Journal of Economic Behavior and Organization 247 (2026) 107558

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Experimental methods: Revisiting the commandments of experimental economics

Uri Gneezy

UC San Diego, Rady School of Management 9500 Gilm, La Jolla 92093, United States

ARTICLE INFO

Keywords:
Experimental methods

ABSTRACT

More than half a century ago, experimental economists came up with a set of rigorous methodological principles that revolutionized the way economists think—and that have defined the field ever since. Pivotal among them are the no-deception norm, the use of abstract instructions, induced-value theory, and the use of incentives in experiments. In this paper I revisit these principles, critically examining their origins and justifications, and explore their stand in current research, given the new challenges posed by online experiments, cross-disciplinary spillovers, and the evolving landscape of behavioral research. I conclude that we have become too rigidly attached to our methodologies and should adopt a more flexible, context-sensitive, and less-dogmatic approach.

1. Introduction

The pioneers of experimental economics faced the challenge of bridging two disciplines—economics and psychology. Their goal was to introduce the experimental method, then common in psychology, into economics. While they could have directly adapted psychological methodologies to economic questions, they instead developed a set of principles tailored to the unique needs of economics research. Leaders like Vernon Smith and Charlie Plott championed four key principles that have since defined the field: *the no-deception norm*, the use of *abstract instructions, induced-value theory*, and the use of *incentives*.

The no-deception norm remains the strongest difference between experimental economics and other disciplines in social science. The reasons used for keeping the norm include ethical integrity, experimental control, and the avoidance of negative externalities. The no-deception norm has taken hold at an almost religious level. As Wilson and Isaac (2007) put it, "In economics all deception is forbidden. Reviewers are quite adamant on this point and a paper with any deception will be rejected." Most journals of economics follow this extreme position (Cook and Yamagishi, 2008). The field's flagship journal, *Experimental Economics*, for instance, does not consider papers for publication that include deception. Debates persist, however, when it comes to what is considered deception. For example, is omission of information deception? And should we apply the same rules that were created for laboratory experiments in the context of online experiments? I argue that in the modern experimental landscape, some justifications for the no-deception norm have

Gary and I shared a passion for exploring experimental methods. Over the years, we collaborated on more than a dozen papers on the subject, with more in the works. Our discussions were often filled with debate, and rather than seeking compromise, we chose to present both sides of our arguments in our writing. I know Gary would have disagreed with some aspects of this paper. Sadly, he passed away before we could have that discussion, leaving me to argue both sides in my own mind. I miss him. I thank Tim Cason, David Cooper, Jim Cox, Dan Friedman, Anya Samek, Jeroen van de Ven, and participants in the conference in honor of Gary Charness. Special thank you to Cathrine Eckel for her help in making the controversial claims I make in the paper more balanced.

E-mail address: ugneezy@ucsd.edu.

https://doi.org/10.1016/j.jebo.2026.107558
Received 17 March 2025; Received in revised form 29 November 2025; Accepted 13 April 2026
Available online 14 May 2026

become less relevant and that using a cost-benefit approach to justifying deception could be more productive, especially in online experiments.

The use of abstract instructions, another core principle, seeks to maintain experimental control by avoiding biases, variability, and demand effects associated with context-rich framing. While abstract language reduces some type of noise associated with different contexts, research demonstrates that meaningful or evocative instructions can reduce other types of noise by improving participant understanding and aligning behavior with real-world scenarios. I argue that in some cases using meaningful, contextually framed instructions can enhance participants' understanding, reduce confusion, and improve data quality, especially when the tasks involve complex reasoning or less-experienced subjects. The key takeaway is that contextual language can change behavior—but that this change might be desirable.

Induced-value theory proposes that monetary rewards should be used to isolate economic behavior. Doing so ensures replicability and control but has limitations. When compared to values influenced by intrinsic motivations, social preferences, or real-world complexities, induced values can be artificial. I argue that in some cases—when measuring preferences, for example—they may even produce results that misrepresent the behavioral tendencies under investigation. Participants enter experiments with their own intrinsic motivations, and researchers then layer additional motives on top of these. The extrinsic incentives we introduce—whether monetary or non-monetary—are different from participants' pre-existing intrinsic motives. Induced-value theory, in its standard form, largely ignores the potential interaction between these two sources of motivation. In my view, some of the most interesting findings in experimental economics arise precisely from studying how intrinsic and extrinsic motivations interact.

Lastly, as with the no-deception norm, experimental economists have embraced the use of incentives at an almost religious level. For example, when discussing financial incentives, Hertwig and Ortmann write, "Experimental economists who do not use them at all can count on not getting their results published" (2001, p390). This statement reflects the view held by some (many?) experimental economists that compensating participants with money is both necessary and sufficient to increase their effort and so improve their performance. Here, again, *Experimental Economics* has embraced the norm: The journal will consider papers for publication only if they involve incentive-based experiments. I show that while incentives sometimes increase effort and reduce noise, their impact is context-dependent. Low incentives can undermine intrinsic motivation, through the crowding-out effect, while high stakes may backfire, by causing participants to "choke under pressure." Furthermore, complex incentivized mechanisms can obscure understanding and produce worse outcomes than simpler, unincentivized approaches. I contend that this heavy reliance on monetary incentives should be reassessed, especially when it compromises clarity and participant engagement.

Although I take a critical approach in this paper to all four of these principles, I wish to emphasize their importance to the development of the field. They have promoted rigor and clarity in experimental design, and in so doing have exerted a positive and formative influence on our field. For example, economic experiments replicate at a significantly higher rate of 61% than the 36% in psychology (Open Science Collaboration. 2015; Camerer et al., 2016; Przemysław, 2021; Brodeur et al., 2025).[1] This difference is probably due to our relatively more rigorous procedures. But it is time for us to adapt to contemporary challenges. We should preserve the values that guided the design of these principles, but at the same time we should make them flexible, responsive to the demands of modern research, and open to refinement.

2. The no-deception norm in experimental economics

2.1. Origin

Early in the history of experimental economics, Vernon Smith made the case that researchers should never conduct experiments that involve deception (Smith, 1976). According to Smith (2017), he was inspired by Sidney Siegel's work (e.g., Siegel, 1961, 1964; Siegel and Fouraker, 1960). This idea had quickly become a strict norm (e.g., Smith 1982; Plott 1982; Wilde 1981) that prevails to this day (e.g., Hertwig and Ortmann, 2008; Ortmann, 2019; Charness, Samek and van de Ven, 2021).

Cooper (2014) offers an interesting insight into how the no-deception norm came to exist. He reports that "E-mails with a number of researchers, especially Charlie Plott, suggest that the norm of not using deception did not leap into existence wholly formed at some specific point in time but, instead, grew organically from Plott's early papers on public choice and preference reversals and Smith's early papers on markets. This gradual development was driven by concerns about a potential loss of control, as discussed in this note, but also reflected a need to differentiate economic experiments from those conducted by psychologists that lacked procedures and controls required for application of economic theories. Given that the target audience at the time consisted largely of economic theorists, there was need to demonstrate beyond a shadow of a doubt that economists were conducting clean tests of the relevant theory." This testimony is interesting because it ties the no-deception norm to early motivations that may have changed with time. Communication with theory is still very important, but the need to prove that experiments are relevant is hopefully less needed.

2.2. What is considered deception in experimental economics?

It is useful to start by distinguishing *lying* from *deception*. I define a lie as a statement the sender believes to be false, regardless of how it influences the receiver's beliefs. Deception can be a lie but does not have to be. It occurs when the sender's statement alters the

[1] The actual successful replication rate in economics might be higher, see Chen et al. (2021).

receiver's beliefs in a way that the sender believes is untrue. For instance, if a 58-year-old man tells a stranger that his age is closer to 40 than 50, he is lying. If, however, he says that his age is closer to 50 than 40, he is not lying—but the statement can still be deceptive, because it may lead the stranger to believe the man is in his late 40 s rather than his late 50 s.

Experimental economists are averse to lying but accept some forms of deception. This suggests that the no-deception norm does not stem from the idea that researchers are generally averse to deceiving. As Hey (1998) puts it, “There is a world of difference between not telling subjects things and telling them the wrong things. The latter is deception, the former is not.” (Note that this statement uses a different definition of deception than mine). And Hertwig and Ortmann (2008, p. 62) point out, “a consensus has emerged across disciplinary borders that intentional provision of misinformation is deception and that withholding information about research hypotheses, the range of experimental manipulations, or the like ought not to count as deception.” Ortmann (2019), writing about what this means on a practical level, proposes that deceptive acts of *commission* should be prohibited but that a minimal set of potentially deceptive acts of *omission* should be allowed.

Charness, Samek and van de Ven (2021) ask members of the field what they consider to be deception. They acknowledge the norm of no lying in experimental economics but show that what crosses the line in this context remains ambiguous. Numerous “gray area” practices of deception are in fact allowed already. These include omitting potentially relevant information or employing language that may be implicitly misleading. A norm exists—but how is it maintained? Charness, Samek and van de Ven (2021) offer their answer: “Following the norm requires an agreement between authors, reviewers, and editors about what constitutes deception.” However, as Ortmann and Hertwig (2002) argue, the benefits of a deception argument may be hard to determine, and the evaluation may be subjective.

2.3. Motivations for the no-deception rule

2.3.1. Ethical reasons

Explaining his decision to adopt the no-deception rule, Smith (2017) writes, “My real reason was that it is immoral.” The ethical argument against deception in experimental economics can take less-strict and more-strict forms. The less-strict form focuses narrowly on payments and states that if the researcher promises certain payoffs to participants, the researcher is ethically obliged to honor that commitment. The more-strict form of the ethical argument is broader and states that it is unethical to deceive regardless of payments. This is the deception-aversion argument (e.g., Gneezy, 2005).

Consider this scenario: You inform participants in an experiment that their payoffs will be determined by a coin toss. If it is heads, they win $10; if it is tails, they receive $0. A severe form of deception would involve not tossing the coin at all and simply telling participants that tails came up. In that case, the ethical breach lies in both the deception itself and the resulting consequences. But what if you skip the coin toss and tell all participants that the outcome was heads, and they won $10? That is still deception, but it benefits the participants financially, so is it unethical? Even when participants are made better off, the issue of dishonesty persists, raising questions about the trade-off between ethical transparency and participant welfare.

2.3.2. Practical reasons: maintain experimental control

2.3.2.1. Credibility. One of Smith’s (2017) arguments is that without the credibility that the no-deception norm establishes, the use of payment could become ineffective: “Otherwise, how can we expect subjects to believe anything we tell them in the instructions?” In a discussion about the timing of payments to participants in probability-matching experiments, Siegel (1964, p.152) warns that “college students are by now so familiar with the notion of an experimental ‘hoax,’ with experimenters who employ deception in the service of their experiments, that they might not believe an experimenter who told them he was going to pay them at the conclusion of their trials.” In summarizing Siegel’s thinking, Svorencík (2016) writes that deception “undermined the relationship between the experimenter and the subjects, who, in consequence, are often too preoccupied with the true agenda of the experiment or doubt the announced relation between actions and rewards, which therefore ultimately jeopardizes experimental control.”

Consider a dictator game in which a sender receives $10 and is asked if they want to give some of it to an anonymous receiver. If the sender suspects deception and believes that there is no actual receiver, they will have no reason to act based on real-world considerations of fairness or benevolence, and their behavior may change as a result of different considerations. For example, Hoffman et al. (1994) designed a dictator game with a strict firewall of anonymity that separates participants’ identities from both the experimenter and other participants, thereby eliminating reputational or other secondary gains that might otherwise influence behavior. Frohlich, Oppenheimer and Moore (2001) tested whether double-blind might make participants less trusting that there is another person down the hall. They find that anonymity changes not only the price of leaving money because of a lack of context (and social distance). The design also appears to induce doubt about whether the other participants really exist. This doubt affects behavior so that the resultant observations are difficult to interpret on the axis of the self-interest/other-regarding dimension.

This example highlights why credibility is important: If participants suspect deceit, it can distort their behavior, making it impossible to accurately measure the motivations the experiment aims to assess.

2.3.2.2. Negative externalities. In any experiment, participants’ beliefs about the experimental setting are shaped both by their prior beliefs and the information provided by the researcher. If current participants believe that a researcher has used deception in past studies, it is reasonable for the participants to suspect that other researchers in the lab might also employ deception in the current study. As a result, participants may partially discount the information provided in the instructions. The researcher then cannot know to what extent participants are discounting the information in the instructions—and that, in turn, means that we cannot know how to think about the results as they relate to the general literature. In private communication, Tim Cason summarizes this argument well: "When I was editor at *Experimental Economics*, people would often ask me if I thought we should publish papers that employ deception. My half-joking reply was, "Perhaps we would publish that paper, but we would desk reject any later paper coming out of that lab.""

This negative-externalities argument seems to be the strongest argument against deception in experiments. Hersch (2015) refers to this argument as the "argument from public bads," noting that "if a researcher deceived subjects in the past, there is a chance that subjects will discount the information that a subsequent researcher provides, thus compromising the validity of the subsequent researcher's experiment."

