

The book cover features an abstract background with a large, vibrant red shape on the left side, transitioning into various shades of blue and purple on the right. The text is overlaid on this background.

Experimental Political Science and the Study of Causality

From
Nature
to the
Lab

Rebecca B. Morton
Kenneth C. Williams

CAMBRIDGE

Subjects' Motivations

In Chapter 1, we observed that one of the big differences between laboratory experiments conducted by political economists and those conducted by political psychologists is the use of financial incentives to motivate subjects. That is, in political economy laboratory experiments subjects' payments for participation are tied to the choices that they make, whereas in political psychology experiments subjects are typically paid a flat fee for participation or receive class credit. Why is there this difference and does it affect the validity of the experiments? In this chapter we consider these questions. We begin with the reasons why political economists use financial incentives.

10.1 Financial Incentives, Theory Testing, and Validity

10.1.1 How Financial Incentives Work in Theory Testing

Most political economy experiments involve either theory testing or stress tests of theories (see Chapter 6). The theories are largely based on formal theoretical foundations. The emphasis of the research is often on political and economic institutions (i.e., election systems, legislative committees, stock markets, first-price auctions, etc.) and the behavior of actors within those institutions. The theories make relationship (either comparative static or dynamic) and point predictions about how these institutions will affect human behavior. We discuss these types of predictions in Chapter 6.

Importantly, the theories assume that subjects have assigned particular values to each outcome in the theory, and that, given these values, the institutional differences have predictable effects on the subjects' choices. To conduct a theory-testing experiment, then, an experimentalist would like to induce subjects to have the same value orderings over outcomes as assumed in his or her theory, as we noted in Chapter 6. Moreover, the experimenter

wants to populate her institution with actors who make coherent and interpretable decisions. Doing so increases the construct validity (Chapter 7) of the experiment because it reduces the disconnects between the motivations of subjects and those assumed by the theory.

One way to induce these values is to use financial incentives. How does this work? Suppose an experimenter wants to construct an institution in the laboratory (such as an election system) and then manipulate certain variables (voting rules) and observe the effects on subjects' choices of variations on that institution (varying whether voters vote sequentially or simultaneously) while holding other variables constant (majority rule voting). The experimenter then wants to populate the experimental institution with actors to bring the institution to life (i.e., the experimenter needs voters to make decisions under the two different institutional procedures). In this case the focus is on how institutions affect human behavior (the experimenter wants to examine the impact of using sequential voting as opposed to simultaneous voting). This comparison is possible when the subjects' values for the outcomes of the voting are held constant.

That is, suppose the theory's prediction is that when voters have incomplete information about the options before them and place different values on the different outcomes, the option that is most preferred by voters in pairwise comparisons (the option that would win if it faced each other option one by one or the Condorcet winner) is more likely to be chosen under sequential voting than under simultaneous voting.¹ Suppose that an experimentalist conducts an experiment testing this prediction with three options labeled Blue, Yellow, or Green. If the experimentalist uses *experimenter-induced values*, then he or she assigns a financial value for each of the possible outcomes for each of the subjects. The experimenter can assign these values so that there is disagreement among the subjects, as assumed by the theory, and also control the information voters have about these values.

Definition 10.1 (Experimenter-Induced Values): *When an experimenter assigns specific financial values to outcomes in an experiment. These values are usually assigned in the context of a theory-testing experiment and the values are designed to mirror the assumed preferences of the actors in the theory.*

For example, the experimenter might have a total of 15 subjects divided into groups of 5 each. The experimenter might assign the first group to

¹ See Condorcet (1785).

each receive \$3 if Blue wins, \$2 if Yellow wins, and \$1 if Green wins. The experimenter might assign the second group to each receive \$1 if either Blue or Green wins, and \$3 if Yellow wins. And finally the experimenter might assign the last group to each receive \$1 if Blue wins, \$2 if Yellow wins, and \$3 if Green wins. In this setup we have both disagreement over the values of the outcomes and Yellow is the Condorcet winner if the subjects vote according to their induced values. That is, if Blue and Yellow were the only candidates, 5 of the subjects would vote for Blue and 10 would vote for Yellow; if Yellow and Green were the only candidates, again 10 of the subjects would vote for Yellow. The experimenter can then hold the information and the disagreement constant by holding the payoffs constant and compare the choices of the subjects under the two different voting systems. The financial incentives are often also called performance-based incentives. If the experimenter-induced values work (we define shortly what we mean by "work"), then the experimenter has achieved a high level of construct validity and can make the comparison between the voting systems.

