [Peysakhovich and Naecker \(2017\)](#) find that the sparsity assumptions incorporated in the latter leads to substantially lower predictive power in the domain of predicting choice under risk.

Table B.3: Welfare gains, welfare losses, and tests for undue inducement in Experiment 1 based on certainty equivalents predicted by Ridge-regression

Variable	(1) $\hat{\beta}_0^G$	(2) $\hat{\beta}_0^L$	(3) $\hat{\beta}_1^G$	(4) $\hat{\beta}_1^L$	(5) $\underline{\alpha}$	(6) $\bar{\alpha}$	(7) Condition (4) satisfied	(8) p -value
<i>Consequences</i>								
Compensable								
Contemporaneous								
Prior failure probability								
80%	0.084*** (0.025)	-0.191*** (0.053)	0.045*** (0.011)	0.016 (0.018)	1.572 (0.995)	0.305* (0.094)	No	0.223
50%	0.108*** (0.025)	-0.195*** (0.058)	0.168*** (0.024)	0.036** (0.015)	1.273*** (0.155)	0.356*** (0.092)	No	0.000
20%	0.258*** (0.029)	-0.054** (0.024)	0.481*** (0.039)	0.013** (0.006)	1.029*** (0.013)	0.828 (0.068)	No	0.012
Delayed								
Prior failure probability								
80%	0.040*** (0.015)	-0.141*** (0.054)	0.089*** (0.020)	0.022** (0.011)	1.325*** (0.235)	0.221*** (0.093)	No	0.000
50%	0.146*** (0.031)	-0.227*** (0.087)	0.185*** (0.025)	0.016 (0.032)	1.096*** (0.208)	0.391*** (0.108)	No	0.006
20%	0.194*** (0.027)	-0.123*** (0.044)	0.342*** (0.033)	-0.017 (0.020)	0.952*** (0.053)	0.613*** (0.099)	No	0.000
Noncompensable								
Contemporaneous								
Prior failure probability								
80%	0.057*** (0.020)	-0.268*** (0.081)	0.008 (0.009)	-0.057* (0.032)	0.122 (0.132)	0.177* (0.071)	Yes	0.724
50%	0.109*** (0.026)	-0.207*** (0.061)	0.043*** (0.013)	-0.042* (0.025)	0.505*** (0.180)	0.345*** (0.092)	No	0.453
20%	0.202*** (0.033)	-0.065** (0.027)	0.243*** (0.028)	0.001 (0.008)	1.003*** (0.033)	0.756*** (0.088)	No	0.017
Delayed								
Prior failure probability								
80%	0.027** (0.012)	-0.288*** (0.095)	0.022** (0.010)	-0.028 (0.024)	0.443* (0.251)	0.086* (0.046)	No	0.162
50%	0.146*** (0.032)	-0.214*** (0.066)	0.084*** (0.020)	-0.026 (0.032)	0.760*** (0.226)	0.406*** (0.101)	No	0.130
20%	0.158*** (0.032)	-0.184*** (0.055)	0.184*** (0.029)	0.008 (0.016)	1.042*** (0.096)	0.461*** (0.097)	No	0.000

Note Each row corresponds to a separate treatment within the Incentive First Condition. Coefficients in each row are estimated in a separate SUR-regression. Columns 1 and 2 show the level of mean welfare gains and losses, respectively, at the €1 incentive. Columns 3 and 4 show the change in mean welfare gains and losses, respectively as the incentive increases by €1. Column 5 shows the least weight α that needs to be placed on those who lose from the transaction such that an increase in the incentive is deemed welfare-decreasing. Column 6 shows the maximal weight α that may be placed on those who lose from the transaction such that the transaction is admissible at sufficiently low incentives. Reported significance levels in columns 5 and 6 reflect the tests that $\underline{\alpha} = 0$ and $\bar{\alpha} = 1$, respectively. Column 7 indicates whether $T = \bar{\alpha} - \underline{\alpha}$ is nonnegative. Column 8 lists p -values for two sided tests of the Null hypothesis that $T = 0$. Standard errors in parentheses, clustered by subject. Standard errors in columns 5, 6, and 8 are calculated by the Delta method.

B.3 Welfare analysis, ex-post benchmark

Here, I perform an analysis parallel to Table 3 with an ex-post definition of individual welfare. Specifically, for a subject who accepts the transaction $w_i = m$ if the state is good. If the state is bad, then $w_i = -d + m$ or $w_i = -d$ in the Compensable and Noncompensable conditions, respectively. If the subject rejects, then $w_i = 0$. I use this definition for the Contemporaneous and the Delay condition. In the Delay condition, this definition involves the normative position that subjects should not discount future losses. The Undue Inducement Condition is violated throughout.

Table B.4: Welfare gains, welfare losses, and tests for undue inducement Experiment 1 based on the ex-post welfare benchmark

Variable	(1) $\hat{\beta}_0^G$	(2) $\hat{\beta}_0^L$	(3) $\hat{\beta}_1^G$	(4) $\hat{\beta}_1^L$	(5) $\underline{\alpha}$	(6) $\bar{\alpha}$	(7) Condition (4)	(8) satisfied p -value
<i>Consequences</i>								
Compensable								
Contemporaneous								
Prior failure probability								
80%	0.106 (0.067)	-0.230*** (0.078)	0.134*** (0.050)	0.041* (0.021)	1.435*** (0.388)	0.316*** (0.169)	No	0.010
50%	0.144* (0.079)	-0.205** (0.091)	0.459*** (0.076)	0.051** (0.026)	1.124*** (0.075)	0.412*** (0.173)	No	0.001
20%	-0.025 (0.058)	-0.137** (0.068)	0.231*** (0.064)	0.032 (0.021)	1.163*** (0.127)	-0.219 (0.656)	No	0.020
Delayed								
Prior failure probability								
80%	0.219*** (0.082)	-0.056 (0.039)	0.468*** (0.078)	-0.019 (0.035)	0.961*** (0.071)	0.797 (0.131)	No	0.368
50%	0.190** (0.082)	-0.033 (0.038)	0.794*** (0.065)	-0.005 (0.018)	0.994*** (0.022)	0.853 (0.163)	No	0.431
20%	0.104*** (0.029)	-0.201*** (0.060)	0.133*** (0.041)	-0.038 (0.029)	0.776*** (0.153)	0.340*** (0.095)	No	0.039
Noncompensable								
Contemporaneous								
Prior failure probability								
80%	0.239*** (0.034)	-0.107*** (0.040)	0.387*** (0.055)	-0.012 (0.017)	0.970*** (0.041)	0.691*** (0.084)	No	0.011
50%	0.466*** (0.049)	-0.007 (0.007)	0.624*** (0.060)	-0.007 (0.007)	0.989*** (0.011)	0.984 (0.015)	No	0.816
20%	0.112*** (0.028)	-0.197*** (0.066)	0.138*** (0.037)	-0.027 (0.021)	0.836*** (0.116)	0.362*** (0.088)	No	0.002
Delayed								
Prior failure probability								
80%	0.253*** (0.039)	-0.099** (0.048)	0.465*** (0.058)	-0.014 (0.017)	0.971*** (0.036)	0.719*** (0.108)	No	0.042
50%	0.353*** (0.041)	-0.045 (0.028)	0.598*** (0.054)	0.007 (0.008)	1.012*** (0.013)	0.886 (0.064)	No	0.100
20%	-	-	-	-	-	-	-	-

Notes Each row corresponds to a separate treatment within the Incentive First Condition. Estimates in the bottom row are missing because the incidence of losses is too low for the model to be estimated. Coefficients in each row are estimated in a separate SUR-regression. Columns 1 and 2 show the level of mean welfare gains and losses, respectively, at the €1 incentive. Columns 3 and 4 show the change in mean welfare gains and losses, respectively as the incentive increases by €1. Column 5 shows the least weight α that needs to be placed on those who lose from the transaction such that an increase in the incentive is deemed welfare-decreasing. Column 6 shows the maximal weight α that may be placed on those who lose from the transaction such that the transaction is admissible at sufficiently low incentives. Reported significance levels in columns 5 and 6 reflect the tests that $\underline{\alpha} = 0$ and $\bar{\alpha} = 1$, respectively. Column 7 indicates whether $T = \bar{\alpha} - \underline{\alpha}$ is nonnegative. Column 8 lists p -values for two sided tests of the Null hypothesis that $T = 0$. Standard errors in parentheses, clustered by subject. Standard errors in columns 5, 6, and 8 are calculated by the Delta method.

C Experiment 2: Additional Analysis

C.1 Randomization check

The four treatments are balanced across demographic characteristics. Table C.1 displays summary statistics of these variables by treatment. For each variable, the table reports the p -value of an F -test for differences in the mean value of the variable across treatments. Of 24 tests conducted, one is significant at the 5% level.

Table C.1: Summary statistics and randomization check.