Jamison, Karlan and Schechter (2008) investigated the problem by intentionally deceiving some participants in one study, informing them about the deception afterward, and then observing their behavior in a subsequent study. They found that significantly fewer participants chose to return to the lab after being deceived, and that many of those who did choose to return altered their behavior during the follow-up games. These results offer support for the negative externalities problem in experimental research. Ortmann and Hertwig (2002) offer a more complex view: "First, the evidence suggests that the experience of having been deceived generates suspicion that in turn is likely to affect the judgment and decision making of a non-negligible number of participants. Second, we find little evidence for the reputational spillover effects…" See also the discussion in Svorenčík and Maas (2016) regarding the origin of these concerns.

There are some creative ways to avoid deception. For example, Eckel and Wilson (2007) designed experiments to test the impact of social status on learning in a coordination game. In the design, sometimes the agent was another player in the game and sometimes the other agent was a simulated player. To avoid deception, they told participants about this design feature and asked them to play both real and not real opponent, paying them only for the real one.

As the dictator-game example suggests, credibility is built over time. Even if in a certain experiment's credibility is not a major concern now, participants in the experiment are likely to remember that they were deceived in the past, which will affect whether and how they participate in the future. The reputation of the lab is also likely to be damaged, such that future experiments will be compromised.

2.4. Cost-benefit trade-offs and gray areas in deception

In their survey, Charness et al. (2021) asked experimental economists and students about gray areas in research practices that were related to the no-deception norm. They found significant heterogeneity in respondents' choices. The data suggest that participants perceive costs and benefits as important factors, indicating that certain practices may be considered acceptable if the research topic is significant and no alternative methods are available for collecting data—in other words, they adopt a cost-benefit approach.

Even though this cost-benefit approach is at odds with the ethical argument against deception, it is in line with current practices in experimental economics. As Hersch (2015) writes, "Experimental economists can still get their work published even when they tell their subjects things that while not explicitly false are nevertheless misleading."

Certain subsets of experimental work even allow outright lying to participants. Audit studies are a case in point. Since the 1960s, audit studies have used field experiments to investigate hard-to-detect behaviors such as racial and gender discrimination in real-world contexts (e.g., Daniel 1968). In these studies, researchers randomize one or more characteristics of individuals, either real or hypothetical, and send them into real-world settings to observe the impact of those characteristics on specific outcomes. For example, audit studies of race and gender discrimination might send research assistants out to pretend they are buying a car or applying for a job (e.g., Daniel 1968; Wienk, 1979; Ayres and Siegelman 1995). In doing so, they are lying, because they do not actually intend to buy a car or accept a job. In contrast with the lab setting, these lies are costly. For example, a car salesperson is wasting time on them instead of selling a car to a real customer. Why is it acceptable to lie in this line of research in economics but not in others? Does a different ethical code apply?

If ethical arguments cannot be identified as the source of today's no-deception norm in experimental economics, the possible sources we are left with are the experimental-control arguments outlined above: credibility and negative externalities. Charness et al. (2021) sum the situation up succinctly, writing, "Some researchers are willing to make a tradeoff between the costs and benefits of deception, so that a reasonable argument can be made for some gray-area violations being appropriate." An example of such a pragmatic approach is Cason and Wu (2019) who discuss deception in agricultural and resource economics (ARE) journals. They argue that, to maintain credibility, disciplinary ARE journals may choose to adhere to the established norms against deception. In contrast, interdisciplinary ARE journals may have more flexibility, though clear guidelines can help minimize inconsistencies in reviewer expectations.

2.5. Deception in psychological research and online experiments

Deception has long been a fundamental component of psychological research (Stricker, 1967), and there is little evidence that its use has diminished over time (Ortmann and Hertwig, 2002; 2008; Kimmel, 2001). In fact, even today researchers are being advised and encouraged to find ways to employ deception more effectively (Olson and Raz, 2021).

The current ethical guidelines of psychological societies do not prohibit the use of deception. However, they try to limit its use to studies of significant importance where deception is deemed unavoidable, and they mandate the implementation of additional safeguards to protect participants (see Hilbig and Thielmann, 2021). For example, the American Psychological Association (2017) states that deception is justified only if "effective non-deceptive alternative procedures are not feasible."

To complicate matters from the credibility and negative-externalities perspectives, psychologists who deceive their participants are required to debrief them about the deception. For example, the American Psychological Association (2017) requires that researchers "explain any deception…to participants as early as is feasible, preferably after their participation, but no later than after the data collection, and permit participants to withdraw their data." Other psychological societies take a similar stand (Hilbig and Thielmann, 2021).

This approach raises an obvious question: If deception is widely used in psychology, to what degree do its credibility and negative-externality effects influence the participants in experimental-economics experiments? Is banning deception within experimental economics alone enough to prevent any spillover of negative effects from psychology?

To assume so requires researchers to believe that there is minimal or no crossover between psychology and economics in terms of participant expectations. For this assumption to hold, Grether and Plott (1979) point out, participants would need to be able to distinguish between economics and psychology experiments and to apply skepticism only to psychology experiments. Whether participants can actually do this is an open question.

2.5.1. Online experiments

Many experimental economists have made efforts to convince participants that deception does not occur in their labs. However, this argument becomes more difficult to sustain with the shift toward online experiments, which have significantly expanded the scope of behavioral research in ways that make it hard for participants to clearly differentiate among experiments from various fields. In most cases, participants in these online experiments are not informed about the no-deception norm that is central to experimental economics, but even if they are, they might be skeptical of its enforcement, given their exposure to other fields where deception is common practice, as in the case of the many online experiments run by psychologists. In the evolving context of online studies, it is increasingly hard to expect that participants will be aware of the varying standards across different research fields, let alone capable of distinguishing among them.

If participants cannot reliably differentiate among fields or trust the guidelines being followed, their behavior may be influenced by a general skepticism rather than the specific rules of the study they are participating in. This undermines the assumption that researchers who adhere to the no-deception norm have greater control over their experiments. In fact, the rise of online experiments introduces more spillover effects from disciplines that permit deception, making it harder for researchers in fields like experimental economics to argue the integrity of their data and the authenticity of participants' responses.

2.6. Way forward

The no-deception norm in experimental economics was introduced to safeguard experimental control by ensuring credibility and avoiding negative externalities, such as participant skepticism, which could compromise the validity of future studies. However, the boundary between acceptable deception and outright lying is ambiguous, and gray areas exist. Much of the current motivation for the no-deception rule revolves around maintaining control over participant behavior and preventing the erosion of trust in the experimental process.

Experimental economists have tried but largely failed to shift the broader acceptance of deception in other disciplines (e.g., Ortmann and Hertwig, 2002). The rise of online experiments has further complicated the issue. With the increasing overlap among disciplines—especially in cases involving psychology, where deception is more common—participants may struggle to distinguish between different experimental norms. This makes it harder for experimental economists to maintain the no-deception norm and raises questions about how experimental economics can navigate the evolving landscape of behavioral research without compromising the trust and authenticity that have long been cornerstones of the field.

In my view, a balanced, "split-personality" approach is the right way forward. We should aim to preserve traditional physical-experimental economics labs, where the strict no-deception rule is upheld and the lab's reputation remains untarnished. For experiments that require careful control, such as social preferences games, it makes sense to conduct them in these physical-experiment labs, where experiments can be tightly controlled to maintain the integrity of the research. However, for other types of experiments, such as coordination games, where experimental control over social preferences is less critical, the use of online labs—despite being noisier and less controlled—may be more appropriate.

Given that participants in online experiments are familiar with looser norms from other fields, it seems unrealistic to demand stricter rules from experimental economists in these settings. Instead, adopting a more flexible, cost-benefit approach for online experiments, while maintaining stricter control in traditional labs, may be the most pragmatic path forward. In these experiments, experimental economists could adopt a no-lying norm, where outright falsehoods are prohibited but some level of deception is tolerated. This approach would acknowledge that deception is already accepted in other social-science fields and would amount to an "If you can't beat them, join them" approach.

A potentially important contribution could be the creation of an online lab that mimics the rules of the physical experimental economics lab. That is, an online platform with some strict rules that are common knowledge.[2]

[2] I thank Jeroen van de Ven for suggesting this approach.

3. Abstract instructions

3.1. Origin

Should instructions in experiments be framed using context-rich language, or should they be kept abstract to maintain experimental control? Smith (1976) had a strong opinion regarding this question. “The experimenter,” he wrote, “may be tempted to add ‘realism’ by giving the abstract experimental commodity a name such as ‘wheat’, or otherwise attempt to use instructions to simulate the alleged circumstances of a particular market. This runs the danger of so enriching induced values that control over valuation is lost… Consequently, it may be preferable not to embellish the instructions with well-intentional attempts at ‘realism’. Let the explicit reward structure be the singular source of valuation, insofar as this is possible.”

Smith’s (1976) approach dominates experimental economics, and research on the effects of context in instructions is limited. For example, Croson (2005) discusses three reasons why economists often avoid adding context in experiments. The first reason involves the generality of theory: Economic theories aim to predict behavior universally, so experiments testing these theories should be context-neutral. In private communication, Charles Noussair commented that “If a theory applies to all contexts, context can be freely included and the test should be valid. The point for me is that theories often have nothing to say about context so it should not be included in a fair test of a theory on its own terms. Including context in a test of a theory, especially when the context is thought to influence behavior, introduces elements that the theory did not model. In essence, it is a joint test of the model and the proposition that there are no effects of context.” The second reason is that context can increase data variability. For instance, using specific terms like "going to court" may introduce noise if participants have differing views, which complicates statistical analysis. Finally, context can lead to demand effects or biases, altering participant behavior in predictable ways based on the context, which risks skewing results.

This section is based on Alekseev, Charness, and Gneezy (2017) who categorize context into three types: *abstract, meaningful*, and *evocative*. Abstract context uses neutral terms like “Player A” and “Choose X,” which are unrelated to real-life experiences. Meaningful context uses terms like “buyers and sellers,” which relate to real-world scenarios but do not evoke strong emotions. Evocative context uses emotionally charged terms like “bribes” or “climate change,” which can elicit strong reactions from participants. The discussion below will be based on this categorization.

3.2. Learning and understanding

Contextual language can improve participants’ understanding and accelerate learning, especially in tasks requiring complex reasoning. A classic example is the Wason selection task (Wason and Shapiro, 1971), which is a test of contingent reasoning in which many participants fail.

Enke et al. (2023) compared two versions of the task. The abstract version used instruction with cards and color. The social-content version (adapted from Cosmides, 1989), used an example that we can relate to—drinking age in a bar. Enke et al. (2023) used three levels of incentives: hypothetical, standard lab level, and few months of salary. The incentives had no significant impact on success (we will return to this in the incentives section of this paper). But, as can be seen in Fig. 1, the difference in results between the abstract and the intuitive task was striking: For different levels of incentives, the fraction of correct answers was about 20% for the abstract instructions and 60% for the ones with context. This is in line with Griggs and Cox (1982), who showed that participants were more successful in solving logical problems when they were presented with thematic labels, such as cities or modes of transportation, than when presented with abstract labels, such as letters or numbers. Context reduced confusion and so helped participants make better decisions.

This example shows that context can change behavior, but it is possible that this change could be desirable. Do the results from the abstract task tell us more about how participants reason than those from the intuitive task?

While context often aids understanding, it can also interfere with well-established results. For example, Cooper and Kagel (2003, 2009) conducted an experiment that studied signaling games and found that using contextual language like “existing firm” and “entrant” improved strategic play compared to abstract terms like “Player A” and “Player B.” However, as they showed in a second experiment, if the context is changed mid-experiment, it can disrupt the learning process, slowing down participants’ ability to develop strategies.

These examples show that abstract versus context rich instructions can result in different outcomes, but whether this difference is desirable is case-dependent.

3.3. Changing social behavior

Contextual instructions can influence the social dynamics of games, particularly in emotionally charged scenarios that involve factors such as bribery or pollution. Abstract instructions may obscure the social components inherent in these situations, reducing the external validity of the results. For instance, Liberman et al. (2004) showed that simply labeling a prisoner’s-dilemma game as “The Community Game” instead of “The Wall Street Game” increased cooperation rates among participants. Similarly, Burnham et al. (2000) found that participants in a trust game were more trusting and trustworthy when called “partners” instead of “opponents.” Instructions could also act as a coordination device, influencing participants’ beliefs about how others will behave rather than altering their preferences. For example, in a “Community Game,” participants are more likely to expect cooperation from others, leading them to cooperate as well, helping the players to coordinate on cooperation. Ellingsen et al. (2012) provided direct evidence supporting this idea in a prisoner’s-dilemma experiment where the framing of the game impacted participants’ expectations and behavior. In contrast,

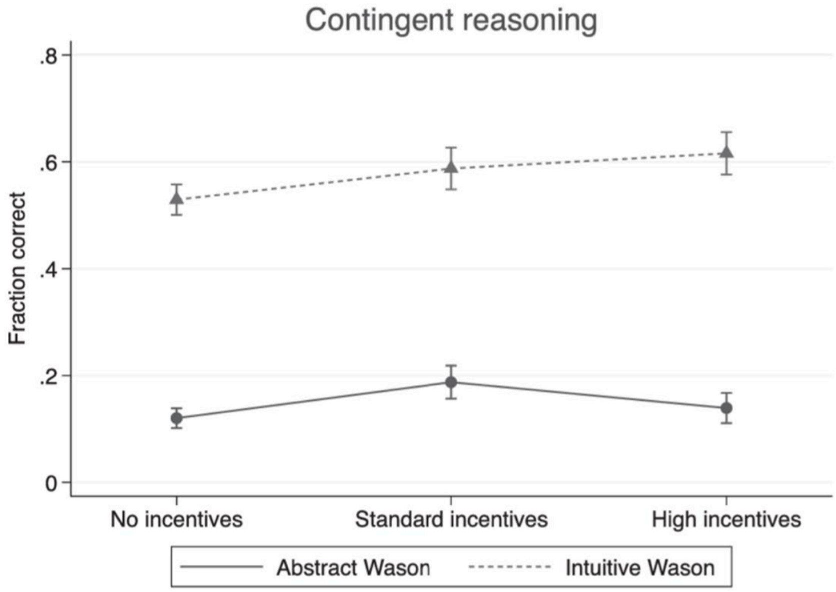

Fig. 1. (from Enke et al., 2023): Performance in the Wason Selection Task for abstract versus intuitive instructions and for three levels of incentives.

when Dufwenberg et al. (2011) studied common-pool/public-good games, they found that labeling the games as "the experiment" as opposed to "the community experiment" lowered participants' beliefs and contributions.