What happens if the experimentalist simply pays the subjects a flat fee for participating and the outcome of the voting has no extrinsic value for the subjects – the experimenter does not explicitly attempt to induce values for the outcomes? It would be more difficult for the experimentalist to evaluate the theory. First, the experimentalist would have to figure out the subjects' preferences over the three options independent of the voting system – figure out the subjects' intrinsic motivations in the experiment to evaluate the theory's predictions. Assuming the experimentalist could do so, what happens if all the subjects are simply indifferent between the choices? Or what happens if in sequential voting all the subjects have the same values but in simultaneous voting the subjects disagree over the best option? The subjects may have as a goal finishing the experiment as soon as possible, which may outweigh intrinsic values they have over the colors, leading them to make choices that are easiest given the experimental setup. Or the subjects may be taking part in repeated elections with randomization as described earlier and may just vote for the candidate who lost the previous election because he or she feels sorry for that candidate. All of these things would create a disconnect between the theory tested and the experimental design, lessening the construct validity of the results.

10.1.2 Financial Incentives Versus Intrinsic Motivations

Some psychologists argue that reward schemes based on financial incentives may actually cause subjects to perform poorly in an experiment.

Psychologists differentiate between intrinsic and extrinsic motivations. Ryan and Deci (2000, p. 55) note: "The basic distinction is between intrinsic motivation, which refers to doing something because it is inherently interesting or enjoyable, and extrinsic motivation, which refers to doing something because it leads to separable outcomes." The authors go on to note that extrinsically motivated actions can be performed with "resentment, resistance, and disinterest". Some psychologists argue that when money is contingent on the actions of subjects within an experiment, then the intrinsic motivation is replaced by extrinsic motivation and the performance of subjects is negatively affected. Deci (1971, p. 108) comments: "If a person is engaged in some activity for reasons of intrinsic motivation, and if he begins to receive the external reward, money, for performing the activity, the degree to which he is intrinsically motivated to perform activity decreases."

A number of studies by psychologists have found evidence that financial incentives lower task performance by crowding out intrinsic motivations. Most of this research focuses on individualized decision making rather than on choices within the context of a group or game situation as in political economy experiments. A recent example is Heyman and Ariely's (2004) study of the consequences of varying payment levels on the performance of subjects engaged in individualized tasks which ranged from boring, repetitive ones to solving puzzle problems that progressed in difficulty during the experiment. They studied the effects of a small payment, a sizable one, and whether the payment was money or candy. They also ran the experiment without paying subjects for performance.

Heyman and Ariely found that when subjects were not given incentive payments (either money or candy), the number of completed tasks was higher than with small incentive payments. Furthermore, when the incentive payment was not explicitly monetary (i.e., candy), the performance was higher than in the small-monetary-payment condition. Increasing incentive payments of both types increased performance, although not always reaching the levels of task performance in the control condition with no payment. These results support the contention that financial incentives crowd out intrinsic motivations and lead to worse task performance.²

² Gneezy and Rustichini (2000a) found similar results when they compared no-payment treatments to insignificant small monetary payments. A reanalysis of the data by Rydval and Ortman (2004) suggests that these differences are more reflective of cognitive differences across subjects rather than payment treatment effects.

Four possible explanations for why explicit financial incentives may worsen task performance in experiments have been proffered. One is that the cognitive effort induced by the incentives may be counter-productive, causing subjects to "overthink" a problem and miss simple solutions as subjects try more complex cognitive strategies to maximize payoffs. Financial incentives may cause subjects to think they should exert more effort than necessary when simpler decision processes such as heuristics are sufficient.³ According to this explanation, we would expect that financial incentives are most harmful for simple, easy tasks or ones where cognitive shortcuts can be effective, even in a situation that is complicated.