<i>Treatment condition</i>					
Incentive	\$30	\$3	\$30	\$3	
Video	Yes	Yes	No	No	
Variable	Mean				p -value ^a
Male	0.55	0.53	0.54	0.54	1.00
Age	21.43	22.01	21.37	21.30	0.34
<i>Ethnicity</i>					
African-American	0.05	0.06	0.07	0.07	0.70
Caucasian	0.57	0.51	0.59	0.56	0.28
East Asian	0.19	0.26	0.19	0.23	0.22
Hispanic	0.07	0.08	0.04	0.04	0.99
Indian	0.03	0.04	0.04	0.07	0.56
Other	0.08	0.05	0.07	0.04	0.70
Monthly spending in USD	251.72	301.40	289.07	288.42	0.44
Year of study ^b	3.50	3.60	3.61	3.47	0.32
Graduate student	0.13	0.15	0.13	0.05	0.07
<i>Field of study</i>					
Arts and humanities	0.16	0.09	0.13	0.11	0.04
Business or economics	0.27	0.36	0.34	0.43	0.09
Engineering	0.20	0.16	0.11	0.12	0.49
Science	0.21	0.23	0.27	0.23	0.47
Social science (excluding business and economics)	0.17	0.17	0.15	0.11	0.59
Political orientation ^c	0.50	0.32	0.27	0.09	0.08
Raven's score ^d	14.77	14.76	14.69	14.68	1.00
CRT score ^e	3.76	3.80	3.50	3.22	0.08
<i>Experience with insects as food (1 = Yes, 0 = No)</i>					
Has intentionally eaten insects before	0.19	0.22	0.19	0.20	0.71
Grown up in culture that practices entomophagy	0.15	0.14	0.13	0.15	0.92
Grown up eating mostly western foods	0.81	0.73	0.82	0.78	0.06
Had a pet that fed on store-bought insects	0.25	0.25	0.21	0.26	0.68
Knew that this study concerns insect eating	0.20	0.30	0.26	0.29	0.23

^a p -value of the test of joint significance of a regression of the indicated variable on treatment dummies.

^bYear of study only includes undergraduate students.

^cPolitical orientation is measured on a scale of -2 (conservative) to 2 (liberal).

^dRaven's score is measured on a scale of 0 to 24.

^eCRT score refers to performance on version of the test by Toplak et al. (2014). It is measured on a scale of 0 to 6.

C.2 Welfare analysis accounting for noisy elicitation

To account for noisy elicitation of reservation prices, I employ the same method as in Experiment 1. Within each treatment, I use ridge regression to predict subject i 's reservation price for species s using her reservation prices for all species except s , using the treatment indicators, and all two-way and three-way interactions between these variables (as well as squared and cubed values of each variable) that do not involve the dependent variable.⁶

I calculate Marshallian surplus by combining observed participation decisions with predicted reservation prices. Panel A of Table C.2 displays the results. The magnitude of welfare losses is weakly decreasing in the incentive amount for each species, which implies that the Undue Inducement Condition (4) is violated throughout.

I perform a parallel analysis using as a welfare benchmark the reservation prices subjects revealed in the (\$3, No Video)-treatment. I first predict reservation prices in that treatment, as well as in the Video condition. I then perform the percentile-match described in Section 4.2 using predicted rather than observed reservation prices. Panel B of Table C.2 displays the results. I find once more that the Undue Inducement Condition (4) is violated for each species.

⁶I account for upper censoring by fitting a log-normal distribution and replacing censored values by the conditional expectation, as explained in footnote 68.

Table C.2: Welfare effects by species in Experiment 2, controlling for noisy elicitation using ridge-regression predicted reservation prices

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\hat{\beta}_0^G$	$\hat{\beta}_0^L$	β_1^G	β_1^L	$\underline{\alpha}$	$\bar{\alpha}$	Condition (4)	
							satisfied	p -value
A. Within-treatment benchmark								
2 house crickets	0.000 (0.522)	-2.451*** (0.628)	0.278*** (0.027)	0.044 (0.033)	1.188*** (0.166)	0.000*** (0.213)	No	0.000
5 superworms	0.000 (0.427)	-1.775*** (0.314)	0.197*** (0.022)	0.061*** (0.016)	1.453*** (0.189)	0.000*** (0.240)	No	0.000
3 silkworm pupae	-0.000 (0.481)	-3.737*** (0.753)	0.249*** (0.025)	0.056 (0.039)	1.293*** (0.262)	-0.000*** (0.129)	No	0.000
2 mole crickets	0.000 (0.351)	-5.481*** (0.943)	0.153*** (0.018)	0.065 (0.049)	1.742* (0.965)	0.000* (0.064)	No	0.071
2 field crickets	-0.000 (0.568)	-2.755*** (0.675)	0.323*** (0.030)	0.043 (0.035)	1.152*** (0.144)	-0.000*** (0.206)	No	0.000
B. No Video, \$3 incentive benchmark								
2 house crickets	0.000 (0.425)	-3.308*** (0.694)	0.189*** (0.022)	0.050 (0.036)	1.363*** (0.354)	0.000*** (0.128)	No	0.000
5 superworms	-0.000 (0.289)	-2.735*** (0.470)	0.104*** (0.015)	0.057** (0.024)	2.208* (1.187)	-0.000* (0.106)	No	0.060
3 silkworm pupae	0.000 (0.344)	-4.338*** (0.800)	0.140*** (0.018)	0.037 (0.042)	1.362** (0.545)	0.000** (0.079)	No	0.013
2 mole crickets	0.000 (0.179)	-6.896*** (1.021)	0.054*** (0.009)	0.039 (0.053)	3.444 (11.484)	0.000* (0.026)	No	0.764
2 field crickets	-0.000 (0.469)	-3.701*** (0.667)	0.237*** (0.024)	0.065* (0.035)	1.380*** (0.278)	-0.000*** (0.127)	No	0.000

Notes: Each row corresponds to a separate SUR-regression using data from the Video condition. Columns 1 and 2 show the level of mean welfare gains and losses, respectively, at the \$3 incentive. Columns 3 and 4 show the change in mean welfare gains and losses, respectively as the incentive increases by \$1. Column 5 shows the least weight α that needs to be placed on those who lose from the transaction such that an increase in the incentive is deemed welfare-decreasing. Column 6 shows the maximal weight α that may be placed on those who lose from the transaction such that the transaction is admissible at sufficiently low incentives. Reported significance levels in columns 5 and 6 reflect the tests that $\underline{\alpha} = 0$ and $\bar{\alpha} = 1$, respectively. Column 7 indicates whether $T = \bar{\alpha} - \underline{\alpha}$ is nonnegative. Column 8 lists p -values for two sided tests of the Null hypothesis that $T = 0$. Standard errors in parentheses, clustered by subject. Standard errors in columns 5, 6, and 8 are calculated by the Delta method.

D Experiment 3: Sequential Information Search

The experiment in this section follows the same design as Experiment 1, but it provides subjects with a different information acquisition technology.⁷ While the results in Experiment 1 might be driven by the selection of information structures available to the subjects, this is not a concern in the current experiment. In particular, subjects can access information structures that are symmetric (encouraging and discouraging signals are equally likely ex ante), and they can choose not only the kind but also the amount of information (in the sense that they can implement information structures that can be Blackwell-ordered). In contrast to the first experiment, information costs are no longer induced, but are given by the subjects' own disutility of effort. Moreover, while subjects in the first experiment can, in principle derive Bayesian posteriors by simple arithmetic, this is not possible in the present experiment. As a result, correct belief updating might be more difficult, possibly leading to more frequent suboptimal choice.

D.1 Design

The main distinction between Experiment 1 and the present experiment consists in the information acquisition technology. Here, subjects obtain information about whether the venture will be successful by examining a picture of 450 randomly arranged letters as in panel A of Figure D.1. Subjects know that if the venture is a success, the picture contains 50 letters *G* and 40 letters *B* (for “good” and “bad”, respectively), and that these numbers are reversed if their venture is a failure.⁸ There is no time constraint. Hence, somebody willing to put in the time and effort to count the letters can know with certainty whether accepting the transaction will lead to a net monetary gain or to a net monetary loss. A black panel covers the grid, but subjects can use their mouse to move a hole across the grid through which they can inspect one letter at a time, as in panel B of Figure D.1.⁹ From the recorded sequence of letters the subject chose to observe, I calculate Bayesian posteriors.¹⁰ By selecting a search strategy, subjects can implement a wide variety of information structures.¹¹

The set of treatments and corresponding parameters are the same as in Experiment 1. The one exception concerns the incentive amounts which are given by \$0.5, \$1, \$5, and \$5.5.¹² Moreover, I use the term *Grid*

⁷A previous version of this paper reports a closely related experiment. While that experiment neither elicited welfare benchmarks nor recorded the information observed by the subjects, the part of the analysis that overlaps across the current and past experiment yields highly similar results.

⁸Dean and Neligh (2017) use a similar paradigm for presenting information.

⁹Each picture is presented with a different background color. Color assignment is randomized on the individual level. For each subject and each decision, a state is drawn independently, and a new picture of randomly scrambled letters is generated. Subjects could not use a text editor to automatically count the letters because they were presented in a picture format (HTML5 Canvas).

¹⁰The likelihood ratio for observing n draws of letters *G* or *B* of which k letters were *G* is given by $D = \frac{P(k|n, s=G)}{P(k|n, s=B)} = \frac{(40-k)! (50-(n-k))!}{(50-k)! (40-(n-k))!}$ as follows directly from the hypergeometric distribution function. I calculate the Bayesian posterior using the formula $\logit[P(s=G|n, k)] = \logit(\mu) + \log(D)$ where $\logit(x) = \log\left(\frac{x}{1-x}\right)$. Each letter on the grid is counted only once, even if the subject moves the cursor over that letter repeatedly.

¹¹Technically, this paradigm is an approximation to the (Wald, 1947) sequential sampling framework, in which any pair of Bayes-consistent posterior beliefs $(\gamma_G, \gamma_B) \in [0, 1]^2$ can be implemented as the informativeness of individual draws approaches zero. The two differences between this design and the Wald framework are, first, that drawing of signals (*B*s or *G*s) occurs without replacement, and second, that the experimental setting is finite.

¹²The \$0.5 and \$5.5 payments appeared to invite boundary effects, so I selected less extreme payment amounts for Experiment 1, which I conducted after the present experiment.

Figure D.1: Presentation of information about the state.

Notes: Panel A: In case the subject will win upon taking the gamble, the picture has 50 letters *G* and 40 letters *B*. In case the subject will lose, it has 40 letters *G* and 50 letters *B*. Panel B: The subject moves the mouse to change the location of the illuminated point. We record the sequence of letters the subject observes.