Barr and Serra (2009) demonstrated the impact of context in a bribery scenario in which bribes imposed negative externalities on others. Participants were less likely to offer bribes, they found, when the game was framed with meaningful terms like "citizen," "public official," and "bribe," compared to abstract terms like "Player A" and "offers." Similarly, Pevnitskaya and Ryvkin (2013) showed that framing an environmental game with terms like "pollution" and "environmental damage" led to less pollution and higher payoffs than a neutral framing. Abbink and Hennig-Schmidt (2006) explored the impact of context in a bribery scenario using a game that mimics trust and reciprocity dynamics. They found no significant differences in the frequency of bribe offers or permission decisions between rich and abstract context. Similarly, Banerjee (2014) found no impact of a change of language on behavior.

Using meaningful terms may sometimes lead to anti-social behavior. Cason and Raymond (2011) studied this effect in an environmental context involving the trade of pollution permits. In their experiment, each participant played the role of a "manager of a power plant." The plant produced a fixed amount of electricity and generated pollution that could be reduced at a cost. They compared two treatments: an environmental treatment, where participants acted as "managers" dealing with pollution permits, and a neutral treatment, where participants traded "coupons" and reported to an "inspector." They found that participants in the environmental treatment were less compliant, under-reporting their pollution levels and demonstrating less environmentally friendly behavior. This contrasts with Pevnitskaya and Ryvkin (2013).

As with learning, these are examples in which context-rich instructions can influence behavior. If one is studying bribes or environmental issues, maybe it is more informative to observe behavior based on meaningful terms?

3.4. Way forward

Meaningful context may improve understanding, which is a desirable outcome. Future research could explore how often biases and instances of "irrational behavior" are simply the result of participants struggling to understand abstract or unclear instructions. For example, Kirchler, Huber and Stöckl (2012) found a significant reduction in bubbles in the experimental asset markets when more meaningful language was used. Using instructions that are more meaningful might change our perception about these findings.

Evocative context presents more challenges. Strong emotional responses may introduce unintended biases or complicate cross-cultural comparisons. Researchers should carefully consider whether evocative language is useful for the research question, as it may increase the degrees of freedom for manipulating results. Future research can investigate this question and see which approach best fits the goal. For example, if the researcher is interested in corruption, words such as "bribe" are more appropriate because they evoke these reactions, not despite it. An interesting parameter in this investigation could be gender, as some studies (e.g., Croson and Gneezy, 2009; Ellingsen et al. 2013) have found that women react differently than men to certain contextual changes.

While abstract instruction is traditionally used for experimental control, meaningful and evocative context can be valuable for improving understanding and external validity. In some cases, using meaningful language will not change behavior. When it does, the researcher needs to decide whether it improved data quality and provided more-accurate modeling of the real-world situations they are interested in. Related to this, see the discussion in Harrison and Rutstroem (2001) who argue that people often bring contextual consideration into the lab, regardless of the level of abstraction of the instructions.

Researchers should weigh the benefits of using context against the potential for bias or confusion and tailor their instructions based on the specific goals of the experiment.

4. The induced-value theory in experimental economics

4.1. Origin

Induced-value theory was developed by Smith (1976), establishing conditions under which experimenters can create controlled incentives for participants in a laboratory setting. The idea is to induce particular values in participants by structuring rewards, usually monetary, so that participants' choices and actions reveal their preferences in a controllable way. As Smith (1976) states, "Control is the essence of experimental methodology, and in experimental exchange studies it is important that one be able to state that, as between two experiments, individual values (e.g., demand or supply) either do or do not differ in a specified way. Such control can be achieved by using a reward structure to induce prescribed monetary value on actions."

Smith (1976) argues that by inducing value, researchers can test economic theories and observe decision-making without the interference of external motivations. Participants become "economic agents" in the experiment, and researchers can observe their preferences, choices, and responses to economic incentives. Smith (1976) emphasizes three elements: *monotonicity, salience* and *dominance*.

In the following decades, the use of induced valuations was commonly applied and rarely challenged in experimental economics (e.g., Plott and Smith, 1978; Isaac and Plott, 1981; Smith, 1982; Friedman, 1991). Proponents have argued that inducing value ensures that findings from these experiments are reliable and applicable to real-world economic behavior, as participants' choices are consistent with the experimental incentives rather than personal biases or uncontrolled factors. For example, in *The Handbook of Experimental Economics*, when Roth and Kagel (2016) wrote about the importance of creating controlled environments in experiments, they implicitly endorsed the application of induced-value theory to ensure that participants' incentives align with the objectives of the study. See also Cassar and Friedman (2004), who described induced-value theory as "arguably the key methodological innovation for experimental economics" (p.26).

4.2. Limitations of induced value theory

In this section I will discuss some limitations of induced-value theory and argue that although induced value theory is useful in some cases, it misses the point in others and should not be the only approach to experimental design. This section will often refer to the following one, which discusses incentives.

4.2.1. Intrinsic motivation

Induced-value theory assumes that in experimental settings, external incentives dominate participants' motivations, which ensures that their decisions will align with the experimenter's design. However, this assumption may not hold when intrinsic motivation—the drive to perform a task for its inherent satisfaction—comes into play.

Intrinsic motivation challenges the basic premise of induced-value theory in several ways. First, it may conflict with external rewards. Smith (1976) argued that intrinsic motivation, such as a desire to perform a task well or simply enjoy it, is typically weak, inconsistent, and unlikely to sustain over time. However, subsequent research (e.g., Frey and Jegen, 2001; Gneezy and Rustichini, 2000—see the discussion of these papers in the Incentives section, below) suggests that intrinsic motivation can be strong, especially when the task is engaging, meaningful, or tied to an individual's sense of self or competence. In these cases, the presence of monetary incentives might "crowd out" intrinsic motivation. For example, individuals who enjoy solving puzzles might lose interest if paid to do so because the payment shifts their perception of the activity from an enjoyable challenge to an economic transaction. Another example is small monetary incentives. Participants might interpret such payments as a signal that the task is unimportant, reducing their effort or engagement.

In some cases, intrinsic factors might lead participants to perform well even without monetary rewards. For example, a participant receiving no monetary reward might nonetheless be driven by social norms or a sense of responsibility, or they might approach the experiment as a challenge to solve or a task to excel in. This persistence of intrinsic motivation can lead to unexpected behaviors that are not aligned with the incentives designed by the experimenter, weakening the explanatory power of induced-value theory.

4.2.2. Social preferences and ethical considerations

The assumption that monetary rewards are the primary driver of participants' decisions may not hold in experiments where moral decisions, social preferences, or other intrinsic values play a significant role. These non-monetary factors can override financial incentives, causing participants' behavior to deviate from what induced-value theory predicts.

While selfish motivations are important, in many cases social preferences are motivations that go beyond self-interest, encompassing behaviors like fairness, altruism, reciprocity, and aversion to inequity. In experiments involving such preferences, monetary incentives often take a back seat to other considerations. Participants may prioritize equitable outcomes over monetary gains (e.g., Fehr and Schmidt, 1999). In experiments where they can choose to share earnings, participants might donate part or all of their rewards to others or a charity, driven by altruistic motives or a desire to create fairness (e.g., Forsythe et al., 1994; Eckel and Grossman, 1996).

In cases where fairness is violated, individuals may punish others even when it is costly to themselves, as shown in the ultimatum game (Güth et al., 1982; van Damme et al., 2014; see the discussion in the Incentives section). Here, participants frequently reject unfair offers, sacrificing monetary rewards to uphold fairness. More generally, reciprocity—the tendency to respond to actions in kind—further complicates the dominance of monetary rewards (see the recent review in Charness and Fehr, 2025). Experiments show that participants may reward or punish others based on perceived fairness, regardless of financial consequences. For example, employees might exert extra effort for an employer who treats them well, even if monetary rewards are fixed. Participants might over-reward cooperative partners or punish selfish ones, even if those actions reduce their own earnings.

Researchers should consider the role of social preferences and how to design experiments that isolate or control for these factors. In many cases, these social preferences are what researchers are interested in Avoiding overly simplistic assumptions about utility functions and considering broader motivations beyond monetary rewards is what we are after.

4.2.3. *Effort function versus real effort*

Charness et al. (2018b) highlight a key difference between stated effort (aligned with induced-value theory) and real effort (aligned more closely with real-world contexts). Induced-value theory assumes that assigning explicit costs and benefits to (stated) effort is sufficient to control participants' motivations, but this approach may often be misguided, for several reasons.

In real-world settings, the cost of effort is often subjective, variable, and influenced by intrinsic motivations or emotional states. Real-effort tasks show that participants' motivation to exert effort often depends on factors such as interest or fatigue, which are not captured by externally imposed reward structures.

Another issue is the static nature of stated-effort experiments, where participants make a single effort decision. Real effort experiments demonstrate that motivation and effort vary dynamically over time and are influenced by changing emotional states. Induced-value theory fails to account for these time-dependent changes, which can lead to results that do not reflect real-world behavior, which often unfolds over extended periods. As in Gneezy and List's (2006) experiment, in which participants' effort initially increased with a positive wage surprise but declined over time as the emotional response to that surprise faded, highlighting the limitations of induced-value theory in capturing the emotional and temporal dynamics of effort.

In some cases, there might not be a difference between real and stated effort. For example, Dutcher et al. (2024) find that "In our data, when controlling for effort costs, we find no differences in behavior between real and stylized effort." Yet, an important criticism discussed by Charness et al. (2018b) is that even the shape of the cost function may differ between people and situations. Some participants may find it enjoyable to engage in a real-effort task (solving mazes, for example), and their enthusiasm will last for a while. Others might find it tedious or boring. Even those who find it interesting to begin with may become fatigued, in which a convex effort function is likely to result. This variability in intrinsic costs creates challenges in modelling behavior purely through induced values.

4.3. *Way forward*

Induced-value theory is a powerful tool for creating controlled experimental environments, but it is of limited value in experiments involving social, moral, or real-effort tasks, which may require non-monetary incentive structures to avoid the risk of biasing the outcomes. Another concern is the crowding out of intrinsic motivations. Understanding and accounting for these motivations is essential to designing experiments that accurately capture participants' behaviors and preferences.

Induced-value theory is used to motivate the reliance on incentives in experiments, and the discussion of both often goes together. For example, Loewenstein (1999) critiqued the use of monetary rewards to induce preferences, arguing that real-world decisions are often influenced by emotions and social context rather than purely financial considerations. He suggested that because induced-value theory may miss key aspects of human motivation, lab experiments that rely on monetary incentives alone have limited external validity. Additionally, he argued that this external-validity problem is particularly salient in experimental economics. In private communication, Jim Cox disagreed with this statement, saying "If a theoretical model fails empirically in a laboratory experiment with induced values then it has failed. If the results in a field experiment are different than in the lab, that is no reason to disregard results from a controlled laboratory experiment. Data from both experiments would be interesting and relevant to drawing conclusions about the empirical validity of theory."

Finally, an important quantification from a private communication with Dan Friedman "The topic surely deserves attention in your paper and beyond… but please don't pretend that experimental economists believe that their subjects are narrowly pecuniary homo economicus agents. Thanks to us experimentalists, few economists of any stripe believe this (except a few old-time finance types perhaps)."

5. Incentives in experimental economics

5.1. *Origin*

Smith (1976) write, "It is often possible in simple-task experiments to get satisfactory results without monetary rewards by using instructions to induce value by role-playing behavior (i.e., 'think of yourself as making a profit of such and such. when…')…but such game values are likely to be weak, erratic, and easily dominated by transactions costs, and subjects may be readily satiated with 'point' profits." As discussed above, in the induced-value section of this paper, Smith's solution is to always use incentives, with money being the king.