A second proposed cause was suggested by Meloy et al. (2006), who found that financial incentives in experiments can elevate a subject's mood, which contributes to worsened task performance. Meloy et al. noted that the effect they and others found may be mitigated if the subjects receive feedback and experience. This suggests that financial incentives interact with feedback and experience, and failure to provide those additional features leads to inaccurate estimates of their effects. It is worth noting that the experiments conducted by economists that demonstrate advantages of financial incentives usually also include feedback and repetition, in contrast to the experiments conducted by psychologists that demonstrate disadvantages of financial incentives in which subjects typically complete tasks without such feedback and repetition. Sprinkle (2000) provided evidence in support of this hypothesis.

Endogeneity of social norm preferences has been projected as a third reason. In this view we think of the experimental subjects as workers and the experimenter as their employer. Some theorists have contended that firms who pay well regardless of performance can motivate workers by inducing them to internalize the goals and objectives of the firm, changing their preferences to care about the firm. If workers are paid on an incentive basis such that lower performance lowers wages, they are less likely to internalize these firm goals and there is less voluntary cooperation in job performance (see Bewley, 1999; James, 2005). Miller and Whitford (2002) made a similar argument about the use of incentives in general in principal agent relationships in politics.

Somewhat related is an explanation suggested by Heyman and Ariely (2004) based on their aforementioned experimental analysis. That is, they contend that, when tasks are tied to monetary incentives, individuals see

³ See Arkes et al. (1999) and Camerer and Hogarth (1999).

the exchange as part of a monetary market and respond to the incentives monotonically, but if the tasks are tied to incentives that do not have clear monetary value, individuals see the exchange as part of a social market and their response is governed by the internalization of social norms outside of the experiment.⁴

Finally, a fourth explanation of crowding out is informational. Benabou and Tirole (2003) showed that when information about the nature of a job is asymmetric, incentive-based payments may signal to workers that the task is onerous and, although increasing compensation increases the probability the agent will supply effort, it also signals to the agent that the job is distasteful and affects their intrinsic motivations to complete the task.

These last two explanations (the social norm perspective and the informational theory) also suggest a nonmonotonic relationship between financial incentives and task performance. That is, when financial incentives are introduced, but are small, subjects' task performance is worsened as compared to the no-payment condition (either because they now think of the exchange with the experimenter as a market one instead of a social one or because they see the task as more onerous than before), but as financial incentives are increased, task performance increases if the financial incentives are sizable enough.

10.1.3 Is Crowding Out by Financial Incentives a Problem?

The relevant question is whether money decreases the performance of subjects in experiments using experimenter-induced financial incentives. To answer this question, we must understand what controls and financial incentives are used for – to reduce performance variability in the data. That is, they are used to reduce randomness caused by subjects making choices outside of the realm of the theory. In a noteworthy study in political science, Prior and Lupia (2005) found that giving subjects financial incentives to give correct answers in a survey experiment on political knowledge induced

⁴ A number of studies show that individuals are more likely to volunteer and contribute to public goods when participation is not tied to financial incentives such as Titmuss's (1970) comparison of blood markets. More recently, Gneezy and Rustichini (2000a) found in a field experiment that the introduction of a fine for parents picking up children late from day-care centers increased the number of parents who came late. Brekke et al. (2003) presented a formal model in which financial incentives can have adverse effects on voluntary contributions because of moral motivations and provided survey evidence on recycling behavior and voluntary community work consistent with the model's predictions. Cappellari and Turati (2004) also found that volunteering in a variety of situations is higher when individuals are intrinsically motivated.

subjects to take more time and to give more accurate responses. Studies by economists suggest that performance-based incentives lead to reductions in framing effects, the time it takes for subjects to reach equilibrium in market experiments, and mistakes in predictions and probability calculations.⁵