First to refer to the case in which subjects examine the grid of letters before learning about the incentive amount for the round (corresponding to the Advisor Choice First condition in Experiment 1).

Because the information technology used here is more time consuming than choosing advisors as in Experiment 1, part 1 consists of only 16 rounds, corresponding to 4 incentive levels, the Incentive First and Grid First condition, and the Compensable and Non-Compensable conditions, administered in random order.

Elicitation of the ex-ante welfare benchmark As is the case in Experiment 1, part 2 of the current experiment presents subjects with lotteries that correspond to the decisions the subject faced in each round of part 1 if she updated in a fully Bayesian manner. The subject proceeds through 16 such decisions corresponding to part 1, as well as through 12 additional decisions for which the success probability is exogenously given by 0.2, 0.5, and 0.8, once for each of the incentive amounts \$1 and \$5 and once for each the Compensable and Non-Compensable condition, respectively. The total of 28 rounds are administered in individually randomized order.

The survey automatically calculates the Bayesian posterior depending on the part of the grid of letters the subject has uncovered during her search. If a subject rapidly moves the mouse across the display, however, letters may be displayed for imperceptibly brief time periods. [Bundesen and Harms \(1999\)](#) show that subjects cannot identify letters with any accuracy if they are displayed for less than 50 milliseconds. Accordingly, I record a letter as seen only if it is displayed for 50 milliseconds or longer. The Bayesian posteriors subjects encounter in part 2 of the experiment are calculated based on this benchmark.

Implementation I conducted this experiment on the Amazon Mechanical Turk online labor market with a total of 386 subjects on November 9-12, 2020.¹³ Each subject receives a \$12 completion payment from which potential losses are discounted. Subjects in the Delay condition receive \$6 of these \$12 with a three months delay (net of any losses they may incur). In addition, subjects receive a base payment of \$1. Subjects could only participate on a desktop or laptop computer with a screen sufficiently large to display all pictures without scrolling.¹⁴

Because the experiment is administered online, attrition is a possible concern. 614 subjects begin the survey on a device that is technically equipped to display the survey. Of these, 70 attrit during the instructions before reaching the first comprehension check. Another 104 do not progress beyond the first comprehension check, and another 11 attrit at or before the second comprehension check. Of these 429 subjects, 405 complete the entire survey. In the instructions, subjects learn whether they will receive some part of the payment with a three months delay, and they learn the prior failure probability they will face. Accordingly, attrition could in principle vary by treatment condition. Amongst the 209 subjects who attrited 99 were in the Contemporaneous condition and 110 were in the Delay condition, whereas 75, 81, and 53 faced an 80%, 50%, or 20% prior failure probability, respectively. Regressing an indicator for whether a subject attrited on an indicator for the Delay condition, the numerical prior failure probability and the interaction between the two has no explanatory power ($p > 0.3$ on all coefficients except for the constant term, and $p > 0.5$ for the F -test of joint explanatory power of all variables compared to a constant-only model).

The median subject completed the experiment in 63.4 minutes; average earnings are \$13.58. Subjects examined each picture for a median time of 12 seconds, with 55 seconds at the 75th percentile and 2 seconds at the 25th percentile.

D.2 Analysis

I begin by studying the effect of incentives on information acquisition. The undue inducement literature is concerned that higher incentives lead to less informed decision making. On the contrary, Panel A of Figure D.2 shows that higher incentives lead to increased search in the Incentive First condition (pooling across all remaining conditions), as measured by the number of informative letters (“B” or “G”) in the string of letters the subject saw. The magnitude of the effect is substantial. In the case of the \$0.5 incentive, subjects parse strings with 12 informative letters on average, which almost doubles to 22 informative letters for the case of a \$5.5 incentive. Hence, higher incentives do not, on average, cause less careful decision making. Moreover, there is no indication that subjects engage in information biasing in the present experiment (in contrast to Experiment 1 in which information biasing is explicit).

Panel B shows the effect of incentives on participation, separately for the Incentive First and for the Grid First condition (pooling across all remaining conditions). Supply responds more strongly if the subject can tailor information search to the incentive amount, but the effect is small (2.8 percentage points as the incentive increases from \$1 to \$5, $p = 0.076$).

¹³I required subjects to be located in the US, to have at least a 99% approval rate on previous tasks completed on the platform, and to have completed at least 500 tasks on the platform.

¹⁴The survey automatically records screen width and height, as well as the device from which it is being taken. The survey does not let the worker participate if the screen is too small, or if it is being taken from a mobile device.

Figure D.2: Results of Experiment 2.

Notes Panel A: Number of informative letters (“B”/“G”) examined by incentive and condition. Panel B: Participation probability by Bayesian posterior. This figure plots estimates of the Bayesian posterior for each level of elicited posterior. Standard errors are clustered by subject.

Table D.1 disaggregates these results across all treatment conditions. Higher incentives uniformly lead to a stronger increase in information acquisition in the Incentive First condition than in the Grid First condition. The effect is statistically significant at the 1% level in four of twelve cases. With three exceptions pertaining to treatments with an 80% prior failure probability, the effect of incentives is slightly stronger in the Incentive First than in the Grid First condition, although the effect reaches statistical significance at the 10% level in only one case.

Broadly, the results of the present experiment are inconsistent with the behavioral part of the Undue Inducement Hypothesis. Nonetheless, it is conceivable that higher incentives affect welfare negatively. I study this question next.

Welfare I perform the same type of welfare analysis as in Experiment 1. All procedural details are identical. Specifically, I let w_i denote subject i ’s ex-ante welfare from her participation decision. If the subject accepted the transaction, then w_i is her certainty equivalent, as measured in part 2 of the experiment, for the gamble that she faced when deciding whether to participate after having studied the picture, where the success probability is explicitly stated, and given by the posterior a Bayesian would have held at that point. If the subject refused to participate, $w_i = 0$. Within each treatment, I define the average welfare gain at incentive m as $g(m) = \frac{1}{n} \sum_{i=1}^n \max(w_i, 0)$ and the average loss as $l(m) = \frac{1}{n} \sum_{i=1}^n \min(w_i, 0)$ where n is the total number of subjects to whom the transaction is offered. I report intercepts at $m = 1$ rather than at $m = 0$, using the notation $\tilde{\beta}_0^w = \beta_0^w + \beta_1^w$, for $w \in \{G, L\}$.

Panel A of Figure D.3 displays the distribution of welfare for the case of compensable consequences separately by incentive category in the Grid First condition, pooling across the remaining conditions. Panel B shows the corresponding distribution for the case of noncompensable consequences in the Incentive First

Figure D.3: Welfare in Experiment 3: Cumulative Distribution Functions.

Notes Panels A and B display cumulative distribution functions of certainty equivalents corresponding to the transactions the subjects faced in part 1 of the experiment, for the compensable, grid-first condition, as well as for the noncompensable, incentive-first condition (pooled across the remaining conditions). For trials in which subjects refused the transaction, the certainty equivalent is set to zero. For easier visibility, I group the \$0.5 and \$1 incentives into the low incentive category, and I group the \$5 and \$5.5 incentives into the high incentive category. Probability mass to the left of zero indicates welfare losses, probability mass to the right of zero indicates welfare gains. Panels C and D display average welfare gains $g(m) = \frac{1}{n} \sum_{i=1}^n \max(w_i, 0)$ and average welfare losses $l(m) = \frac{1}{n} \sum_{i=1}^n \min(w_i, 0)$ by incentive amount, across the same conditions as panels A and B, respectively, averaged across the Prior and Timing conditions. Whiskers indicate 95% confidence intervals. Standard errors clustered by subject.

Table D.1: Behavioral effects in Experiment 3

Dependent variable	(1)	(2)	(3)	(4)
	Information choice		Supply	
	Signal length		Participates in venture	
Consequences	Delayed	Contemporaneous	Delayed	Contemporaneous
Compensable				
Prior failure prob.				
80%	2.196***	1.690**	0.013	-0.012
	0.740	0.677	0.012	0.015
50%	1.231***	1.876***	0.007	0.018
	0.448	0.627	0.012	0.014
20%	0.855	3.488***	0.004	0.020*
	0.818	0.833	0.015	0.012
Non-compensable				
Prior failure prob.				
80%	2.295***	2.966***	-0.005	-0.004
	0.707	0.907	0.013	0.014
50%	0.180	2.933***	0.000	0.011
	0.527	0.753	0.012	0.015
20%	2.859***	3.186***	0.016	0.010
	0.869	0.805	0.014	0.014

Note Each entry in the table displays the estimate of the interaction coefficient of a regression of the dependent variable on the incentive amount, an indicator for the Incentive First condition, and the interaction between the two. Each entry corresponds to a separate OLS regression. Standard errors in parentheses, clustered by subject.

condition.¹⁵ Just as in Experiments 1 and 2, higher incentives greatly increase the average welfare gain and have little, if any, effect on the average welfare loss. Panels C and D display the effect of incentives on mean welfare gains and losses in the same conditions as panels A and B, respectively. Visual inspection shows that the Undue Inducement Condition is violated in each case.

Table D.2 formally tests the Undue Inducement Condition (4) within each treatment condition. As columns 7 and 8 show, the condition is not satisfied in any treatment; the hypothesis that $T = 0$ can be rejected at the 0.001 level in each case.

Overall, this experiment demonstrates that the welfare conclusions of Experiment 1 are robust to a more flexible form of information acquisition.

¹⁵Due to a coding error, certainty equivalents are unavailable for the case of compensable consequences in the Incentive First condition and for the case of noncompensable consequences in the Grid First condition.