Smith's conclusion was inspired by earlier work, including Siegel and Goldstein (1959), who found that in a probability-matching task the probability of correct prediction increased with incentives. Their observation that an increase in payoffs led to an increase in the proportion of rational choices is often cited as evidence for the compelling power of monetary rewards. Siegel made that argument elsewhere, too. For example, Siegel and Fouraker (1960) established the notion that experimental findings are only valid when rewards are paid in cash. Ever since, whenever experimental economists talk about incentives, they have been thinking about money.

Inspired by Siegel's methods, Smith has stressed performance-based payments in his works. Svorenčík (2016) noted that although Siegel was not the first to raise the importance of monetary payments as a way to control participants' motivations (see Moscati, 2016), he was the first to systematically apply it to a series of experiments that came to be widely accepted.

The view that incentives should be used has become the norm. In one of the first experimental-economics textbooks, Davis and Holt (1993) repeatedly emphasized the need for financial incentives in experiments. They argued that because monetary payoffs eliminate concerns about reward heterogeneity and reduce performance variability, economics experiments should almost always include them. They claimed that it is more common to observe large and non-systematic deviations from the norm when financial incentives are absent, and that economic models yield better predictions when participants are paid for performance. While they did not present empirical evidence for their claim, they concluded that the failure to use salient financial rewards was a "fatal error" in economics experiments (Davis and Holt, 1993, p.33).

Since the early days of experimental economics, the effect of financial incentives on behavior has been a standing point of contention between economists and psychologists. In what follows, I discuss some of the arguments on both sides of this argument, highlighting some incentives that can be useful and others that can be harmful.

5.2. Incentives and effort

The notion that incentives influence effort is rooted in the standard model of agency theory in economics. Effort is assumed to be costly, and without extrinsic incentives, effort will be expended at the lowest acceptable level (e.g., Kreps, 1997). The agency model makes two assumptions: first, that higher incentives lead to higher effort, and second, that higher effort leads to better performance. Does this model describe behavior in laboratory settings well? In this section I will explore this question, starting with cases in which, as the model predicts, higher incentives result in higher effort and lower noise. I will then discuss the crowding-out literature, in which financial incentives reduce effort. The crowding-out effect can be avoided by increasing payoffs, but, as I will discuss in the last part of this section, very large incentives can cause participants to choke under pressure.

5.2.1. When incentives increase effort and reduce noise

Experimental economists often argue that participants will not use up their scarce cognitive resources unless they are rewarded financially for their performance. For example, Smith (1991) postulated that the reason why the psychological literature finds that people violate rationality is that psychological experiments were not providing monetary rewards. There was nothing of value worth maximizing, and therefore no reason for participants to behave in value-maximizing ways. According to this view, monetary payoffs would incentivize participants to maximize, and the greater the payoffs, the closer that responses would be to the best response. Similarly, Cox and Sadiraj (2019), in the Handbook for Experimental Economics, in the chapter titled "Incentives," put forward multiple arguments about why the use of incentives is crucial. "If a lazy or distracted subject does not understand the experimental environment or institution," they wrote at one point, "then providing larger payoffs may motivate effort leading to better understanding." At another point, they wrote, "Scaled-up payoffs can reduce decision errors from random choices among actions with trivial economic consequences."

In a similar vein, Smith and Walker (1993) reviewed 31 experimental studies that incorporated monetary rewards and found that higher incentives tend to align participants' behavior more closely with theoretical predictions. They concluded that increasing rewards not only shifts the central tendency of the data toward expected outcomes but also reduces variability, making results more consistent with economic models. Notable empirical evidence includes Binswanger (1981), Plott and Smith (1978), Fiorina and Plott (1978), and Kroll et al. (1988). The central notion is that incorporating incentives in experiments increases effort, and by doing that reduces noise. Psychology research made related findings early on. For instance, Tversky and Edwards (1966) found that paying as little as five cents when a participant makes a correct prediction—or charging them that amount if they make an incorrect prediction—is sufficient to shift the central tendency of the outcomes closer to the optimum.

Hertwig and Ortmann (2001) reported findings that supported this conclusion after reviewing 186 studies published in the *Journal of Behavioral Decision Making* between 1988 and 1997. "We conclude that," they wrote, "although payments do not guarantee optimal decisions, in many cases they bring decisions closer to the predictions of the normative models. Moreover, and equally important, they can reduce data variability substantially." However, of the 186 studies they reviewed, only ten systematically explored the effect of incentives by varying payment and non-payment conditions. Five of these ten studies concluded that there were no significant main effects of financial incentives on performance.

Wilcox (1993) showed that people spend more time thinking about what they are doing (in other words, they invest more effort) when dealing with real money rather than hypothetical rewards. Camerer and Hogarth (1999) reviewed 74 studies that examined the effects of performance-based financial incentives at different levels—none, low, and high. They found that incentives have the largest effect in tasks related to judgement and decision-making, an effect also identified by Hertwig and Ortman. Camerer and Hogarth (1999) reported that although the modal result across 74 studies was no effect on mean performance, incentives led to a convergence of data toward the optimum. This was true, they found, even in cases where incentives did not affect mean performance.

Overall, there is evidence that financial incentives sometimes lead to increased effort, which in turn causes improved task performance, mostly by reducing noise. However, this result is not universal. In many cases, incentives do not lead to more effort or better performance. For example, Bonner et al. (2000) reviewed the literature on incentives in experiments and found that incentives lead to better performance in about half of the reviewed studies. A similar meta-analysis by Jenkins and colleagues (1998) showed that incentives were not related at all to performance *quality*. Another meta-analysis by Guzzo, Jette, and Katzell (1985), on motivating techniques aimed at increasing effort, found that financial incentives had widely fluctuating effects, and that the mean effect was not significantly different from zero.

5.2.2. When incentives reduce effort: crowding out

The crowding out of incentives refers to a phenomenon where external incentives, such as monetary payments, undermine non-monetary drivers of behavior. Instead of reinforcing the desired behavior, these incentives can reduce the likelihood of participants engaging in the behavior for its own sake and can even cause unintended negative outcomes.

In the early 1970s, psychologists began to challenge the assumption that intrinsic and extrinsic motivation always work in harmony to drive behavior (e.g., Deci, 1971; Lepper, Greene, and Nisbett, 1973; Kruglanski, Alon, and Lewis, 1972). Around the same time, economists began exploring similar themes, spurred by Titmuss (1970), who argued that paying individuals to donate blood could undermine established social norms around voluntary contribution.

Extensive evidence has since demonstrated that extrinsic incentives can sometimes undermine performance. For example, Frey and Oberholzer-Gee (1997) demonstrated that offering a community large monetary compensation to host a nuclear waste site signaled that the associated risks were significant. This perceived danger made residents less willing to accept the site, even with the financial incentive. The effect depends on the existence and size of the offered incentives, as agents infer messages about the task or situation from these signals. See also the discussion in Frey and Jegen (2001).

In a laboratory setup, Gneezy and Rustichini (2000) found that even introducing small payments could lead to counterproductive outcomes. In one of their studies participants were asked to solve IQ test questions. The study compared the performance of treatments with no monetary incentive, a small payment, and a larger payment. Participants who received a small payment performed worse than those who were not paid at all. The title of the paper “Pay enough or don’t pay at all” tells the story—once incentives are introduced, higher incentives work better. For similar results, see for example Aksoy, de Oliveira and Eckel (2025) who find that workers are more responsive to higher wages in a non-profit setting, not only in for-profit cases, and the replication in Asulin, Heller and Munichor (2024). This suggests that low monetary rewards might reduce participants' motivation to perform well. There might be different reasons for the reduction in performance as a result of the use of incentives, see for example Rydval and Ortmann (2004) and Gneezy (2023).

The crowding-out effect shows that paying participants in experiments can backfire if the payment undermines intrinsic motivations or alters how participants perceive the task. There could be many reasons for this effect. Low payments can signal that the task is not important (Gneezy, 2023), reframe social or altruistic behaviors as transactions (e.g., helping in an experiment because of curiosity or goodwill versus being paid), or lead participants to focus solely on the monetary reward.

The takeaway is that external incentives (in particular small ones) may not only fail to motivate but can also reduce effort.

5.2.3. When higher effort leads to worse performance: choking under pressure

The phrase "choking under pressure" is often used when heightened motivation and effort result in a decline in performance. Baumeister (1984) argued that choking arises from "explicit monitoring," where individuals focus excessively on task execution, causing a breakdown in automated performance. Individuals overthink their actions, which prevents them from carrying them out smoothly and automatically. Supporting this notion, Beilock and Carr (2001) found that pressure-induced self-focus interferes with procedural memory, which is critical for executing well-practiced tasks like those common in sports or musical performances. Masters (1992) demonstrated that skilled performers choke when they revert to consciously monitoring their movements, a phenomenon referred to as "reinvestment theory."

An alternative pathway is the Yerkes-Dodson Law (1908), an inverse-U model in which high motivation creates excessive arousal that pushes individuals beyond their optimal performance zone. In this model, increasing motivation helps boost performance up to a point and then reduces it. A heightened state of arousal overwhelms cognitive and motor systems, making it difficult to maintain focus or execute precise actions (e.g., Mobbs et al., 2009).

The psychology experiments above used methods other than high financial incentives to increase motivation. In relation to this paper, the use of very high incentives is important. To study how very high performance-based incentives affect participant performance in an experimental setting, Ariely et al. (2009) tested choking under pressure using low, moderate, and high incentives. They found that in cases where cognitive performance is needed, very high incentives may lead to a decline in performance in tasks that draw on skills such as memory, creativity, and motor dexterity. Relatedly, Aarts et al. (2014) showed that the mere promise of a monetary bonus can impair cognitive control and have a detrimental effect on performance. See also Hickman and Metz (2015), Lee et al. (2018) and Castro et al. (2021).

Tasks in economics experiments often require participants to perform under time constraints, solve mathematical problems, or engage in strategic interactions such as bargaining, auction bidding, or investment games. These tasks typically demand a combination of cognitive processing, decision-making, and, occasionally, motor skills, all of which can be impaired under high-pressure conditions. Increased performance-contingent payments may lead participants to overthink their strategies or focus excessively on potential losses, consequently disrupting their ability to perform optimally.

As these results demonstrate, a substantial increase in incentives can lead to an increase in effort—but such incentives can be counterproductive, because higher effort doesn’t necessarily improve performance and in fact can make it worse.

5.3. Measuring preferences

In the previous section, I discussed cases in which participants do not know their preferences and need to expend effort to determine them—effort that could be boosted with incentives. This section focuses on cases in which participants do know their preferences and need to expend no effort to determine them. In theory, this should mean that no incentives are necessary. In some contexts, however, participants may choose to misrepresent their true preferences. What is the impact of financial incentives in these situations? The central question here is not about performance—since there is no agreed-upon “better” preferences—but rather about the accuracy of revealing true preferences. Can incentives help in uncovering those true preferences?

5.3.1. Social preferences

5.3.1.1. Dictator game. As discussed above, the dictator game involves one participant (the "dictator") who is endowed with a sum of money and decides whether to allocate a portion to another participant (the "receiver"). This game is widely used to assess altruism. Intuitively, it seems plausible that participants would be more generous with hypothetical money than with real money, and less generous with higher amounts. Yet dictator-game experiments have produced mixed findings regarding the effect of incentives on behavior.

Some studies suggest that participants are indeed less generous with real money, as in Sefton (1992), Forsythe et al. (1994), Dana et al. (2006), List and Cherry (2008), Engel (2011), Leibbrandt, Maitra, and Neelim (2015), and Schier, Ockenfels, and Hofmann (2016). Other studies have found no significant difference with respect to incentives. For instance, Ben-Ner et al. (2008) reported that participants in a dictator game exhibited similar behavior regardless of whether real money was involved. With stakes set at $10, the group that played for real money and the group that played hypothetically did not behave statistically differently. Additional examples that have studied different levels of incentives (rather than contrasting real and hypothetical ones) include Carpenter, Verhoogen, and Burks (2005), Heinz, Juranek, and Rau (2012), Tisserand, Cochard, and Le Gallo (2015), and Keuschnigg, Bader, and Bracher (2016). These findings are discussed in reviews by Karagözoğlu and Urhan (2017) and Larney, Rotella, and Barclay (2019).

In a meta-analysis that included experiments with stakes ranging from $0 to $130 (mean $21.77), Engel (2011) reported the following finding: “Higher stakes reduce the willingness to give. If there is more to gain, dictators keep more, not only in absolute, but also in relative terms (although the effect is very small).” In their meta-analysis, Larney et al. (2019) made a similar finding.

Despite these mixed findings, the practice of incentivizing participants with monetary compensation remains standard. The general rule of Hertwig and Ortmann (2001) is valid here: If experimental economists want to publish the results of a dictator-game experiment, they need to have run the experiment using real money.

5.3.1.2. Ultimatum game. In the ultimatum game, one player (the "proposer") is given a sum of money and decides how to split it with another player (the "responder"). The responder can either accept or reject the offer. If accepted, the money is divided as proposed; if rejected, neither player receives money. Since its introduction by Güth et al. (1982), the game has been widely used to study fairness, strategic reasoning, and the influence of social preferences on decision-making.

Empirical findings indicate that, contrary to the predictions of a purely selfish profit-maximizer model, proposers frequently offer a substantial portion of the total amount—typically 40–50%—to responders. Similarly, responders often reject low offers, prioritizing fairness and the punishment of inequity over monetary gain.