Furthermore, a growing number of field and marketing experiments show that choices made by subjects in hypothetical situations are significantly different from the choices made by subjects in comparable real situations in which financial incentives are involved, suggesting that using hypothetical situations in place of financial incentives leads to biased and inefficient predictions about behavior. Bishop and Heberlein (1986) showed that willingness-to-pay values of deer-hunting permits were significantly overstated in a hypothetical condition as compared to a paid condition. List and Shogren (1998) found that the selling price for a gift is significantly higher in real situations than in hypothetical ones. List (2001) demonstrated that in a hypothetical bidding game bids were significantly higher than in one in which real payments were used. In marketing research, Ding et al. (2005) presented evidence that shows significantly better information is gathered on subjects' preferences over different attributes of meal choices when the meals are not hypothetical but real. And Voelckner (2006) found significant differences between consumers' reported willingness to pay for products in hypothetical choice situations as compared to real choices across a variety of methods used to measure willingness to pay in marketing studies. In a recent meta-analysis of experiments on preference reversals (situations in which individuals express preferences over gambles that are at odds with their rankings of the gambles individually), Berg et al. (2010) showed that when financial incentives are used, the choices of the individuals are reconcilable with a model of stable preferences with errors, whereas the choices of individuals where such incentives are not used cannot be reconciled. Thus, the evidence appears to support the conclusions of Davis and Holt (1993, p. 25): "In the absence of financial incentives, it is more common to observe nonsystematic deviations in behavior from the norm."

Fortunately, several systematic reviews of the literature have examined this question. In one survey article, Smith and Walker (1993b) examined 31 economic experimental studies on decision costs and financial incentives

⁵ See Brase et al. (2006); Hogarth et al. (1991); Gneezy and Rustichini (2000a); Levin et al. (1988); List and Lucking-Reiling (2002); Ordóñez et al. (1995); Parco et al. (2002); Wilcox (1993); and Wright and Aboul-Ezz (1988).

and concluded that financial incentives bolstered the results. They noted (pp. 259–260):

A survey of experimental papers which report data on the comparative effects of subject monetary rewards (including no rewards) show a tendency for the error variance of the observations around the predicted optimal level to decline with increased monetary reward. . . . Many of the [experimental] results are consistent with an “effort” or labor theory of decision making. According to this theory better decisions – decisions closer to the optimum, as computed from the point of view of the experimenter/theorist – require increased cognitive and response effort which is disutilitarian. . . . Since increasing the reward level causes an increase in effort, the new model predicts that subject’s decisions will move closer to the theorist’s optimum and result in a reduction in the variance of decision error.

This conclusion has found support elsewhere. Camerer and Hogarth (1999) reviewed a wide range of studies and found that higher financial incentives lead to better task performance. Hertwig and Ortmann (2001), in a similar review, found that when payments were used, subjects’ task performances were higher. Hertwig and Ortmann (2001) conducted a 10-year review of articles published in the *Journal of Behavioral Decision Making* (JBDM) and reviewed articles that systematically explored the effect of financial incentives on subject behavior. Similar to Smith and Walker’s assessment, they “conclude that, although payments do not guarantee optimal decisions, in many cases they bring decisions closer to the predictions of the normality model. Moreover, and equally important, they can reduce data viability substantially” (Hertwig and Ortmann, 2001, p. 395).

Of particular interest is the systematic review of Cameron and Pierce (1994, 1996) of approximately 100 experiments in social psychology and education. These researchers found that “[financial] rewards can be used effectively to enhance or maintain intrinsic interest in activities. The only negative effect of reward occurs under a highly specific set of conditions that be easily avoided” (Cameron and Pierce, 1996, p. 49). The negative effect that Cameron and Pierce make mention of is “when subjects are offered a tangible reward (expected) that is delivered regardless of level of performance, they spend less time on a task that control subjects once the reward is removed” (Cameron and Pierce, 1994, p. 395). In other words, flat payment schemes hinder subjects’ performance. Although some quibble about the methodology employed in these studies, it is clear that financial incentives based on performance have not had as negative an impact on subjects’ behavior as some psychologists have argued.