Table D.2: Welfare effects in Experiment 3

Variable	(1) $\tilde{\beta}_0^G$	(2) $\tilde{\beta}_0^L$	(3) β_1^G	(4) β_1^L	(5) $\underline{\alpha}$	(6) $\bar{\alpha}$	(7) Condition (4)	(8) satisfied p -value
<i>Consequences</i>								
Compensable (Grid First)								
Contemporaneous								
Prior failure probability								
80%	0.081*** (0.017)	-0.579*** (0.134)	0.129*** (0.027)	-0.083*** (0.031)	0.609*** (0.110)	0.123*** (0.034)	No	0.000
50%	0.116*** (0.019)	-0.498*** (0.120)	0.217*** (0.029)	-0.015 (0.030)	0.934*** (0.124)	0.190*** (0.048)	No	0.000
20%	0.231*** (0.029)	-0.251*** (0.076)	0.500*** (0.040)	0.016 (0.021)	1.034*** (0.044)	0.480*** (0.084)	No	0.000
Delayed								
Prior failure probability								
80%	0.091*** (0.034)	-0.424*** (0.141)	0.137*** (0.028)	-0.059*** (0.017)	0.701*** (0.068)	0.177*** (0.077)	No	0.000
50%	0.140*** (0.039)	-0.312*** (0.092)	0.211*** (0.028)	-0.092*** (0.028)	0.696*** (0.075)	0.309*** (0.092)	No	0.001
20%	0.314*** (0.068)	-0.368*** (0.120)	0.418*** (0.042)	-0.005 (0.018)	0.987*** (0.041)	0.460*** (0.097)	No	0.000
Noncompensable (Incentive First)								
Contemporaneous								
Prior failure probability								
80%	0.067*** (0.018)	-0.417*** (0.109)	0.118*** (0.026)	-0.012 (0.022)	0.907*** (0.155)	0.138*** (0.048)	No	0.000
50%	0.102*** (0.022)	-0.359*** (0.091)	0.215*** (0.028)	-0.025 (0.024)	0.896*** (0.090)	0.221*** (0.061)	No	0.000
20%	0.238*** (0.033)	-0.208*** (0.067)	0.433*** (0.048)	0.010 (0.020)	1.024*** (0.048)	0.533*** (0.092)	No	0.000
Delayed								
Prior failure probability								
80%	0.091** (0.037)	-0.512*** (0.170)	0.100*** (0.026)	-0.011 (0.025)	0.898*** (0.198)	0.151*** (0.067)	No	0.000
50%	0.125*** (0.041)	-0.322*** (0.086)	0.181*** (0.025)	-0.040 (0.024)	0.820*** (0.095)	0.279*** (0.089)	No	0.000
20%	0.330*** (0.065)	-0.420*** (0.114)	0.469*** (0.044)	0.032*** (0.012)	1.074*** (0.031)	0.440*** (0.082)	No	0.000

Note Each row corresponds to a separate treatment. Coefficients in each row are estimated in a separate SUR-regression. Columns 1 and 2 show the level of mean welfare gains and losses, respectively, at the €1 incentive. Columns 3 and 4 show the change in mean welfare gains and losses, respectively as the incentive increases by €1. Column 5 shows the least weight α that needs to be placed on those who lose from the transaction such that an increase in the incentive is deemed welfare-decreasing. Column 6 shows the maximal weight α that may be placed on those who lose from the transaction such that the transaction is admissible at sufficiently low incentives. Reported significance levels in columns 5 and 6 reflect the tests that $\underline{\alpha} = 0$ and $\bar{\alpha} = 1$, respectively. Column 7 indicates whether $T = \bar{\alpha} - \underline{\alpha}$ is nonnegative. Column 8 lists p -values for two sided tests of the Null hypothesis that $T = 0$. Standard errors in parentheses, clustered by subject. Standard errors in columns 5, 6, and 8 are calculated by the Delta method.

E Experimental Materials

E.1 Video transcriptions

The following is a transcription of the videos used in that experiment. The videos are available at <https://youtu.be/HiNnbYuuRcA> (“Why you may want to eat insects”) and <https://youtu.be/ii4YSG0EcRY> (“Why you may not want to eat insects”).

Transcription: Why You May Want to Eat Insects Five reasons you should consider eating insects. For your own personal health, and for the overall health of the planet, and, most importantly, for your pleasure, you should be eating more insects. This isn’t meant as a provocative, theoretical idea. Here are five very serious reasons why you should consider increasing your insect intake.

First, insects can be yummy. You’d think that insects would have a pungent, unusual aroma. But they are actually very tasty, and considered a delicacy in many parts of the world. Also, like tofu, they often take on the flavor of whatever they’re cooked with. That’s why we are on the verge of a real insectivorous moment in consumer culture. The Brooklyn startup Exo just started selling protein bars made from ground cricket flour, and the British company Ento sells sushi-like bento boxes with cricket-based foods. The restaurant Don Bugito in San Francisco’s Mission district offers creative insect-based foods inspired by Mexican pre-hispanic and contemporary cuisine. “I am trying to bring a solution into the food market which is introducing edible insects” [Monica Martinez, owner of Don Bugito]. New cookbooks are entering the market, such as Daniella Martin’s “Edible”, or van Huis et al.’s “The insect cookbook”. Don Bugito’s reviews on yelp are glowing. Most Americans need some courage to take a bite. But once they do, they are pleasantly surprised. Morgane M., from Sunnyvale, CA describes her experience: “I saw their stand at the Ferry Building farmers market and decided to take the plunge. I tried the chili-lime crickets and they were surprisingly good! For the curious-but-apprehensive: the chili-lime crickets taste like flavorful, super crunchy (almost flaky) chips. That’s it. If you’ve ever had super thin tortilla chips, you’ll have an idea what to expect.” Other people liked them even more. For example Nelson Q. from Las Vegas, NV: “This Pre-Hispanic Snackeria has made me a fan They had the most interesting menu items of the evening at Off The Grid ... Would I try insects again??? Yessir!...ALOHA!!! ” Rodney H. from San Francisco agrees: “It’s great! And the mealworms add kind of a nice, savory quality to it. You never would guess that you’re eating an insect.”

Second, insects are a highly nutritious protein source. “Insects are actually the most ... one of the most efficient proteins on the planet” [Monica Martinez]. It turns out that pound for pound, insects provide much higher levels of protein compared to conventional meats like beef, chicken, and fish. While eggs consist to just 12% proteins, and beef jerky clocks in at 33%, a single pound of cricket flour has 65% protein. That’s twice as much as you get in beef jerky! Insects also have much higher levels of nutrients like calcium, iron, and zinc. They are also good sources of vitamin B12. That’s an essential vitamin that’s barely found in any plant-based foods and thus can be difficult for vegans to come by.

Third, our objection to eating insects is arbitrary. Your first reaction to this movie was probably a sense of dislike. But there’s nothing innate about that reaction. For one, billions of people already eat insects in Asia, Africa, and Latin America every day. More generally, the animals considered to be fit for consumption

vary widely from culture to culture for arbitrary reasons. Most Americans consider the idea of eating horses or dogs repugnant, even though there's nothing substantial that differentiates horses from cows. Meanwhile, in India, eating cows is taboo, while eating goat is common. These random variations are the results of cultural beliefs that crystallize over generations. But luckily, these arbitrary taboos can be defeated over time. There was a time when raw fish – served as sushi – was seen as repugnant in mainstream US culture. Now it's ubiquitous. Soon, insects – which are closely related to shrimp – may be elegant hors d'oeuvres.

Fourth, insects are more sustainable than chicken, pork, or beef. "I think the biggest problem for United States right now is we eating too much cattle, too much meat" [Monica Martinez.] Insects are a serious solution to our increasingly pressing environmental problems. It takes 2000 gallons of water to produce a single pound of beef, and 800 gallons for one pound of pork. How much do you think is required for a pound of crickets? One single gallon! Producing a pound of beef also takes thirteen times more arable land than raising a pound of crickets. It needs twelve times as much feed, and produces 100 times as much greenhouse gases. These very handsome environmental benefits are why the UN has released a 200 page report on how eating insects could solve the world's hunger and environmental problems just two years ago. Needless to say, the UN strongly advocates for insects as a food source. And it's not just the UN. In 2011, the European Commission has offered a four million dollar prize to the group that comes up with the best idea for developing insects as a popular food.

Five, we already eat insects all the time. The majority of processed foods you buy have pieces of insect in them. The last jar of peanut butter you bought, for instance, may have had up to 50 insect fragments. A bar of chocolate can have about 60 fragments of various insect species. Some experts estimate that, in total, we eat about one or two pounds of insects each year with our food. These insects pose no health risks. The FDA does set limits, but they are simply set for aesthetic reasons in other words, so you don't actually see them mixed into your food.

To summarize, these are five very compelling reasons to give it a try. Five, we already eat insects all the time anyway. Four, insects are more sustainable and ethical than chicken, pork, or beef. Three, our objection to eating insects is completely arbitrary. Two, insects are a highly nutritious protein source. One. "Most of people react really, really positive" [Monica Martinez]. Insects can be very tasty!

Transcription: Why You May Not Want to Eat Insects Four reasons you may want to avoid eating insects. Reason 1. Some cultures eat insects. But to those of us who are not used to it, insects can be... well, see for yourself. [American tourist in China] "Welcome to eating crazy foods around the world with Mike. And we're in China. If I've learned one thing about China it's they will eat absolutely everything. So you have caterpillars and you have butterflies. The pupae is what the caterpillar turns into before it turns into a butterfly. ... they don't look very appealing at all. But ... try everything once. So, up to the face. Hhh." [Eats pupae.] "Not good. Ugh ... it ... it popped. It popped! It's just ... it's just too much for me." [Throws remaining pupae into trash bin.] [Bear Grylls] "Whoa! Ready for this? Oh my goodness! Pfh! This one has been living in there a very, very long time. I'm not gonna need to eat for a week after this. Pfh." [Eats live beetle larva.] "Argh! This actually ranks as one of the worst things I've ever, ever eaten!"