As in the dictator game, the use of real monetary incentives is standard in ultimatum-game experiments. Early papers tested the impact of incentives of behavior. For example, Forsythe et al. (1994) observed no differences in proposer behavior between $5 and $10 stakes, and Hoffman et al. (1996) found no significant differences in offers between $10 and $100 stakes. However, financial limitations often restrict "high-stake" conditions to relatively modest amounts. To overcome this, some researchers have employed hypothetical stakes, allowing for arbitrarily large amounts without incurring extra costs (e.g., Amir, Rand, and Gal, 2012; Ben-Ner, Kramer, and Levy, 2008). These hypothetical conditions assume participants can accurately predict and truthfully report their behavior under real stakes of similar magnitude.

Another strategy for addressing budget constraints involves conducting experiments in developing countries, where lower income relative to western countries enables high-stakes scenarios without excessive costs. For instance, Slonim and Roth (1998) conducted their experiment in the Slovak Republic, observing no effect on proposer behavior. Similar results were found by Cameron (1999), which also concentrated on proposers’ behavior, as they did not have statistical power to study low offers. Andersen et al. (2011) studied stakes effect on responders in Northeast India, finding that rejection rates for low proposals significantly decreased as stakes increased. Leibbrandt, Maitra, and Neelim (2015) reported similar findings. The fact that rejection rates are lower with very high stakes implies that the demand for fairness aligns with the law of demand—when the "price of fairness” increases, fewer individuals "purchase" it.

Larney et al. (2019) conducted a meta-analysis to examine the effect of stake sizes and concluded that there is no strong evidence to suggest that people offer lower proportions in high-stakes ultimatum games. The impact of stake size on proposer behavior appears negligible, even when stakes vary significantly. However, this conclusion pertains only to proposers and does not address responder behavior. Larney et al. (2019) noted that stake size had a slightly greater impact on dictator-game offers than ultimatum-game offers, though the small effect size and overlapping 95% confidence intervals indicate this finding is not statistically conclusive.

Overall, the effect of stake size on proposer behavior in dictator and ultimatum games appears to be in the intuitive direction but minimal—at least for standard lab incentives. There are many other games that measure social preferences, such as the trust game and the deception games, but overall, the conclusion seems to be consistent: Using real incentives has a relatively small effect on outcomes, in particular when considering standard laboratory pay. But this small effect is in the anticipated direction.

5.3.2. *Risk preferences*

Risk attitudes are central to economics, and measuring them is often critical to understanding behavior. Charness, Gneezy, and Imas (2013) surveyed different methods used in experiments, including incentivized and non-incentivized methods. One important conclusion they came to was that incentivizing risk measures comes with a cost—it makes the method more complex, and the complexity often results in noise rather than a true measure of risk attitude (e.g., Charness et al. (2018a)).

A technique often used to measure risk preferences is the popular Holt-Laury multiple price list method (Holt and Laury, 2002). Participants are presented with a series of binary choices between two lotteries, typically labeled Option A (the "safe" option, with a lower but more certain payoff) and Option B (the "risky" option, which a higher but less certain payoff). Participants are told that one of their choices will be randomly selected for pay. As participants move through the series, the probability of the higher payoff in Option B increases, prompting some participants to switch from Option A to Option B. The point at which a participant switches is referred to as the crossover point and is used to measure the participant's risk tolerance.

Holt and Laury (2002) ran treatments in which stakes were increased up to 90 times, both hypothetical and real, and found that the primary incentive effect was observed in the overall levels of risk aversion. Even at the low payoff level, where all prizes were below $4.00, approximately two-thirds of participants displayed risk-averse behavior. With real payoffs, risk aversion increased significantly as payoffs were scaled up by factors of 20, 50, and 90. In contrast, scaling up hypothetical payoffs had little impact on behavior, suggesting that participants' responses to hypothetical risk assessment differ fundamentally from their responses to real stakes.

However, there are problems with the Holt and Laury (2002) measures. Some are theoretical (e.g., Harrison et al., 2005) and some practical. Most notably, some (often many) participants switched more than once—a behavior that violates first-order stochastic domination. Such violations are often avoided artificially, by allowing participants to switch only once or using only their first switching point in the estimations. However, these methods clearly introduce noise generated by participants who do not understand the complex task. The problem of such measurement errors is discussed extensively in the literature (e.g., Harrison, Johnson, McInnes, Rutström, 2005; Loomes, 2005; Sahm, 2012).

Charness et al. (2018a) studied how these errors might distort the representation of economic behaviors. They examined gender differences in risk-taking—a topic where multiple price lists often show no significant differences, contrary to much of the existing literature (e.g., Charness and Gneezy, 2012). They replicated this lack of significant gender differences using the multiple price list mechanism and then compared it to a simplified approach in which participants made only one of the ten decisions, presented in isolation, with each participant making only one choice. This simpler format revealed strong gender differences, with women exhibiting greater risk aversion than men. Charness et al. (2018a) argued that the results from the simpler task better reflect underlying preferences, as the reduced complexity improves comprehension. Friedman et al. (2014) go a step further and propose to return to a simpler sort of scientific theory (not just measurement) of risky choice, one that focuses upon the potentially observable opportunities and constraints facing decision makers.

More generally, experiments on choice under risk often involve multiple decisions from individual participants, with the selection of a payoff mechanism being an essential feature of experimental design. Two questions arise. Is the payoff mechanism incentive compatible with the theoretical model? What are the behavioral properties of the mechanism? Cox, Sadiraj and Schmidt (2015) examine both questions with nine payoff mechanisms but identical decision tasks, finding substantial discrepancies in participants' risk preferences across mechanisms.

Dohmen et al. (2011) explored the relationship between risk preferences elicited through various methods and actual behaviors in the German population. They found that general unincentivized risk questions could predict risk-taking in some contexts. See also Weber et al. (2002) Domain-Specific Risk-Taking scale. Such unincentivized methods are sometimes a useful alternative to complex incentivized ones. Another reason to use hypothetical incentives is to allow studying losses in the lab (e.g., Berns, Capra, Moore and Noussair, 2007). However, Eckel, Wilson and Yang (2025), who study the stability of risk preferences over time, find that risk tolerance measured through incentivized lottery choices are more stable over time than survey-based measures.

A middle point to consider is some less ambitious but simpler incentivized methods for eliciting risk aversion, as discussed in Charness et al. (2013). The challenges of task complexity and participant understanding likely extend to other methods of eliciting economic preferences, such as time preferences. While I did not discuss time preference measures in this paper, I believe that similar conclusions are true there as well.

5.4. *Belief elicitation*

To understand the beliefs of participants, one often has to understand their behavior. There are many incentivized methods designed to elicit these beliefs. Charness et al. (2021) reviewed various approaches to belief measurement, arguing that, as in the risk-measurement example, the cost of incentivizing is often too high in terms of the complexity of instructions— to report beliefs, participants need to pay attention and exert effort (Burfurd and Wilkening, 2021; Massoni et al., 2014). However, as in the risk case, complex elicitation methods introduce their own biases. These biases arise from the inherent complexity of the elicitation process. Efforts to enhance theoretical rigor have sometimes increased this complexity, exacerbating these biases.

Danz et al. (2022) demonstrated this point by empirically testing the Binarized Scoring Rule (BSR), a popular incentivized mechanism for belief elicitation. Their goal was to determine whether providing detailed information about incentives improves reporting accuracy when participants have a known objective prior. This aspect of knowing the objective prior was central to their experiment. They found that informing participants about the incentives increased error rates to over 40%, with only 15% of participants consistently reporting truthfully. They concluded that misreporting stems from a biased perception of incentives. Efforts to debias participants had limited success, and the highest levels of truth-telling occurred when quantitative incentive information was withheld. They observed that gradually introducing information further reduces truth-telling, which further supports the idea that incentives can contribute to false reports. See also Danz et al. (2024) with regard to complexity of incentives.

As when using multiple price lists to measure risk taking, the evidence here points to a measurement cost when complex incentive mechanisms are introduced. But it is impossible to conclude whether this added complexity increases accuracy, because—unlike in Danz et al. (2022)—in most of these cases we do not know what the true beliefs of participants are. Another aspect of complexity caused by some incentivized belief measures is that they may create incentives to distort them by using the payments to hedge payoffs by guessing outcomes in which you get relatively low earnings from the outcome itself. Encouragingly, some simple, user-friendly incentivized methods may outperform introspection, though few studies have directly compared these approaches.

5.5. *Biases*

Biases refer to systematic deviations from the predictions of the neoclassical economic model that result in "errors." The discussion of this type of biases began with Allais (1953) and Ellsberg (1961), who introduced paradoxes challenging subjective expected utility. Tversky and Kahneman (1974) expanded the conversation by identifying heuristics—mental shortcuts the brain uses to simplify information processing—that, when applied inappropriately, lead to systematic deviations from rationality. This line of research has had a lasting impact because failing to account for biases can result in models that poorly predict actual human behavior.

Do incentives help people "get it right"? Tversky and Kahneman used hypothetical scenarios in their studies. Experimental economists have criticized this approach, advocating for incentive-compatible designs where participants' choices directly affected their payoffs. Does this approach eliminate biases? An important example is Grether and Plott (1979), who replicated the preference-reversal phenomenon first observed by Lichtenstein and Slovic (1971). Contrary to their expectations, real incentives did not reduce the bias (see also the discussion in Cox, 2008). This surprising result contributed to Camerer and Hogarth's (1999) conclusion that no replicated study had eliminated rationality violations solely by increasing incentives. A counterexample is Charness et al. (2010), that presents results from a series of experiments testing the conjunction fallacy. Adopting the experimental framework of Tversky and Kahneman (1983), they show that when participants are given modest incentives, the share of individuals violating the conjunction principle is substantially lower than in the original findings.

Despite the significance of this topic, systematic evidence on how incentives influence biases remains limited. Enke et al. (2023) addressed this gap by examining cognitive biases in a Nairobi laboratory. Their study tested three pay levels: hypothetical scenarios (with low and high hypothetical stakes), small incentives (standard lab pay for a few hours), and large incentives (exceeding a month's pay). They found that while cognitive effort, as measured by response time, increased by 40% under very high stakes, performance improved only marginally or not at all.

While Enke et al. (2023) studied mistakes, Gneezy et al. (2025) explored this interaction between incentives and biases in the context of the Allais paradox. They observed that deviations from expected utility *increased* as stakes rose. Their findings also indicated that when approximating behavior in high-stakes Allais experiments—which are often prohibitively expensive—hypothetically high stakes are more accurate than the commonly used approach of real but low stakes.

In summary, these findings suggest that even very high incentives fail to eliminate biases, challenging the assumption that increased stakes invariably lead to more rational decisions.

5.6. *Hypothetical bias*

A special group of biases are not violations of the standard model, but rather differences between stated preferences and actual behavior—known as hypothetical biases. Much like experiments, surveys are frequently employed by researchers to elicit preferences. Unlike incentivized experiments, however, surveys capture stated rather than revealed choices. This distinction has long fueled debate about the extent to which survey data reflect “true preferences.”

A prominent example is the literature on contingent valuation (CV), a widely used survey method designed to measure willingness to pay for nonmarket goods such as environmental quality (e.g., Mitchell and Carson, 1989). To improve the credibility of stated preference surveys, standards have been developed to mitigate distorted valuations, such as the guidelines issued by a panel chaired by Kenneth Arrow and Robert Solow in response to the Exxon Valdez oil spill (Arrow et al., 1993), which sought to establish best practices for estimating consumers’ valuations of environmental damages.

The CV method typically relies on hypothetical scenarios in which respondents indicate whether they would pay a given amount for a specified good. Despite its influence, the approach is controversial, with critics questioning whether hypothetical responses map onto real economic behavior (Diamond and Hausman, 1994; Hanemann, 1994). To address this concern, laboratory and field experiments have directly compared CV responses with actual purchase or donation decisions. For instance, in some studies participants are offered the opportunity to buy a consumer good or contribute to a public good immediately after responding to a CV-style hypothetical question, allowing researchers to compare hypothetical “yes” responses to real choices at the same price point.

Experimental evidence consistently suggests that CV responses overestimate real WTP (the so-called “hypothetical bias”). Studies of private goods (e.g., Bishop and Heberlein, 1979; Cummings, Harrison and Rutström, 1995) and public goods (e.g., Champ et al., 1997) document higher stated than actual WTP. Blumenschein et al. (1998) similarly compared hypothetical responses with real purchase decisions for a consumer good, confirming that hypothetical “yes” rates systematically exceeded actual purchases. Importantly, however, they could not reject the null hypothesis that "definitely sure" yes responses correspond to real purchasing behavior.

On the other side of this debate are researchers who argue that hypothetical stakes suffice in decision-making experiments. For example, Kühberger (2001) writes "Decision making can be studied using hypothetical payoffs because it is hypothetical to its very core. However, the core process can be influenced by contextual features. As there is no theory for these contextual features, a "do-it-both-ways" rule amounts to a waste of money. If we had such a theory, doing it both ways would be unnecessary." See also Read (2005) and Gilis and Hettler (2007).