To ensure that positive intrinsic behavior is not crowded out, an experimenter can attempt to make the experiment interesting and avoid repetitive

tasks. As noted earlier, repetition is a hallmark for many experimental designs in political science and economics that test formal models. However, if the experiment is not interesting and the subjects are simply performing the same task repeatedly, then we can imagine cases in which intrinsic motivation will decrease and subjects will become bored and perform poorly. To avoid this type of behavior, experimental designs can incorporate greater randomness in treatments so that subjects are engaged in different tasks. Then performance-based financial incentives can ensure that the experiment is an interesting and enjoyable task for the subjects.

10.1.4 Induced Value Theory

The theory that reward media such as financial incentives can induce subjects to have preferences as theoretically assumed is called induced value theory. This theory was posited by Nobel Laureate Vernon Smith in a series of articles (Smith, 1976, 1982; Smith and Walker, 1993a). Smith (1982, p. 931) comments:

Control over preferences is the most significant element distinguishing laboratory experiments from other methods of economic inquiry. In such experiment, it is of the greatest importance that one be able to state that, as between two experiments, individual values (or derivative concepts such as demand or supply) either do or do not differ in a specified way. This control can be exercised by using a reward structure and a property right system to induce prescribed monetary value on (abstract) outcomes.

Or, as Friedman and Sunder (1994) note: “The key idea in induced-value theory is that proper use of a reward medium allows an experimenter to induce prespecified characteristics in experimental subjects, and the subject’s innate characteristics become largely irrelevant.” Therefore, if subjects’ motivations in the experiment are guided by the reward mechanism, then other factors such as altruism, revenge, and naivety will be ruled out.

Induced Value Theory postulates that four conditions should be considered when attempting to induce experimental motivations by a reward medium (such as money) in the laboratory. First, if a reward medium is *monotonic* then subjects prefer more of the medium to less. When financial incentives are used, monotonicity requires that subjects prefer more money to less. Second, if a reward medium is *salient*, then the rewards are a by-product of a subject’s labor or the choices he or she makes during the experiment. Reward mechanisms that are salient are also referred to as performance-based incentives because subjects earn rewards in the

experiment based on the decisions that they make. For example, in the experiment described earlier, the subjects would receive the dollar values assigned as payment for the election in which they participated depending on which candidate won the most votes. In cases for which the researcher uses repetition, usually the subjects' rewards may be accumulated over the experiment. Alternatively, sometimes a researcher may randomly choose one of the choices of the subjects for one of the periods to reward, as we discuss later. Third, if a reward medium is *private*, then interpersonal utility considerations are minimized. That is, subjects are unaware of what other subjects are awarded. And fourth, if a reward medium is *dominant*, then the choices made in the experiment are based solely on the reward medium and not some other factors such as the rewards earned by other subjects (i.e., a subject is not concerned about the payoffs of other subjects).

Definition 10.2 (Monotonicity): *Given a costless choice between two alternatives, identical except that the first yields more of the reward medium than the second, the first will be preferred over or valued more than the second by any subject.*

Definition 10.3 (Salience): *The reward medium is consequential to the subjects; that is, they have a guaranteed right to claim the rewards based upon their actions in the experiment.*

Definition 10.4 (Dominance): *The reward structure dominates any subjective costs (or values) associated with participation in the activities of the experiment.*

Definition 10.5 (Privacy): *Each subject in an experiment is only given information about his or her own payoffs.*

Smith did not specify that these four conditions were necessary conditions to control subject behavior but rather only sufficient conditions (Smith, 1982). Guala (2005) points out that these conditions are not hardened rules but actually precepts or guidelines on how to control preferences in experiments. He states (p. 233):

[F]irst, the conditions identified by the precepts [of induced value theory] were not intended to be necessary ones; that is, according to the original formulation, a perfectly valid experiment may in principal be built that nevertheless violates some or all of the precepts. Second, the precepts should be read as hypothetical conditions ("if you want to achieve control, you should do this and that") and

should emphatically not be taken as axioms to be taken for granted. . . . Consider also that the precepts provide broad general guidelines concerning the control of individual preferences, which may be implemented in various ways and may require ad hoc adjustment depending on the context and particular experimental design one is using.