Reason 2. Insects have many body parts. Most of those parts we do not usually eat in other animals. Let's see those parts... [Biology student] "Let's take a closer look at some of the structures we see on this grasshopper. So the first thing I want to point out is that it has six legs. There are two pairs. Here is the pair of hindlegs. There's a pair in the middle here, on the middle segment of the thorax. Ok, those are the midlegs. And then there's another pair on the front here, those are the forelegs or prolegs. Ok? So there's six altogether, all insects have six legs, or three pairs of legs, it's characteristic of the class. Ok? So we also can see, right up here, there are a pair of wings. On each side of the body there are two wings. The forewing, k? – as in the one in front – and this is the hindwing down here, ok? So there are four wings on this animal. Other insects only have two, some have none. Now we'll move up to the head. The first thing you'll notice is this pair of long antennae. Ok, we've seen antennae in other animals. So, clearly, those are involved ... they have a sensory function. They're usually involved in a tactile, or a touch sensory function. Some of them are used in chemoreception, which would be like a smell or taste. And speaking of sensory organs, we got one more here, which we would be remiss to not mention, uhm, which is the large compound eye here. So, I've made an incision on the dorsal surface of this grasshopper. Ok? And I've peeled back the exoskeleton. And before I go digging too much, uhm, it's going to be difficult to see many structures, but on these individuals it's very easy to see, uhm, all of these very large and pronounced little sort of tubular looking structures. There's one right there. Those are all eggs."

Reason 3. When you eat an insect, you eat ALL of it. In particular, its digestive system, including its stomach, intestine, rectum, anus, and whatever partially digested food is still in there. [Biology student] "Now, if we move on to the digestive system... there is a mouth, of course, we talked about that being down here, ok? The mouth opens into a small pharynx, ok? And then it basically opens up into this large, dark, thin-walled sack right here, ok? This is the crop. Ok, so this is basically a food storage pouch right in here. So ... getting to the stomach, that's what we find next, this thin-walled, sort of darker colored sack right here, which I've just broken a little bit, that, uhm, is the stomach, all in here, ok? Below the stomach we find this slightly darker and a bit more muscular tube right here. That is the intestine. And the intestine opens into a short rectum and an anus."

Reason 4. Edible insects are perfectly safe to eat. Nonetheless, we tend to associate insects with death and disease. Even if we know that eating some insects is harmless, this association is difficult to overcome. [Nature film maker] "Just a few days ago, one of those gaur was killed by a tiger in the night. This carcass is now probably about five days old, and, as you can see, absolutely riving with maggots of many different species."

E.2 Instructions for Experiment 1

Enter the following number into the field below

847924

WELCOME

This study is run in collaboration with the University of Zurich.

The study will take between 60 and 90 minutes to complete. You will be able to continue with the study only once you pass two **comprehension checks** about the instructions. Hence, please pay attention.

You will receive a completion payment of € 9 for completing this study.

In addition, you receive an allowance of € 6. Depending on the decisions you make in this study and on luck, that allowance will be increased, reduced, or remain unchanged.

We will pay your total earnings to you through PayPal. We will initiate payment no more than 4 business days after your completion of this study.

*This study involves **NO DECEPTION**. Whenever we tell you that something will happen if you make some decision, then that thing will happen if you make that decision. If we tell you something will happen with an X% chance, then that thing will happen with an exactly X% chance.*

At the end of the survey, you will click a link to a separate survey. On that link, you will enter the information required to process the study payment.

Your study data are completely anonymous. The research team will not have access to the data required for processing payment.

If you have any questions about any part of this study, please send an email to sandro.ambuehl@econ.uzh.ch

By clicking the "continue" button below, you consent to participating in this decision making study.

Your survey ID

The ID you will need to enter for the processing of payments is

1606224523032

We will show you this ID again at the end of the experiment where you will receive the link to enter your payment information.

Please take a screenshot or a picture of this ID.

If, during the course of the experiment, your computer experiences a problem, it is safe to close this window / your browser. To continue with the survey, simply click on the study link in the invitation email again. The survey will automatically continue at the point where you left off.

Parts and Choices

This study has two parts. The first part has 18 rounds. The second part has 26 rounds.

One part and **one single decision** from that part, determined randomly, will **entirely determine your payment** from this study.

Hence, you should make each decision as if it is the one that counts - because it might be!

The ventures

In each round of part A you will decide whether to participate in a venture. Ventures can be successes or failures. You will want to participate in successful ventures, and you will want to avoid participating in failures. We'll explain details on the next page.

At the beginning of each stage, you will randomly draw one of 20 ventures, pictured like this:

- ✔ = venture that will succeed
- ✘ = venture that will fail

In each round, 20% of the ventures will be successes, the remaining 80% will be failures.

You will first click a button that will hide whether a given venture is a success or a failure. After clicking, each venture will look grey. Then you will shuffle the ventures three times and pick one at random. That venture will be yours for the round. If you participate in that venture and it is a success, you will gain money. If you participate in it but it is a failure, you will lose money.

Click to see how this will work.

Hide success / failure information

You will first click a button that will hide whether a given venture is a success or a failure. After clicking, each venture will look grey. Then you will shuffle the ventures three times and pick one at random. That venture will be yours for the round. If you participate in that venture and it is a success, you will gain money. If you participate in it but it is a failure, you will lose money.

Click to see how this will work.

If this were a real round, the venture in the rectangle would be yours. Click Next to continue.

How much money will I gain or lose from a venture?

If you decide not to participate in the venture, your study payment will neither grow nor shrink.

If you decide to participate, the following will happen.

First, you will **receive a payment \$X** for participating in the venture (the amount X will vary across the rounds). We call this the **venture participation payment**.

Second, the venture will either succeed or fail.

- If the venture succeeds, nothing else will happen.
- If the venture fails, you have to **pay damages of € 6**.

Different ventures in different rounds

In some rounds, *if the venture fails, you can keep the venture participation payment* and use it to pay towards the damages. For instance, if the venture participation payment is € 1, and the venture succeeds, you get to keep the € 1. If the venture fails, you'll have to pay the damages of € 6. Using the venture-participation payment of € 1, you will be left with a total loss of $(6 - 1) = € 5$.

In other rounds, *if the venture fails, you will lose the venture participation payment, and you will still have to pay the damages*. In the example above, if the venture fails, you would then be left with a total loss of € 6.

We will tell you in each round, whether you will keep the venture participation payment or lose it in the event that you participate in a failing venture.

The amount of damages, € 6, will be the same in every round.

Chance that the venture will succeed

In each round, you will decide about a new venture.

Any venture will **succeed** with **20%** chance,
any venture will **fail** with **80%** chance,

Hence, any venture is four times as likely to fail as it is to succeed.

Whether the venture in a given round will succeed or fail does not depend on anything that happened in any other round.

Important: The advisors

In each round, before you decide whether to participate in the venture, you can consult an advisor. The advisor will either recommend that you participate in the venture, or he will recommend that you don't participate.

There are two advisors, the bold advisor, and the cautious advisor. Both advisors know something about the venture, but neither can perfectly predict whether your venture is a success or a failure.

←

→

←

→

The **bold advisor** 😊 is likely to tell you to participate:

- If the venture is a success, the bold advisor will certainly advise you to participate in the venture.
- If the venture is a failure, there is a 50% chance that he nonetheless (and mistakenly) advises you to participate in the venture.

The **cautious advisor** 😐 is likely to tell you not to participate in the venture.

- If the venture is a failure, the cautious advisor will certainly advise that you do *not* participate in the venture.
- If the venture is a success, there is a 50% chance that the cautious advisor nonetheless (and mistakenly) advises you *not* to participate in the venture.

In each round, you will choose whether to get a recommendation from the bold advisor or from the cautious advisor. In each round, you can select only one advisor.

Like in real life, choosing your advisor carefully in each round is important.

Thinking through this decision carefully will help you get the kind of payment you would like from this study, on average.

In some rounds, you will know how the magnitude of the venture participation payment **before** you choose your advisor. In other rounds, you will learn that number only **after** picking an advisor. (In each round, the venture participation payment will lie between € 0 and € 5, determined randomly by the computer, drawn without replacement.)

Summary

Here's how each round proceeds:

1. You learn how much money you will get as venture-participation payment if you decide to participate in the venture.
2. You choose an advisor. (In some rounds you'll do that before step 1.)
3. You get a recommendation from the advisor you've chosen.
4. You decide whether or not to participate in the venture.

Recall:

In each round, the chance that the venture in that round is a success is 20%.

Comprehension check 1 of 2

Please mark all the correct statements, and only the correct statements. If you do not succeed, please review the instructions until you can correctly mark all the statements (click the back button).

If you feel you have understood the instructions, but still cannot continue, send an email to sandro.ambuehl@econ.uzh.ch.

☐ Whatever I gain or lose will be added to or discounted from the study payment that I would otherwise get	☐ My choice of advisor is irrelevant	☐ The venture will succeed whenever the advisor tells me to participate in it.
☐ The venture will fail whenever the advisor tells me not to participate in it.	☐ The bold advisor 😊 will more likely recommend I participate in the venture than the cautious advisor	☐ IF the venture will FAIL, the cautious advisor 😊 will certainly recommend I refuse it, but the bold advisor 😊 will, with a 50% chance, mistakenly recommend I take it.
☐ The cautious advisor 😊 will more likely advise me not to participate in the venture than the bold advisor	☐ The chance that venture will succeed is 20% in each round.	☐ IF the venture will SUCCEED, the bold advisor 😊 will certainly recommend I take it, but the cautious advisor 😊 will, with a 50% chance, mistakenly recommend that I do not participate in the venture.