Murphy, Allen, Stevens and Weatherhead (2005) argue that the beliefs regarding how much individuals overstate their economic valuation is around a factor of two or three. They then report the results of a meta-analysis of hypothetical bias across 28 stated-preference valuation studies that report monetary WTP using the same elicitation mechanism for both hypothetical and actual values. The analysis yields 83 observations, with a median ratio of hypothetical to actual WTP of 1.35, concluding that choice-based elicitation mechanisms are important in reducing bias, but not as much as some argue.

The discussion highlights the difference between quantitative and qualitative differences. Quantitative effects are the extent to which the magnitude of behavior changes with real versus hypothetical payoffs. Qualitative effects are whether incentives alter the direction of behavior. Theoretical and empirical arguments suggest that monetary incentives are more likely to influence the size of an effect than its sign. If so, hypothetical incentives may suffice when the research question concerns directional validity (e.g., whether fairness concerns increase giving), but real and sufficiently high incentives are required when precise quantitative estimates are needed (e.g., how much individuals donate to charity). This distinction shows the importance of considering both qualitative and quantitative validity in evaluating the external validity of laboratory findings (Levitt and List, 2007; Kessler and Vesterlund, 2015; Masclet and Rebieres, 2023).

As discussed above, most laboratory experiments rely on relatively small stakes. As such, their results may lack quantitative realism even if their qualitative insights are robust, which raises the broader question of whether monetary incentives are always necessary, especially given their costs (Camerer and Hogarth, 1999; Hertwig and Ortmann, 2001; Read, 2005). Alternative approaches—such as clearer instructions or additional time—may achieve similar improvements at lower cost. Moreover, hypothetical incentives can enable larger and more diverse samples which may enhance external validity. Balancing the benefits of monetary incentives against their limitations and costs thus remains a central challenge for experimental economics.

5.7. *Way forward*

The evidence presented here illustrates the varied effects of incentives across different contexts. Incentives often increase effort in multiple dimensions, such as simplifying participant recruitment and ensuring attention when the stakes are sufficiently high. However, the rigid stance that only incentivized experiments should be published is, in my view, unjustified. For one, this approach disregards a significant portion of psychological research, where unincentivized experiments have yielded valuable insights. Additionally, incentives can create barriers, particularly for early-career researchers with limited resources. Incentives can also backfire in certain situations by reducing effort or causing choking. These findings raise important questions about the universal effectiveness of incentives on effort and performance in experiments. Importantly, incentivizing participants can increase the complexity of an experiment and risk confusing them.

While incentivizing participants should remain the standard practice in experimental economics, flexibility is warranted. We should recognize the value of experiments that incorporate unincentivized components when appropriate. For instance, a study might incentivize the primary behavior under investigation but employ unincentivized methods to measure ancillary factors such as risk preferences or beliefs. Embracing such pragmatic approaches can broaden the scope and applicability of experimental research without compromising its rigor.

6. Concluding remarks

The emergence of experimental economics over half a century ago fundamentally transformed the field of economics, steering it toward a more behavioral and empirically grounded approach. By challenging traditional simplifying assumptions and demonstrating systematic deviations through rigorous lab studies, experimental economics has reshaped how economists understand human behavior. Consider the assumption of selfishness. Thanks to experimental economics, we know more about the importance of fairness considerations and how they affect behavior in labor markets. The findings from experimental economics established the field as a cornerstone of modern economic thinking.

Experimental economics is no longer an emerging discipline. It has matured and evolved alongside technological advancements. In addition, the world has changed—most notably, a large portion of experiments have moved to online labs. There are many differences between lab and online experiments: Participants in online experiments may be less attentively (Anderhub et al. 2001), it is also harder to create common knowledge among participants online or to verify understanding. Are the norms online different than those in the lab? Some arguments, most notably the repeated game and negative externalities seem to be less convincing online. We have very little control over what the participants are exposed to in other experiments online. Yet the methodological principles that once helped establish the field's credibility continue to serve as benchmarks for experimental rigor. These principles should be re-evaluated in light of new contexts and tools. Most relevant is the development of automated agents. Data quality could be compromised by these automated agents, with participants who rely on LLMs without researchers' awareness. The current situation is far from a steady state, with evolving developments and counter measures. For example, Celebi et al. (2025) identify behavioral patterns that serve as effective data-quality indicators, and suggest ways to use them to improve data collection online.

I have examined four of what I think are the most important principles of experimental economics: No-deception, abstract instructions, induced value, and incentives. As in the case of the ten commandments, some violations are more tolerable than others ("You shall not bear false witness" is important, but most people will agree that "You shall not murder" is more important). Applied to experimental economics, violations of the use of abstract instructions or induced values are accepted with some reservations. However, as mentioned above, no-deception and incentives are still necessary conditions for publishing experimental papers in economics.

Why are the norms in economics and psychology different, and what can each field learn from the other? For example, experimental economics has benefited greatly from the use of induced values and incentives—not only for aligning participants' motivations, but also because these methods require researchers to specify games with precision: who the players are, what are the strategies, and what payoffs follow. The difference from psychology is likely to be related to the economists' stronger training and reliance on formal models. This difference in approach has enhanced clarity and interpretability of results and is one of the field's major advantages. At the same time, strict adherence to incentivizing every action can constrain research. It may overcomplicate designs and limit the exploration of phenomena that are difficult to capture with incentive compatible designs. Taken together, the fields have much to learn from one another: economists can appreciate the richness that comes from looser, more exploratory designs, while psychologists can benefit from the rigor and clarity induced by structured, incentive-based methods.

This paper advocates for a pragmatic and open-minded approach to experimental methods. Referees and editors, as gatekeepers of the discipline, should critically assess deviations from traditional rules on a case-by-case basis. If such deviations are well-reasoned and enhance the validity or relevance of the findings, they should be embraced rather than dismissed. Experimental economics must remain adaptive to maintain its relevance and continue driving innovation within the broader economics profession.

Declaration of competing interest

I declare no conflict of interest associated with the submission "Experimental Methods: Revisiting the Commandments of Experimental Economics."

Data availability

No data was used for the research described in the article.

References

Aarts, E., Wallace, D.L., Dang, L.C., Jagust, W.J., Cools, R., D'Esposito, M., 2014. Dopamine and the cognitive downside of a promised bonus. Psychol. Sci. 25 (4), 1003–1009. https://doi.org/10.1177/0956797613517240.

Abbink, K., Hennig-Schmidt, H., 2006. Neutral versus loaded instructions in a bribery experiment. Exp. Econ. 9 (2), 103–121. https://doi.org/10.1007/s10683-006-5385-z.

Aksoy, B., de Oliveira, A.C.M., Eckel, C., 2025. Workers' response to monetary incentives in for-profit and non-profit jobs. Econ. Inq. 63 (3), 830–845.

Alekseev, A., Charness, G., Gneezy, U., 2017. Experimental methods: when and why contextual instructions are important. J. Econ. Behav. Organ. 134, 48–59. https://doi.org/10.1016/j.jebo.2018.02.024.

Allais, M., 1953. Le Comportement de l'Homme Rationnel devant le Risque: critique des Postulats et Axiomes de l'Ecole Americaine. Econometrica 21 (4), 503–546. https://doi.org/10.2307/1907921. JSTOR.

American Psychological Association, 2017. Ethical Principles of Psychologists and Code of Conduct. American Psychological Association. https://www.apa.org/ethics/code.

Amir, O., Rand, D.G., Gal, Y.K., 2012. Economic games on the internet: the effect of $1 stakes. PLoS. One 7 (2), e31461. https://doi.org/10.1371/journal.pone.0031461.

Anderhub, V., Müller, R., Schmidt, C., 2001. Design and evaluation of an economic experiment via the internet. J. Econ. Behav. Org. 46 (2), 227–247.

Andersen, S., Ertaç, S., Gneezy, U., Hoffman, M., List, J.A., 2011. Stakes matter in ultimatum games. Am. Econ. Rev. 101 (7), 3427–3439. https://doi.org/10.1257/aer.101.7.3427.

Ariely, D., Gneezy, U., Loewenstein, G., Mazar, N., 2009. Large stakes and big mistakes. Rev. Econ. Stud. 76 (2), 451–469. https://doi.org/10.1111/j.1467-937x.2009.00534.x.

Asulin, Y., Heller, Y., Munichor, N., 2024. Comparing the effects of non-monetary incentives and monetary incentives on prosocial behavior. Eur. Econ. Rev. 165, 104740.

Arrow, K., Solow, R., Portney, P., Leamer, E., Radner, R. & Schuman, H. (1993). Report of the NOAA panel on contingent valuation.

Ayres, I., Siegelman, P., 1995. Race and gender discrimination in bargaining for a new car. Am. Econ. Rev. 85 (3), 304–321. https://www.jstor.org/stable/2118176.

Barr, A., Serra, D., 2009. The effects of externalities and framing on bribery in a petty corruption experiment. Exp. Econ. 12 (4), 488–503. https://doi.org/10.1007/s10683-009-9225-9.

Banerjee, R., 2014. On the interpretation of bribery in a laboratory corruption game: moral frames and social norms. Exp. Econ. 2, 1–28.

Baumeister, R.F., 1984. Choking under pressure. J. Pers. Soc. Psychol. 46 (3), 610–620.

Beilock, S.L., Carr, T.H., 2001. On the fragility of skilled performance: what governs choking under pressure? J. Exp. Psychol.: Gen. 130 (4), 701–725. https://doi.org/10.1037/0096-3445.130.4.701.

Ben-Ner, A., Kramer, A., Levy, O., 2008. Economic and hypothetical dictator game experiments: incentive effects at the individual level. J. Socio Econ. 37 (5), 1775–1784. https://doi.org/10.1016/j.socec.2007.11.004.

Berns, G., Capra, C., Moore, S., Noussair, C., 2007. A shocking experiment: new evidence on probability weighting and common ratio violations. Judgm. Decis. Mak. 234–242.

Binswanger, H.P., 1981. Attitudes toward risk: theoretical implications of an experiment in rural India. Econ. J. 91 (364), 867. https://doi.org/10.2307/2232497.

Bishop, R.C., Heberlein, T.A., 1979. Measuring values of extramarket goods: are indirect measures biased? Am. J. Agric. Econ. 61, 926–930.

Blumenschein, K., Johannesson, M., Blomquist, G.C., Liljas, B., O'Conor, R.M., 1998. Experimental results on expressed certainty and hypothetical bias in contingent valuation. South. Econ. J. 65 (1), 169–177.

Bonner, S.E., Hastie, R., Sprinkle, G.B., Young, S.M., 2000. A review of the effects of financial incentives on performance in laboratory tasks: implications for management accounting. J. Manag. Account. Res. 12 (1), 19–64. https://doi.org/10.2308/jmar.2000.12.1.19.

Brodeur, A., Mikola, D., & Cook, N. (2025). Mass reproducibility and replicability: a new hope. IZA discussion Paper No. 16912.

Burfurd, I., Wilkening, T., 2021. Cognitive heterogeneity and complex belief elicitation. Exp. Econ. 25 (2). https://doi.org/10.1007/s10683-021-09722-x.

Burnham, T., McCabe, K., Smith, V.L., 2000. Friend-or-foe intentionality priming in an extensive form trust game. J. Econ. Behav. Organ. 43 (1), 57–73. https://doi.org/10.1016/s0167-2681(00)00108-6.

Camerer, C.F., Hogarth, R.M., 1999. The effects of financial incentives in experiments: a review and capital-labor-production framework. J. Risk. Uncertain. 19 (1/3), 7–42. https://doi.org/10.1023/a:1007850605129.

Camerer, C.F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., Wu, H., 2016. Evaluating replicability of laboratory experiments in economics. Science (1979) 351 (6280), 1433–1436. https://doi.org/10.1126/science.aaf0918.

Cameron, L.A., 1999. Raising the stakes in the ultimatum game: experimental evidence from Indonesia. Econ. Inq. 37 (1), 47–59.

Carpenter, J., Verhoogen, E., Burks, S., 2005. The effect of stakes in distribution experiments. Econ. Lett. 86 (3), 393–398. https://doi.org/10.1016/j.econlet.2004.08.007.

Cason, T.N., Wu, S.Y., 2019. Subject pools and deception in agricultural and resource economics experiments. Environ. Resour. Econ. 73, 743–758. https://doi.org/10.1007/s10640-018-0289-x.

Cason, T.N., Raymond, L., 2011. Framing effects in an emissions trading experiment with voluntary compliance. Res. Exp. Econ. 14, 77–114. https://doi.org/10.1108/s0193-2306(2011)0000014006.

Cassar, A., Friedman, D., 2004. Economics Lab. Routledge.

Castro, J.F., Yamada, G., Contreras, H., Linares, F., Watson, H., 2021. Tracking the pressure: the effect of monetary incentives on effort, emotions, and performance. J. Neurosci. Psychol. Econ. 14 (4), 234–258. https://doi.org/10.1037/npe0000148.

Celebi, C., Exley, C., Harrs, S., Kivimaki, H., Serra-Garcia, M. & Yuso, J. (2025). Mission possible: data quality in online surveys. https://www.dropbox.com/scl/fi/t7kl81ml9fzzdbfs6w6if/DataQuality_Oct2025.pdf?rlkey=hpat3sn4wdckk9l34eccsxqld&e=2&dl=0.