These guidelines were set over a quarter of a century ago when the use of financial incentives in experiments was still relatively new. How do these guidelines hold up today for political scientists who wish to use experimenter-induced values? What implications do they have for experimental design choices?

Monotonicity and Salience

In our view, the two conditions of monotonicity and salience are intricately related. Given that empirical evidence rather overwhelmingly suggests that sufficient financial incentives can work to create experimenter-induced values, the conditions of monotonicity and salience together raise the following questions for experimentalists who use financial incentives: (1) How much total should subjects expect to earn on average for them to value their participation? (2) How much should subjects' choices affect their payoffs? We consider these questions in order.

How Much Should Subjects Be Paid on Average?

Undergraduate Subject Pools. When using undergraduate students, the standard norm among experimental political economists is to structure the experimental payments so that on average subjects earn 50% to 100% above the minimum wage per hour (see Friedman and Sunder, 1994, p. 50). But this is only a rule of thumb. Does this amount have any empirical justification? Gneezy and Rustichini (2000b) conducted experiments that considered how varying the reward medium affected student performance. They conducted an experiment in which students answered questions on an IQ test. Subjects were randomly assigned to one of four treatments that varied the reward medium. In all the treatments, subjects were given a flat sum payment and treatments varied over an additional amount that the subjects could earn depending on whether they answered questions correctly. In the first treatment, subjects were not given an additional opportunity to earn more; in the second treatment, subjects were given a small amount for each question they got correct; in the third treatment, subjects were given a substantial amount for each question they answered correctly; and in the fourth treatment, subjects were given three times the amount given in the third treatment for each correct question. The authors found that the

performance on the IQ tests of subjects in treatments 1 and 2 was essentially the same and significantly worse than in treatments 3 and 4.

The interesting finding is that there was no difference between the high-payoff conditions in treatments 3 and 4. Hence, what mattered in the experiment was that subjects who received substantive rewards performed better than subjects with minimum or no rewards, but there was no difference between the two types of substantive rewards. This finding suggests that financial incentives in the laboratory are not strictly monotonic in the sense that increasing the reward medium will increase the performance of subjects. Rather the subjects only have to perceive that the reward medium is sufficient. The authors foreshadow their conclusion with the title of their paper: "Pay enough or don't pay at all." This research suggests that the rule of thumb of "twice the minimum wage per hour" may be appropriate.

However, in contrast to these results, as observed earlier, Bassi et al. (2010; see Example 6.3) conducted an experiment on a voting game in which they varied the financial incentives paid to subjects. In one treatment, subjects were paid only a flat fee for participating; in another treatment, subjects were paid a normal experimental payment; and in a third treatment, the subjects were paid double the normal experimental payment. The authors also considered the effect of increasing the complexity of the voting game by reducing the information available to voters. They found a monotonic relationship between financial incentives and the tendency of voters to choose as predicted by the game-theoretic model. Furthermore, they found that this tendency was particularly strong in the complex game with incomplete information. These results suggest that, in game-theoretic experiments, particularly complex ones, increasing financial incentives does increase the attention of voters to the task. This analysis suggests that in complex games the researcher may want to pay subjects more than the standard twice the minimum wage.

Nonstudent Subject Pools. A more complex question is how much to pay nonstudent subject pools in the laboratory and how that would affect the comparison to student subjects. For example, in Palacios-Huerta and Volij's experiment with soccer players (Example 9.2), the soccer player subjects were paid the same amount as the students, yet arguably on average their income was significantly higher.⁶ The payments both the students and the

⁶ The incomes are not public information, but the authors estimate that the average income of the soccer players, excluding extra money for endorsements and the like, was between 0.5 and 2 million dollars.