Your confidence about the venture

In each round, after you have decided whether to participate in the venture, we will ask you how certain you are that the venture in that round will succeed (or would succeed, in case you rejected it), in a question such as this:

This venture will ...												
definitely succeed	most likely succeed	very likely succeed	quite likely succeed	fairly likely succeed	slightly more likely succeed than fail	slightly more likely fail than succeed	fairly likely fail	quite likely fail	very likely fail	most likely fail	definitely fail	
succeed with 100% chance	succeed with 90- 99% chance	succeed with 80- 89% chance	succeed with 70- 79% chance	succeed with 60- 69% chance	succeed with 50- 59% chance	succeed with 40- 49% chance	succeed with 30- 39% chance	succeed with 20- 29% chance	succeed with 10- 19% chance	succeed with 1- 10% chance	succeed with 0% chance	
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

Your answer to this question may fully determine your payment (see below). In this case, depending on your answer and on luck, your payment will rise by € 5 or fall by € 5.

The payment procedure is designed such that it in your best interest select the answer you genuinely believe is true.

If you believe, for example, that it is about 75% likely that the venture will succeed, it is in your best interest to select “quite likely succeed (70 - 80%)”. If you believe, for example that it is about 25% likely that the venture will succeed (that is, you believe it is about 75% likely that the venture will fail), then it is in your best interest to select “quite likely fail (20 - 30%)”.

If this part determines your payment, then you will be paid for **EITHER** for your decision whether or not to participate in the venture in a random round (with a 4 in 5 chance), **OR** for the answer you give to the question above in a random round (with a 1 in 5 chance), but **never for both**.

[Click here for further details about the payment mechanism](#)

[Click here for further details about the payment mechanism](#)

Here are the details about the payment mechanism and about WHY it is in your best interest to answer this questions according to you true beliefs.

In case you are interested, here are the details about the payment mechanism and about WHY it is in your best interest to answer this questions according to you true beliefs.

The payment procedure works like this. (This procedure is known as the "probabilistic quadratic scoring rule." The first version of it was introduced by Glenn W. Brier in 1950.)

For most choices you can select, there is a range of chances (for example 50 - 60%). Your payment is determined by the number in middle of the range you select (for example 55%, if you select the range 50 - 60 %). Suppose you select a choice for which the middle of the range is some number X. The computer will randomly and secretly draw another number Y between 0 and 100. If the number the computer randomly draws is the larger one, that is if $Y > X$, then you will win € 5 with chance Y in 100 (and lose € 5 if you don't win). If the number you stated is the larger one, that is, if $X > Y$, then you will win if the venture will succeed. So if X is your genuine belief that the venture will succeed, you will win with chance X or with chance Y, whichever of the two is larger.

Why is it in my best interest to answer this question according to my genuine beliefs?

Simply, the reason is that you lower your chance of winning if you state a chance that is lower than you genuinely believe, and you also lower your chance of winning if you state something that is higher than you genuinely believe. So the best you can do is state what you genuinely believe.

(continues on next page)

(continued from previous page)

Simply, the reason is that you lower your chance of winning if you state a chance that is lower than you genuinely believe, and you also lower your chance of winning if you state something that is higher than you genuinely believe. So the best you can do is state what you genuinely believe.

To see why, it's best to go through an example.

Here's why you lose from stating a chance that is higher than you genuinely think is true. For example, suppose you genuinely believe the chance that the venture will succeed is 60%, but in the survey question you select something higher, say 90%. Suppose the number Y that the computer draws is between 60% and 90%, let's say it is 80%. This is lower than what you've indicated in the survey question (you've indicated 90%), so you will not play the computers' bet. Instead, you will win if the venture succeeds, which you genuinely think is true with only a 60% chance. The computers' bet would have given you a higher, 80%, chance instead. Hence, you hurt your chance of winning by stating the venture will more likely succeed than you genuinely think.

And here's why you lose from stating a lower chance than you genuinely think is true. For example, suppose again you genuinely believe the chance that the venture will succeed is 60%, but in the survey question you select a lower chance, say 10%. Suppose the number Y that the computer draws is between 10% and 60%, let's say it is 30%. That is higher than what you told us (which is 10%), so you will play the computers' bet and win with chance 30%. That is lower than if you had instead won depending on whether the venture will succeed, which, according to your genuine belief, has a 60% chance. Hence, you hurt your chance of winning by indicating a lower chance than you genuinely think.

Therefore, the best you can possibly do is to select exactly the answer that corresponds to your genuine beliefs.

If you have any questions about this payment mechanism, please send an email to sandro.ambuehl@econ.uzh.ch.

Recall:

Your payment is determined by **one single decision from one single round** of this study, determined at random.

Hence, you should make each decision as if it is the one that counts - because it might be!

Comprehension check 2 of 2

Please mark all the correct statements, and only the correct statements.

If you do not succeed, please review the instructions until you can correctly mark all the statements.

- ☐ At the end of the study, the computer will randomly select ONE round. My payment will depend on what happened during that SINGLE round alone.
- ☐ When I am asked about my confidence about the venture, I will earn the most if I state something a little higher than I truly think
- ☐ At the end of the experiment, the computer will randomly select TWO rounds. My payment will depend on what happened during BOTH of these rounds.
- ☐ At the end of the study, the computer will sum up all my wins and losses over all the rounds. My payment will depend on the total amount of those wins and losses.
- ☐ When I am asked about my confidence about the venture, I will earn the most if I state exactly what I truly think
- ☐ When I am asked about my confidence about the venture, I will earn the most if I state something a little lower than I truly think

Part A of this study starts now.

The following six screenshots display one arbitrary round. The remaining rounds proceed in a similar way.

Round 1 of 16

If you participate in the venture, you will get a venture participation payment of

€ 2

If the venture fails, you will have to pay **damages** of € 6. But you can **keep the venture participation payment**, and use it to pay towards the damages.

If you do not participate in the venture, your bonus will be unaffected.

Randomly draw the venture for this round

✓ = venture that will succeed

✗ = venture that will fail

Shuffle the ventures three times. Then click the button to randomly select one of the ventures.

Hide success / failure information

Randomly draw the venture for this round

- ✔ = venture that will succeed
- ✖ = venture that will fail

Shuffle the ventures three times. Then click the button to randomly select one of the ventures.

Shuffle all ventures

Randomly draw the venture for this round

- ✔ = venture that will succeed
- ✖ = venture that will fail

Shuffle the ventures three times. Then click the button to randomly select one of the ventures.

The venture in the rectangle is yours in this round

Select your advisor

If you participate in the venture, you will
get € 2

Recall: the venture will succeed with 50% chance, it will fail with 50% chance.

IF...	Ask the bold advisor.	Ask the cautious advisor.
Advice you'll get	Advice you'll get	Advice you'll get
If venture succeeds Keep €2	100% chance: "Take it!" (good advice)	50% chance: "Take it!" (good advice) 50% chance: "Don't take it!" (too pessimistic)
If venture fails Keep €2 Pay damages €6 Net loss €4	50% chance: "Don't take it!" (good advice) 50% chance: "Take it!" (too optimistic)	100% chance: "Don't take it!" (good advice)

If you participate in the venture, you will
get € 2

If the venture fails, you will pay **damages of € 6** towards which you can use the venture-participation payment (so you **lose a total of € 4**).

Recommendation

The **cautious advisor** recommends:

Your decision:

Participate
in the venture
☐

Don't participate
in the venture
☐

How sure are you whether the venture in this round would succeed, given the recommendation you got?

(The **cautious advisor** recommended that you don't participate).

This venture will ...

definitely fail	most likely fail	very likely fail	quite likely fail	fairly likely fail	slightly more likely fail than succeed	slightly more likely succeed than fail	fairly likely succeed	quite likely succeed	very likely succeed	most likely succeed	definitely succeed
succeed with 0% chance	succeed with 1-10% chance	succeed with 10-19% chance	succeed with 20-29% chance	succeed with 30-39% chance	succeed with 40-49% chance	succeed with 50-59% chance	succeed with 60-69% chance	succeed with 70-79% chance	succeed with 80-89% chance	succeed with 90-99% chance	succeed with 100% chance
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

The number below each option is the percentage chance with which you think that the venture will succeed.

Recall that it is your best interest to state your genuine belief, that there's a 1 in 5 chance that you will be paid according to this type of decision, and that depending on your decision, you could win or lose € 5.

If you like, you may now go back to change your decision on whether to participate in the venture or not.

Part B of this study.

This part of this study has 26 rounds. In each of them, you will face several decisions such as the one below. In each decision, you choose between a lottery and a sure amount of money. In some decisions, the sure amount of money is a gain, in other decisions, the sure amount of money is a loss.

Here is an example of one such decision (X and Y will be replaced with specific numbers):

Which of the following two options do you prefer?

Receive \$X with p% chance,
lose \$Y with q% chance.
☐

Lose \$1 for sure
☐

Payment for this part

If your study payments are determined according to this part of the study, here's what will happen.

The computer will randomly select one of the decisions you have made in this part. If you chose the gamble, you will play the gamble. Gains will be added to your budget of € 6, losses will be discounted. If you rejected the gamble, your payment will grow or shrink by the money amount you chose instead of the gamble.

Hence, you should make every decision as if it is the one that counts, because it might be!

In each round, you will proceed through a list of questions like this:

Question 1.1

Which of the following two options do you prefer?

Lose \$5 with 90% chance
receive \$0 with 10% chance.

Lose \$4 for sure

☐

☐

Question 1.2

Which of the following two options do you prefer?

Lose \$5 with 90% chance
receive \$0 with 10% chance.

Receive \$5 for sure

☐

☐

Question 1.3

...