Champ, P.A., Bishop, R.C., Brown, T.C., McCollum, D.W., 1997. Using donation mechanisms to value nonuse benefits from public goods. J. Environ. Econ. Manage 33, 151–162.

Charness, G., Karni, E., Levin, D., 2010. On the conjunction fallacy in probability judgment: new experimental evidence regarding Linda. Games Econ. Behav. 68 (2), 551–556.

Charness, G., Fehr, E., 2025. Social preferences: fundamental characteristics and economic consequences. J. Econ. Lit. 63 (2), 440–514.

Charness, G., Gneezy, U., 2012. Strong evidence for gender differences in risk taking. J. Econ. Behav. Organ. 83 (1), 50–58. https://doi.org/10.1016/j.jebo.2011.06.007.

Charness, G., Gneezy, U., Imas, A., 2013. Experimental methods: eliciting risk preferences. J. Econ. Behav. Org. 87, 43–51.

Charness, G., Eckel, C., Gneezy, U., Kajackaite, A., 2018a. Complexity in risk elicitation may affect the conclusions: a demonstration using gender differences. J. Risk. Uncertain. 1–17.

Charness, G., Gneezy, U., Henderson, A., 2018b. Experimental methods: measuring effort in economics experiments. J. Econ. Behav. Organ. 149, 74–87. https://doi.org/10.1016/j.jebo.2018.02.024.

Charness, G., Samek, A., van de Ven, J., 2021. What is considered deception in experimental economics? Exp. Econ. 25 (2), 385–412. https://doi.org/10.1007/s10683-021-09726-7.

Chen, R., Chen, Y., Riyanto, Y.E., 2021. Best practices in replication: a case study of common information in coordination games. Exp. Econ. 24, 2–30.

Cook, K.S., Yamagishi, T., 2008. A defense of deception on scientific grounds. Soc. Psychol. Q. 71 (3), 215–221. https://doi.org/10.1177/019027250807100303.

Cooper, D.J., Kagel, J.H., 2003. The impact of meaningful context on strategic play in signaling games. J. Econ. Behav. Organ. 50 (3), 311–337. https://doi.org/10.1016/s0167-2681(02)00025-2.

Cooper, D.J., Kagel, J.H., 2009. The role of context and team play in cross-game learning. J. Eur. Econ. Assoc. 7 (5), 1101–1139. https://doi.org/10.1162/jeea.2009.7.5.1101.

Cooper, D.J., 2014. A note on deception in economic experiments. J. Wine Econ. 9 (2), 111–114.

Cosmides, L., 1989. The logic of social exchange: has natural selection shaped how humans reason? Studies with the Wason selection task. Cognition 31 (3), 187–276. https://doi.org/10.1016/0010-0277(89)90023-1.

Cox, J.C. (2008). Preference reversal. Chapter 104 in the handbook of experimental economics results, Volume 1, 2008, Pages 967–975.

Cox, J.C., Sadiraj, V., Schmidt, U., 2015. Paradoxes and mechanisms for choice under risk. Exp. Econ. 18, 215–250.

Cox, J.C., Sadiraj, V., 2019. Incentives. Handbook of Research Methods and Applications in Experimental Economics. Edward Elgar Publishing, pp. 9–27. https://doi.org/10.4337/9781788110563.00008.

Croson, R., 2005. The method of experimental economics. Int. Negotiation 10 (1), 131–148. https://doi.org/10.1163/1571806054741100.

Croson, R., Gneezy, U., 2009. Gender differences in preferences. J. Econ. Lit. 47 (2), 448–474. https://doi.org/10.1257/jel.47.2.448.

Cummings, R.G., Harrison, G.W., Rutstrom, E.E., 1995. Homegrown values and hypothetical surveys: is the dichotomous choice approach incentive-compatible? Am. Econ. Rev. 85, 260–266.

Dana, J., Weber, R.A., Kuang, J.X., 2006. Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. Econ. Theory. 33 (1), 67–80. https://doi.org/10.1007/s00199-006-0153-z.

Daniel, W.W. (1968). Racial discrimination in England: based on the PEP report.

Danz, D., Vesterlund, L., Wilson, A.J., 2022. Belief elicitation and behavioral incentive compatibility. Am. Econ. Rev. 112 (9), 2851–2883. https://doi.org/10.1257/aer.20201248.

Danz, D., Vesterlund, L., Wilson, A.J., 2024. Evaluating behavioral incentive compatibility: insights from experiments. J. Econ. Perspect. 38 (4), 131–154.

Davis, D.D., Holt, C.A., 1993. Experimental Economics. Princeton University Press.

Deci, E.L., 1971. Effects of externally mediated rewards on intrinsic motivation. J. Pers. Soc. Psychol. 18 (1), 105–115.

Diamond, P.A., Hausman, J.A., 1994. Contingent valuation: is some number better than No number? J. Econ. Perspect. 8 (4), 45–64.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G.G., 2011. Individual risk attitudes: measurement, determinants, and behavioral consequences. J. Eur. Econ. Assoc. 9 (3), 522–550. https://doi.org/10.1111/j.1542-4774.2011.01015.x.

Dufwenberg, M., Gächter, S., Hennig-Schmidt, H., 2011. The framing of games and the psychology of play. Games. Econ. Behav. 73 (2), 459–478. https://doi.org/10.1016/j.geb.2011.02.003.

Dutcher, E.G., Salmon, T.C., Saral, K.J., 2024. Is "real" effort more real? Exp. Econ. 27, 1001–1032.

Eckel, C.C., Grossman, P.J., 1996. Altruism in anonymous dictator games. Games. Econ. Behav. 16 (2), 181–191. https://doi.org/10.1006/game.1996.0081.

Eckel, C.C., Wilson, R.K., 2007. Social learning in coordination games: does status matter? Exp. Econ. 10, 317–329.

Eckel, C.C., Wilson, R.K., Yang, N., 2025. The evolution of risk attitudes: a panel study of the university years. J. Risk. Uncertain. 70, 225–248.

Ellingsen, T., Johannesson, M., Mollerstrom, J., Munkhammar, S., 2013. Gender differences in social framing effects. Econ. Lett. 118 (3), 470–472. https://doi.org/10.1016/j.econlet.2012.12.010.

Ellingsen, T., Johannesson, M., Mollerstrom, J., Munkhammar, S., 2012. Social framing effects: preferences or beliefs? Games. Econ. Behav. 76 (1), 117–130. https://doi.org/10.1016/j.geb.2012.05.007.

Ellsberg, D., 1961. Risk, ambiguity, and the savage axioms. Q. J. Econ. 75 (4), 643–669. https://doi.org/10.2307/1884324.

Engel, C., 2011. Dictator games: a meta study. Exp. Econ. 14 (4), 583–610. https://doi.org/10.1007/s10683-011-9283-7.

Enke, B., Gneezy, U., Hall, B., Martin, D., Nelidov, V., Offerman, T., van de Ven, J., 2023. Cognitive biases: mistakes or missing stakes? Rev. Econ. Stat. 105 (4), 1–818832. https://doi.org/10.1162/rest_a_01093.

Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. Q. J. Econ. 114 (3), 817–868.

Fiorina, M.P., Plott, C.R., 1978. Committee decisions under majority rule: an experimental study. Am. Polit. Sci. Rev. 72 (2), 575–598. https://doi.org/10.2307/1954111.

Forsythe, R., Horowitz, J.L., Savin, N.E., Sefton, M., 1994. Fairness in simple bargaining experiments. Games. Econ. Behav. 6 (3), 347–369. https://doi.org/10.1006/game.1994.1021.
Frey, B.S., Jegen, R., 2001. Motivation crowding Theory. J. Econ. Surv. 15 (5), 589–611. https://doi.org/10.1111/1467-6419.00150.
Frey, B.S., Oberholzer-Gee, F., 1997. The cost of price incentives: an empirical analysis of motivation crowding-out. Am. Econ. Rev. 87 (4), 746–755.
Friedman, D., 1991. A simple testable model of double auction markets. J. Econ. Behav. Organ. 15 (1), 47–70. https://doi.org/10.1016/0167-2681(91)90004-h.
Friedman, D., Isaac, M., James, D., 2014. Risky Curves: On the Empirical Failure of Expected Utility. Routledge.
Frohlich, N., Oppenheimer, J., Moore, B., 2001. Some doubts about measuring self-interest using dictator experiments: the costs of anonymity. J. Econ. Behav. Organ. 46 (3), 271–290.
Gillis, M.T., Hettler, P.L., 2007. Hypothetical and real incentives in the ultimatum game and Andreoni's public goods game: an experimental study. East. Econ. J. 33 (4), 491–510.
Gneezy, U., 2005. Deception: the role of consequences. Am. Econ. Rev. 95 (1), 384–394. https://doi.org/10.1257/0002828053828662.
Gneezy, U., 2023. Mixed Signals. Yale University Press.
Gneezy, U., Halevy, Y., Hall, B., Offerman, T. & van de Ven, J. (2025). How real is hypothetical? A high-stakes test of the Allais paradox. Mimeo.
Gneezy, U., List, J.A., 2006. Putting behavioral economics to work: testing for gift exchange in labor markets using field experiments. Econometrica 74 (5), 1365–1384. https://doi.org/10.1111/j.1468-0262.2006.00707.x.
Gneezy, U., Rustichini, A., 2000. Pay enough or don't pay at all. Q. J. Econ. 115 (3), 791–810. https://doi.org/10.1162/003355300554917.
Grether, D.M., Plott, C.R., 1979. Economic theory of choice and the preference reversal phenomenon. Am. Econ. Rev. 69 (4), 623–638. https://doi.org/10.2307/1808708. JSTOR.
Griggs, R.A., Cox, J.R., 1982. The elusive thematic-materials effect in Wason's selection task. Br. J. Psychol. 73 (3), 407–420. https://doi.org/10.1111/j.2044-8295.1982.tb01823.x.
Güth, W., Schmittberger, R., Schwarze, B., 1982. An experimental analysis of ultimatum bargaining. J. Econ. Behav. Organ. 3 (4), 367–388. https://doi.org/10.1016/0167-2681(82)90011-7.
Guzzo, R.A., Jette, R.D., Katzell, R.A., 1985. The effects of psychologically based intervention programs on worker productivity: a meta-analysis. Pers. Psychol. 38 (2), 275–291. https://doi.org/10.1111/j.1744-6570.1985.tb00547.x.
Hanemann, V.M., 1994. Valuing the environnent through contingent valuation. J. Econ. Perspect. 8 (4), 19–43.
Harrison, G.W., Rutström, E.E., 2001. Doing it both ways - experimental practice and eeuristic context. Behav. Brain Sci. 24 (3), 413–414.
Harrison, G.W., Johnson, E., Mcinnes, M.M., Rutström, E.E., 2005. Risk aversion and incentive effects: comment. Am. Econ. Rev. 95 (3), 897–901. https://doi.org/10.1257/0002828054201378.
Heinz, M., Juranek, S., Rau, H.A., 2012. Do women behave more reciprocally than men? Gender differences in real effort dictator games. J. Econ. Behav. Organ. 83 (1), 105–110. https://doi.org/10.1016/j.jebo.2011.06.015.
Hersch, G., 2015. Experimental economics' inconsistent ban on deception. Stud. Hist. Philos. Sci. Part A 52, 13–19. https://doi.org/10.1016/j.shpsa.2015.04.005.
Hertwig, R., Ortmann, A., 2001. Experimental practices in economics: a methodological challenge for psychologists? Behav. Brain Sci. 24 (3), 383–403. https://doi.org/10.1017/s0140525×01004149.
Hertwig, R., Ortmann, A., 2008. Deception in experiments: revisiting the arguments in its defense. Ethics Behav. 18 (1), 59–92.
Hey, J.D., 1998. Experimental economics and deception: a comment. J. Econ. Psychol. 19 (3), 397–401.
Hickman, D.C., Metz, N.E., 2015. The impact of pressure on performance: evidence from the PGA TOUR. J. Econ. Behav. Organ. 116, 319–330. https://doi.org/10.1016/j.jebo.2015.04.007.
Hilbig, B.E., Thielmann, I., 2021. On the (Mis)use of deception in web-based research. Zeitsch. Psychol. 229 (4), 225–229. https://doi.org/10.1027/2151-2604/a000466.
Hoffman, E., McCabe, K., Shachat, K., Smith, V.L., 1994. Preferences property rights and anonymity in bargaining games. Games. Econ. Behav. 7 (3), 346–380.
Hoffman, E., McCabe, K., Smith, V.L., 1996. Social distance and other-regarding behavior in dictator games. Am. Econ. Rev. 86 (3), 653–660. https://www.jstor.org/stable/2118218.
Holt, C.A., Laury, S.K., 2002. Risk aversion and incentive effects. Am. Econ. Rev. 92 (5), 1644–1655. https://doi.org/10.1257/000282802762024700.
Isaac, R.M., Plott, C.R., 1981. Price controls and the behavior of auction markets: an experimental examination. Am. Econ. Rev. 71 (3), 448–459. https://doi.org/10.2307/1802791. JSTOR.
Jamison, J., Karlan, D., Schechter, L., 2008. To deceive or not to deceive: the effect of deception on behavior in future laboratory experiments. J. Econ. Behav. Organ. 68 (3–4), 477–488. https://doi.org/10.1016/j.jebo.2008.09.002.
Jenkins Jr., G.D., Mitra, A., Gupta, N., Shaw, J.D, 1998. Are financial incentives related to performance? A meta-analytic review of empirical research. J. Appl. Psychol. 83 (5), 777–787. https://doi.org/10.1037/0021-9010.83.5.777.
Karagözoğlu, E., Urhan, Ü.B., 2017. The effect of stake size in experimental bargaining and distribution games: a survey. Group. Decis. Negot. 26 (2), 285–325. https://doi.org/10.1007/s10726-016-9490-x.
Kessler, J., Vesterlund, L., 2015. In: Frechette, G., Schotter, A. (Eds.), The External Validity of Laboratory Experiments: The Misleading Emphasis on Quantitative Effects. Handbook of Experimental Economic Methodology. Oxford University Press.
Keuschnigg, M., Bader, F., Bracher, J., 2016. Using crowdsourced online experiments to study context-dependency of behavior. Soc. Sci. Res. 59, 68–82. https://doi.org/10.1016/j.ssresearch.2016.04.014.
Kimmel, A.J., 2001. Deception in marketing research and practice: an introduction. Psychol. Mark. 18 (7), 657.
Kirchler, M., Huber, J., Stöckl, T., 2012. Thar she bursts: reducing confusion reduces bubbles. Am. Econ. Rev. 102 (2), 865–883. https://doi.org/10.1257/aer.102.2.865.
Kreps, D.M., 1997. Intrinsic motivation and extrinsic incentives. Am. Econ. Rev. 87 (2), 359–364.
Kroll, Y., Levy, H., Rapoport, A., 1988. Experimental tests of the separation theorem and the capital asset pricing model. Am. Econ. Rev. 78 (3), 500–519. https://doi.org/10.2307/1809148.
Kruglanski, A.W., Alon, S., Lewis, T., 1972. Retrospective misattribution and task enjoyment. J. Exp. Soc. Psychol. 8 (6), 493–501. https://doi.org/10.1016/0022-1031(72)90075-3.
Kühberger, A., 2001. Why use real and hypothetical payoffs? Behav. Brain Sci. 24 (3), 419–420.
Larney, A., Rotella, A., Barclay, P., 2019. Stake size effects in ultimatum game and dictator game offers: a meta-analysis. Organ. Behav. Hum. Decis. Process. 151, 61–72. https://doi.org/10.1016/j.obhdp.2019.01.002.
Lee, T.G., Acuña, D.E., Kording, K.P., Grafton, S.T., 2018. Limiting motor skill knowledge via incidental training protects against choking under pressure. Psychon. Bull. Rev. 26 (1), 279–290. https://doi.org/10.3758/s13423-018-1486-x.
Leibbrandt, A., Maitra, P., Neelim, A., 2015. On the redistribution of wealth in a developing country: experimental evidence on stake and framing effects. J. Econ. Behav. Organ. 118, 360–371. https://doi.org/10.1016/j.jebo.2015.02.015.
Lepper, M.R., Greene, D., Nisbett, R.E., 1973. Undermining children's intrinsic interest with extrinsic reward: a test of the "overjustification" hypothesis. J. Pers. Soc. Psychol. 28 (1), 129–137. https://doi.org/10.1037/h0035519.
Levitt, S., List, J.A., 2007. What do laboratory experiments measuring social preferences reveal about the real world? J. Econ. Perspect. 21 (2), 153–174.
Liberman, V., Samuels, S.M., Ross, L., 2004. The name of the game: predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. Personality Soc. Psychol. Bull. 30 (9), 1175–1185. https://doi.org/10.1177/0146167204264004.
Lichtenstein, S., Slovic, P., 1971. Reversals of preference between bids and choices in gambling decisions. J. Exp. Psychol. 89 (1), 46–55. https://doi.org/10.1037/h0031207.
List, J.A., Cherry, T.L., 2008. Examining the role of fairness in high stakes allocation decisions. J. Econ. Behav. Organ. 65 (1), 1–8. https://doi.org/10.1016/j.jebo.2003.09.021.