Important: **Each row of the list is a *separate* decision.** All of your decisions are choices between a gamble and a sure gain (or loss) of money.

In particular, in a list like the above, your decision is NOT whether you prefer receiving \$5 to losing \$4 for sure (we already know the answer to that question).

Instead, in the above table you see two questions. In the first question (top half), you decide whether you'd rather play the gamble or whether you'd rather lose \$4 for sure. In the second question (bottom half) you choose whether you'd rather play the gamble or whether you'd rather receive \$5 for sure.

Questions in which both alternatives are unattractive

In rows of some tables you may dislike both the gamble and the sure amount. For instance, in the first row of the list above, both options involve losing money. In these cases, too, it will be in your own best interest to choose whichever option you like more (that is, whichever option you dislike less). The reason is that if that first row happens to be randomly selected to determine your study payment, you'll be better off having selected the option you dislike less of the two.

Your choices have *no bearing* on which round or which row of which list will be selected for determining your payment.

That selection is entirely random.

Your decisions for part B start now.

In each question, choose the option you genuinely prefer.

In each question you will decide between losing / receiving a certain amount of money and the following lottery (which is the same in all questions on this page):

Receive € 4 with 11% chance,
lose € 2 with 89% chance.

Each row is a separate decision!

Question 1.1

Which of the following **two** options do you prefer?

Receive € 4 with 11% chance,
lose € 2 with 89% chance.

☐

Receive € 4.5 for sure

☐

Question 1.2

Which of the following **two** options do you prefer?

Receive € 4 with 11% chance,
lose € 2 with 89% chance.

☐

Lose € 3 for sure

☐

Question 1.3

Which of the following **two** options do you prefer?

Receive € 4 with 11% chance,
lose € 2 with 89% chance.

☐

Receive € 0.5 for sure

☐

This page includes an additional 8 questions of a similar format, displayed in random order. The survey continues with 25 additional similar pages.

You have completed the decision making part of this study.

We would now like to ask you some questions. Your answers will not affect your payment.

A bat and a ball cost \$5.50 in total. The bat costs \$5.00 more than the ball.
How many cents does the ball cost?

If it takes 10 machines 10 minutes to make 10 widgets,
how many minutes would it take 100 machines to make 100 widgets?

In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 24 days for the
patch to cover the entire lake, how many days would it take for the patch to cover half of the lake?

If you're running a race and you pass the person in second place, what place are you in?

- ☐ First
☐ Second
☐ Third

A farmer had 15 sheep and all but 8 died. How many are left?

Emily's father has three daughters. The first two are named April and May. What is the third daughter's
name?

How many cubic meters of dirt are there in a hole that is 3 meters deep, 3 meters wide, and
3 meters long?

Questions about yourself

Please describe yourself truthfully.

I see myself as someone who...

	Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly
... Can be somewhat careless	☐	☐	☐	☐	☐
... Tends to be disorganized	☐	☐	☐	☐	☐
... Perseveres until the task is finished	☐	☐	☐	☐	☐
... Does a thorough job	☐	☐	☐	☐	☐
... Is easily distracted	☐	☐	☐	☐	☐
... Is a reliable worker	☐	☐	☐	☐	☐
... Makes plans and follows through with them	☐	☐	☐	☐	☐
... Tends to be lazy	☐	☐	☐	☐	☐
... Does things efficiently	☐	☐	☐	☐	☐

What is your gender?

- ☐ Male
- ☐ Female
- ☐ Other (e.g. genderqueer)

What is your age?

What is your native language?

- ☐ German
- ☐ English
- ☐ Other. Please indicate.

What type of degree are you currently working towards?

- ☐ Undergraduate degree (BA, BSc, etc.)
- ☐ Masters degree (MA, MSc, etc.)
- ☐ Doctorate
- ☐ I am not a student

Have you ever taken a course in statistics?

- ☐ Yes
- ☐ No

Have you ever taken a course in probability theory?

- ☐ Yes
- ☐ No

If you know it, please enter the name of the following mathematical formula (i.e. the last name of the person after whom the result is named):

What type of degree are you currently working towards?

- ☐ Undergraduate degree (BA, BSc, etc.)
- ☐ Masters degree (MA, MSc, etc.)
- ☐ Doctorate
- ☐ I am not a student

Have you ever taken a course in statistics?

- ☐ Yes
- ☐ No

Have you ever taken a course in probability theory?

- ☐ Yes
- ☐ No

If you know it, please enter the name of the following mathematical formula (i.e. the last name of the person after whom the result is named):

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

At which faculty do you study?

- ☐ Wirtschafts- und Sozialwissenschaftliche Fakultät
- ☐ Rechtswissenschaftliche Fakultät
- ☐ Medizinische Fakultät
- ☐ Philosophische Fakultät
- ☐ Mathematisch-naturwissenschaftliche Fakultät
- ☐ Humanwissenschaftliche Fakultät
- ☐ I am not a student

In which Bundesland have you completed your Abitur?

- ☐ Baden-Württemberg
- ☐ Bayern
- ☐ Berlin
- ☐ Brandenburg
- ☐ Bremen
- ☐ Hamburg
- ☐ Hessen
- ☐ Mecklenburg-Vorpommern
- ☐ Niedersachsen
- ☐ Nordrhein-Westfalen
- ☐ Rheinland-Pfalz
- ☐ Saarland
- ☐ Sachsen
- ☐ Sachsen-Anhalt
- ☐ Schleswig-Holstein
- ☐ Thüringen
- ☐ I have completed the International Abitur
- ☐ I do not have an Abitur
- ☐ I prefer not to say

What is the average of your grades in the Abitur?

Which grade did you achieve in your Abitur in **mathematics**?

What grade did you achieve in your Abitur in **German literature and language**?

Did you complete **mathematics** as Leistungskurs in your Abitur?

- ☐ Yes
- ☐ No
- ☐ I do not have an Abitur

Did you complete **German literature and language** as Leistungskurs in your Abitur?

- ☐ Yes
- ☐ No
- ☐ I do not have an Abitur

How much money do you spend per month, on average (including rent, food, transportation, etc.)?

How much money do you earn per month, on average, through your own work?

How much money do you receive from your parents each month?

What is the approximate wealth of your parents (including houses and other real estate)?

Your payment for this study.

As explained in the instructions, the computer randomly selected one round of the experiment to determine your payment.

You are paid by one of your decisions whether to take one of the gambles. You decided to take the gamble. You won. Hence, X will be added to your allowance.

Your total bonus from this study is therefore €Y

In addition, you will receive the completion payment of € 9.

Click Next so you can enter your information for receiving your study payment.

E.3 Instructions for Experiment 2

This section reproduces the instructions for the Treatment condition. The instructions for the Control condition are identical, except that no video is mentioned.

Welcome

This is a study of individual decision-making and behavior. Money earned will be paid to you in cash at the end of this experiment.

YOUR DECISIONS IN THIS EXPERIMENT

Your decisions in this experiment determine your payment for this study, and what kind and amount of food products you will eat. There is *no deception* in this experiment: Whenever we tell you that something will happen if you choose a given option, then that thing *will* happen if you choose that option. Whenever we tell you that the computer will do something with a specified chance, then the computer *will* do that thing with exactly that chance. Anything else would violate the IRB protocol under which we run this study (IRB34001).

WHAT YOU MAY AND MAY NOT DO

On your desk you find a pen, paper, and a calculator. You may make use of any of them if you wish. You are not required to do so.

This is a study of **individual** decision-making, which means you are **not** allowed to talk during the study. If you have any questions, please raise your hand and we will come and answer your questions privately.

Please do not use cell phones or other electronic devices until after the study is over. Also, do not browse the internet, or check your emails. If we do find you doing one of these things, the rules of the study require us to deduct \$20 from your payment. The only exception to this rule concerns the calculators that are provided on your desk.

Payment

This study has three parts. Parts 1 and 3 concern food items. Part 2 does not concern food items. Each part has several steps or rounds.

You will be paid at least \$15 if you complete this experiment (this includes your showup payment of \$5), regardless of your choices.

In addition, you receive payment for each part.

Part 1

1. You automatically receive \$20 if you complete part 1 *and* follow through with the decisions you make in this part.
2. Depending on your choices and luck, you may win additional money.

Part 2

You may win or lose up to \$10 from in this part. This depends both on your choices and on luck. Any earnings will be added to your payment from part 1. Any losses will be discounted from your payment in part 1.

Part 3

Depending on your choices and luck, you may earn additional money from this part.

At the end of this experiment, the computer will randomly select *exactly one* decision from part 1, *exactly one* decision from part 2, and *at most one* decision from part 3. Only these three decisions will determine your payment and consumption of food items in this study. Hence,

*In each part, you should make each decision
as if it is the only one that counts—because it might be!*

The payment for participation in this study will be paid to you in cash at the end of this study.

Decision Lists

All parts of this experiment involve decision lists similar to the one below. What X is will vary across different decision lists.

Do X. In exchange receive \$0.	☐ ☐	Do not participate in the transaction on the left
Do X. In exchange receive \$2.	☐ ☐	Do not participate in the transaction on the left
Do X. In exchange receive \$4.	☐ ☐	Do not participate in the transaction on the left
Do X. In exchange receive \$6.	☐ ☐	Do not participate in the transaction on the left
Do X. In exchange receive \$8.	☐ ☐	Do not participate in the transaction on the left

Each line is a separate decision.

On each line, you are asked to select either the option on the right, or the option on the left.

If at the end of the study, the computer randomly chooses your payment from this study to be determined by a decision list, here's what will happen: The computer will randomly draw one line from that price list. Your payment will be determined according to the decision you made on that line. Your choices have *absolutely no* bearing on the line the computer may select.

*Hence it is in your best interest to select on each line
the option you genuinely prefer!*

For instance, suppose that you filled in the decision list like this:

Do X. In exchange receive \$0.	☐ ☒	Do not participate in the transaction on the left
Do X. In exchange receive \$2.	☐ ☒	Do not participate in the transaction on the left
Do X. In exchange receive \$4.	☒ ☐	Do not participate in the transaction on the left
Do X. In exchange receive \$6.	☒ ☐	Do not participate in the transaction on the left
Do X. In exchange receive \$8.	☒ ☐	Do not participate in the transaction on the left

Suppose that the computer randomly selects the decision on the first line to be carried out. On this line you selected the option on the right. Therefore, you will not do X, and you will neither earn or nor lose any additional money.

Suppose instead that the computer randomly selects the third line for payment. On that line, you selected the option on the left. Therefore, you will do X, and \$4 will be added to your earnings from this experiment.

Most people begin a decision list by choosing the option on the right and then switch to choosing the option on the left (as in the example above). For your convenience, once you click a given option, the computer automatically selects the option on the right on all lines above the line you clicked, and automatically selects the option on the left on all lines below the line you clicked.

Part 1 (involving food items)

Most of the decisions you will make in part 1 are about whether or not you want to consume some food item in exchange for a specified amount of money. Many of the food items may be unfamiliar and / or unappealing to you.

IMPORTANT:

*You will always be given the choice not to eat any food item at all! You will **never** be forced to eat any food item!*

All food items will be eaten in the small waiting room next to this room. At each time, there will be at most one participant in that room. Hence, except for the study staff, consumption of the food items occurs in private.

Recall: This study involves *no deception*. Whenever we tell you that something will happen if you choose a given option, then that thing *will* happen if you choose that option. In particular, if you decide to eat a food item in exchange for a specified amount of money (and that decision is selected to be carried out by the computer), then you will get that amount only if you completely eat the food item.

What happens if during the experiment, I decide to eat some food item for some price, and at the end of the experiment, when I should eat the item, I change my mind?

You are allowed to finish the study without eating the food item. In this case, because you did not follow through with your decisions, you will not receive the \$20 you would have received had you instead completed part 1 *and* followed through with your decision.

Example

Suppose you are given the offer

"Eat [specified food item]. In exchange, receive \$5."

Suppose your decision on this offer is randomly selected to be carried out at the end of this study.

If you **accept** the offer:

- If, at the end of this study, you **follow through** with your decision and eat the food item, you will receive the \$5 you were promised in the offer. In addition, you will receive \$20 for following through with your decision. Hence, your **total** payment from part 1 will be **\$25**.
- If, at the end of this study, you **decide not to eat** the food item, even though you have accepted the offer, then you will *not* receive the \$5 you were promised in the offer. You will also forfeit the \$20 you would have received for following through with your decision. Hence, your **total** payment from part 1 will be **\$0**.

If you **reject** the offer:

- You will receive the \$20 you are promised for following through with your decision. Because you rejected the offer, you will not eat a food item, and you will not receive any additional money for part 1. Hence, your **total** payment for part 1 will be **\$20**.

Main Decisions

Some participants will be presented with offers to eat specified food items in exchange for \$3. For instance, they will decide whether or not to accept offers such as this one:

Eat [specified food item]. In exchange, receive \$3.

Other participants will be presented with offers to eat specified food items in exchange for \$30. Those participants will decide whether or not to accept offers such as this one:

Eat [specified food item]. In exchange, receive \$30

All participants will be offered *exactly the same* food items, and in the same amount, regardless of how much they are paid. The **only** difference is that some participants will be offered a higher amount of money for eating any given food item.

There are five different food items in this study. For each of the food items you will make a decision such as the above. You will either be offered \$3 in exchange for eating the item in **all** these decisions, or you will be offered \$30 in exchange for eating the item in **all** of these decisions.

There is an **80% chance** that at the end of this study, the computer selects one of these five main decisions to be carried out.

Part 1 consists of the following 8 steps.

Step 1: Learn how much money you will be offered in the main decisions

At the beginning of the experiment, you will click a button to make the computer randomly decide whether you will be offered \$30 in all five main decisions, or whether you will be offered \$3 in all five main decisions.

The computer will immediately tell you which one got selected.

Step 2: Video

The food items in this study may be unfamiliar to you. To help you with your decisions, you will choose videos to watch.

You will watch the video you selected *before* you make any decisions regarding food items.

Step 3: Decision Lists, round 1

You will fill in five decision lists similar to the one below, one for each of the five food items.

Eat [specified food item]. In exchange receive \$0.	☐	☐	Do not participate in the transaction on the left
Eat [specified food item]. In exchange receive \$2.	☐	☐	Do not participate in the transaction on the left
Eat [specified food item]. In exchange receive \$4.	☐	☐	Do not participate in the transaction on the left
Eat [specified food item]. In exchange receive \$6.	☐	☐	Do not participate in the transaction on the left
Eat [specified food item]. In exchange receive \$8.	☐	☐	Do not participate in the transaction on the left
...	☐	☐	...

The food items will be described in words. No further information will be given.

There is a **7 percent** chance that at the end of this study, the computer selects one of the decisions you make in one of these five lists to be carried out

Step 4: Main decisions

It is at this stage that you will make the five main decisions that were described before. The food items will be described in words. No further information will be given. There is a **80% chance** that one of the decisions you make in this part will be carried out.

Step 5: Survey

You will answer some survey questions. Your answers do not affect your payment from this study.

Step 6: Food items handout

Each participant will be given five containers, each containing one of the five food items. You may closely inspect the food items.

Step 7: Decision Lists, round 2

You will fill in five decision lists as before, one for each of the five food items.

There is a **7 percent** chance that at the end of this study, the computer selects one of the decisions you make in one of these five lists to be carried out

Step 8: Additional decisions

You will make a number of additional decisions. You will learn what these are later. There is a **6 percent** chance that one of these decisions will be carried out for payment.

Summary of part 1

Let's recap. Here's how part 1 proceeds:

1. You learn how much money you will later be offered in exchange for eating food items (in step 4 below).
2. The food items in this study may be unfamiliar to you. To help you with your decisions, you may select choose videos to watch.
3. **Decision lists, round 1.** You will fill in five decision lists, one for each of five different food items. The food items will be described in words, but no further information will be given.

*There is a **7% chance** that one of these decisions will be carried out.*

4. **Main decisions.** You will make five decisions whether or not to participate a transaction such as this:

"Consume [specified food item]. In exchange, receive \$[payment]".

You will either be offered *payment* = \$30 in all five decisions, or you will be offered *payment* = \$3 for all five decisions. Again, the food items will be described in words, but no further information will be given.

*There is an **80% chance** that one of these five decisions will be carried out.*

5. You will complete a short survey.
6. Food items will be handed out, and you will be able to inspect them as closely as you wish.
7. **Decision lists, round 2.** You will fill in five decision lists, one for each of five different food items.

*There is a **7% chance** that one of these decisions will be carried out.*

8. Additional decisions. These will be explained later.

*There is a **6% chance** that one of these decisions will be carried out.*

At the very end of this study, the computer will randomly select exactly one of your decisions from this part to be carried out, according to the chances specified above.

This chart illustrates the chances that a decision from any given step of part 1 will be chosen to be carried out:

Chance that a decision from a given part will be carried out

- First round of decision lists
- Main decision
- Second round of decision lists
- Additional Decision

Parts 2 and 3

The instructions for parts 2 and 3 of this study will be displayed on your screen right before those parts begin.

You now have 5 minutes to read through these instructions on your own pace.

If you have any questions about this study, please raise your hand. We will come by and answer them privately.

Recall

*This is a study of **individual** decision making. Hence, you are not allowed to talk.*

Please do not use cell phones or other electronic devices until after the study is over. Also, do not browse the internet, or check your emails. If we do find you doing one of these things, the rules of the study require us to deduct \$20 from your payment. The only exception to this rule concerns the calculators that are provided on your desk.

Additional decisions: Estimate what other participants did

48 participants have previously completed this study.
(These participants participated on May 8 and 9, 2015)

Your task is to estimate the least amount of money for which the average participant would eat a given food item.

To be very precise: Suppose we had offered each of the previous participants just enough so (s)he would eat the food item.

How much money would we have spent, on average, per participant, for each food item?

Since all of the previous participants have completed the same decision lists as you, we know the correct answer to this question. We calculated the correct answer only using the choices participants had made BEFORE the food items had been handed out.

You will answer 15 questions in this stage.

Payment for this stage

There is a 6% chance that **one** of your decisions in this part will be carried out.

We told you in the beginning that you would receive \$20 for following through with your decisions. If this stage of the experiment you do not make any decisions regarding food items. Instead, you will get these \$20 if your estimate on the question that is selected for payment is perfect.

If your estimate is off, we will discount money from these \$20. Specifically, for each \$1 your estimate differs from the true amount, we will discount \$0.50 from your payment.

For instance, if you underestimate the true average by \$10, we will discount $10 * \$0.50 = \5 from your payment, and you will earn \$15 for part 1 of this experiment. If you overestimate the true average by \$15, we will discount $15 * \$0.50 = \7.50 from your payment, and you will earn \$12.50 for part 1 of this experiment.

If you have any questions, please raise your hand.

The following five questions are the same as those before.

However:

This time we are asking you only about those participants who were offered **\$30** in the main decisions.

That is, we are asking:

How much would the average participant **amongst those who were offered \$30 in the main decision** at least need to be paid to eat [specified food item].

The following five questions are the same as those before.

However:

This time we are asking you only about those participants who were offered **\$3** in the main decisions.

That is, we are asking:

How much would the average participant **amongst those who were offered \$3 in the main decision** at least need to be paid to eat [specified food item].