Loewenstein, G., 1999. Experimental economics from the vantage-point of behavioural economics. Econ. J. 109 (453), 25–34. https://doi.org/10.1111/1468-0297.00400.
Loomes, G., 2005. Modelling the stochastic component of behaviour in experiments: some issues for the interpretation of data. Exp. Econ. 8 (4), 301–323. https://doi.org/10.1007/s10683-005-5372-9.
Masclet, D., Rebière, T., 2023. Comparing real and hypothetical incentives in giving and money burning experiments. Ann. Econ. Stat. 152, 65–102.
Massoni, S., Gajdos, T., Vergnaud, J.-C., 2014. Confidence measurement in the light of signal detection theory. Front. Psychol. 5. https://doi.org/10.3389/fpsyg.2014.01455.
Masters, R.S.W., 1992. Knowledge, knerves, and know-how: the role of explicit versus implicit knowledge in the breakdown of a complex motor skill under pressure. Br. J. Psychol. 83 (3), 343–358. https://doi.org/10.1111/j.2044-8295.1992.tb02446.x.
Mitchell, R.C., Carson, R.T., 1989. Using Surveys to Value Public Goods the Contingent Valuation Method. Resources for the Future, Washington DC.
Mobbs, D., Hassabis, D., Seymour, B., Marchant, J.L., Weiskopf, N., Dolan, R.J., Frith, C.D., 2009. Choking on the money. Psychol. Sci. 20 (8), 955–962. https://doi.org/10.1111/j.1467-9280.2009.02399.x.
Moscati, I., 2016. Measuring the economizing mind in the 1940s and 1950s: the Mosteller-Nogee and Davidson-Suppes-Siegel experiments to measure the utility of money. Hist. Polit. Econ. 48 (suppl 1), 239–269. https://doi.org/10.1215/00182702-3619298.
Murphy, J.J., Allen, P.J., Stevens, T.H., Weatherhead, D., 2005. A meta-analysis of hypothetical bias in stated preference valuation. Environ. Resour. Econ. 30, 313–325.
Olson, J.A., Raz, A., 2021. Applying insights from magic to improve deception in research: the swiss cheese model. J. Exp. Soc. Psychol. 92, 104053. https://doi.org/10.1016/j.jesp.2020.104053.
Open Science Collaboration, 2015. Estimating the reproducibility of psychological science. Science (1979) 349 (6251), aac4716.
Ortmann, A., 2019. Deception. Edward Elgar Publishing EBooks. https://doi.org/10.4337/9781788110563.00009.
Ortmann, A., Hertwig, R., 2002. The costs of deception: evidence from psychology. Exp. Econ. 5 (2), 111–131. https://doi.org/10.1023/a:1020365204768.
Pevnitskaya, S., Ryvkin, D., 2013. Environmental context and termination uncertainty in games with a dynamic public bad. Environ. Dev. Econ. 18 (1), 27–49. https://doi.org/10.1017/s1355770×12000423.
Plott, C.R., Smith, V.L., 1978. An experimental examination of two exchange institutions. Rev. Econ. Stud. 45 (1), 133–153. https://doi.org/10.2307/2297090.
Plott, C., 1982. Industrial Organization theory and Experimental Economics. J. Econ. Lit. 20 (4), 1485–1527.
Przemysław, G.H., 2021. Reproducibility and replicability crisis: how management compares to psychology and economics – a systematic review of literature. Euro. Manag. J. 39, 577–594.
Read, D., 2005. Monetary incentives, what are they good for? J. Econ. Methodol. 12, 265–276.
Roth, A.E., Kagel, J.H., 2016. The Handbook of Experimental Economics, Volume 2. Princeton University Press.
Rydval, O., Ortmann, A., 2004. How financial incentives and cognitive abilities affect task performance in laboratory settings: an illustration. Econ. Lett. 85 (3), 315–320.
Sahm, C.R., 2012. How much does risk tolerance change? Q. J. Finance 02 (04), 1250020. https://doi.org/10.1142/s2010139212500206.
Schier, U.K., Ockenfels, A., Hofmann, W., 2016. Moral values and increasing stakes in a dictator game. J. Econ. Psychol. 56, 107–115. https://doi.org/10.1016/j.joep.2016.06.004.
Sefton, M., 1992. Incentives in simple bargaining games. J. Econ. Psychol. 13 (2), 263–276. https://doi.org/10.1016/0167-4870(92)90033-4.
Siegel, S., 1961. Decision making and learning under varying conditions of reinforcement*. Ann. N. Y. Acad. Sci. 89 (5), 766–783. https://doi.org/10.1111/j.1749-6632.1961.tb20177.x.
Siegel, S., 1964. Choice, Strategy, and Utility. McGraw-Hill, New York.
Siegel, S., Fouraker, L.E., 1960. Bargaining and Group Decision Making: Experiments in Bilateral Monopoly. McGraw-Hill.
Siegel, S., Goldstein, D.A., 1959. Decision-making behavior in a two-choice uncertain outcome situation. J. Exp. Psychol. 57 (1), 37–42. https://doi.org/10.1037/h0045959.
Slonim, R., Roth, A.E., 1998. Learning in high stakes ultimatum games: an experiment in the Slovak Republic. Econometrica 66 (3), 569. https://doi.org/10.2307/2998575.
Smith, V.L., 1976. Experimental economics: induced value theory. Am. Econ. Rev. 66 (2), 274–279.
Smith, V.L., 1982. Microeconomic systems as an experimental science. Am. Econ. Rev. 72 (5), 923–955. https://www.jstor.org/stable/1812014.
Smith, V.L., 1991. Rational choice: the contrast between economics and psychology. J. Polit. Econ. 99 (4), 877–897. https://doi.org/10.1086/261782.
Smith, V.L., 2017. Tribute to Sidney Siegel (1916-1961). South. Econ. J. 83 (3), 664–667. https://doi.org/10.2307/26632369. JSTOR.
Smith, V.L., Walker, J.M., 1993. Monetary rewards and decision cost in experimental economics. Econ. Inq. 31 (2), 245–261. https://doi.org/10.1111/j.1465-7295.1993.tb00881.x.
Stricker, L.J., 1967. The true deceiver. Psychol. Bull. 68 (1), 13–20. https://doi.org/10.1037/h0024698.
Svorenčík, A., 2016. The Sidney Siegel tradition: the divergence of behavioral and experimental economics at the end of the 1980s. Hist. Polit. Econ. 48 (suppl 1), 270–294. https://doi.org/10.1215/00182702-3619310.
Svorenčík, A., Maas, H. (Eds.), 2016. The Making of Experimental Economics: Witness Seminar on the Emergence of a Field. Springer.
Tisserand, J.-C., Cochard, F., Le Gallo, J., 2015. Altruistic or Strategic Considerations: A meta-Analysis on the Ultimatum and Dictator Games. CRESE, Université de Franche-Comté, Besançon.
Titmuss, R.M., 1970. The Gift Relationship : from Human Blood to Social Policy. Allen And Unwin, London.
Tversky, A., Edwards, W., 1966. Information versus reward in binary choices. J. Exp. Psychol. 71 (5), 680–683. https://doi.org/10.1037/h0023123.
Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. Science (1979) 185 (4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124.
Tversky, A., Kahneman, D., 1983. Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. Psychol. Rev. 90 (4), 293.
van Damme, E., Binmore, K.G., Roth, A.E., Samuelson, L., Winter, E., Bolton, G.E., Ockenfels, A., Dufwenberg, M., Kirchsteiger, G., Gneezy, U., Kocher, M.G., Sutter, M., Sanfey, A.G., Kliemt, H., Selten, R., Nagel, R., Azar, O.H., 2014. How Werner Güth's ultimatum game shaped our understanding of social behavior. J. Econ. Behav. Organ. 108, 292–318. https://doi.org/10.1016/j.jebo.2014.10.014.
Wason, P.C., Shapiro, D., 1971. Natural and contrived experience in a reasoning problem. Q. J. Exp. Psychol. 23 (1), 63–71. https://doi.org/10.1080/00335557143000068.
Weber, E.U., Blais, A.-R., Betz, N.E., 2002. A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors. J. Behav. Decis. Mak. 15 (4), 263–290. https://doi.org/10.1002/bdm.414.
Wienk, R.E., 1979. Measuring Racial Discrimination in American Housing Markets: The Housing Market Practices Survey (Vol. 444). Department of Housing and Urban Development, Office of Policy Development and Research, Division of Evaluation.
Wilcox, N.T., 1993. Lottery choice: incentives, complexity, and decision time. Econ. J. 103 (421), 1397. https://doi.org/10.2307/2234473.
Wilde, L.L., 1981. On the use of laboratory experiments in economics. Philosophy in Economics: Papers Deriving from and Related to a Workshop on Testability and Explanation in Economics Held at Virginia Polytechnic Institute and State University, 1979. Springer Netherlands, Dordrecht, pp. 137–148.
Wilson, R.K., & Isaac, R.M. (2007). Political economy and experiments. In the political economist: newsletter of the section on political economy.
Yerkes, R.M., Dodson, J.D., 1908. The relation of strength of stimulus to rapidity of habit-formation. J. Comparat. Neurol. Psychol. 18 (5), 459–482.