REMEDIYING HYPOTHETICAL BIAS IN JURY SIMULATION RESEARCH

Min Ding
701A BAB
Marketing Department
Smeal College of Business
The Pennsylvania State University
University Park, PA 16802
Phone (814) 865-0622 Fax (814) 865-3015 email mqd9@psu.edu

Daniel R. Cahoy
310J BAB
Insurance and Real Estate Department
Smeal College of Business
The Pennsylvania State University
University Park, PA 16802
Phone (814) 865-6205 Fax (814) 865-6284 email dancahoy@psu.edu

March 1, 2005

* Dan Cahoy is Assistant Professor of Business Law and Min Ding is Assistant Professor of Marketing. The authors acknowledge financial support from the Smeal Small Research Grant Program, and helpful comments from Gary Bolton, Keith Crocker, Robert A. Prentice, Abdullah Yavas, and the participants of the Smeal Faculty Luncheon Colloquia.
ABSTRACT

It has been well documented in many contexts that systematic hypothetical bias exists when subjects are not incentive aligned. In situations where real world actions may impact the welfare of the decision maker, experimenters avoid hypothetical bias by creating an incentive structure that utilizes compensation or rewards based on participant’s actions and the outcomes in the experiment. However, no existing work has examined the use of such incentives in the context of a qualitatively different type of behavior in which real world decisions are not driven by self-interest, but rather by the impact on the welfare of others. A prime example is decision making in simulated or “mock” juries. In this paper, we propose a novel incentive structure that aligns experimental subjects to real world decisions that influence the welfare of two other parties. We present the results of two experiments using this incentive structure that demonstrate the existence of the hypothetical bias and the effectiveness of the incentive structure.
1. Introduction

A well-characterized problem in the experimental literature regarding human behavior is the difficulty in encouraging study participants in a laboratory setting to reveal their real-world responses. For example, while an economic theory underlying an observed problem may appear sound, attempting to replicate the results under controlled conditions is often frustrated by the fact that participants know the study is hypothetical. In a hypothetical setting, important real-world incentives may be non-existent, causing a participant to make choices other than he or she would if the conditions were real. In other words, responses that are incentive-compatible with those non-existent incentives may not occur, making the overall results of the experiment questionable.¹

One of the most important contexts for this phenomenon is experimental studies to uncover the biases, preconceptions and emotional triggers that influence the behavior of juries.² These studies are critically important to the judicial system for determining the best ways to mitigate (or at least anticipate) the effects of racism, sexism and economic bias. They are similarly vital for parties involved in litigation for predicting the likely outcome of a jury trial to


use in fashioning a litigation or negotiation strategy. To conduct controlled research with replicable results, simulated or “mock” juries are generally used. However, researchers in the field acknowledge that the hypothetical nature of jury simulation studies casts doubt on the results. The U.S. Supreme Court has even echoed such concerns. Unfortunately, existing economic models for ameliorating the effects of hypothetical bias are not useful in this context due to the unique incentive structure of juries. Specifically, juries are not motivated by the potential for personal gain, but instead are called upon to make just decisions for others. A novel approach to capture this incentive is called for.

This paper proposes a remedy for the problem of hypothetical bias in jury studies that is translatable to existing modalities. We begin by reviewing relevant literature, followed by a brief discussion of a mechanism designed to align incentives in jury studies. Finally, we present

---


4 See, e.g., Lockhart v. McCree, 476 U.S. 162, 171 (1986) (in criticizing jury studies cited to the court, Chief Justice Rehnquist noted that they “were based on the responses of individuals randomly selected from some segment of the population, but who were not actual jurors sworn under oath to apply the law to the facts of an actual case involving the fate of an actual capital defendant. We have serious doubts about the value of these studies in predicting the behavior of actual jurors.”).

5 The term “jury” is used herein to refer to petit juries as opposed to grand juries. A petit jury is a panel of persons from the community empanelled to determine factual issues in the context of a trial, a long-standing tradition in the United States. See Colgrove v. Battin, 413 U.S. 149, 176-77 (1973) (noting that the use of petit juries has been described at least as early as 1164).
the results of two experiments that demonstrate the utility of the incentive structure and provide a roadmap for future work in this area.

2. Relationship to the Previous Literature

Empirical evidence of hypothetical bias has been widely reported in a variety of contexts. Diamond and Hausman\(^6\) reviewed surveys that purported to gauge how much value people placed on public goods, such as cleaning up polluted rivers and lakes. They concluded that the discrepancy between the valuation stated under hypothetical condition and true valuation was so profound that information gleaned from contingent valuation surveys was actually \textit{worse} than no information at all. Bishop and Heberlein’s study found that the amount of money people were willing to spend for deer hunting permits was overstated when compared to a situation in which they would have to actually spend cash.\(^7\) List showed that sports card dealers significantly overstate their bids for a sports card in a hypothetical situation.\(^8\) List and Shogren found the selling price for a gift is significantly higher in an actual situation as compared to a hypothetical.\(^9\) And Ding, et al., recently found that subjects were more price-sensitive and had a different preference structure for Chinese dinner specials when their decisions resulted in real outcome (eating Chinese food).\(^10\)


\(^7\) See Bishop & Heberlein, \textit{supra} note 1, at 146-47.

\(^8\) See List, \textit{supra} note 1, at 1501-02.

\(^9\) List & Shogren, \textit{supra} note 1, at 1354.

\(^10\) Ding, et al., \textit{supra} note 1.
There are several possible specific factors that may cause or contribute to hypothetical bias. For example, participants in hypothetical settings are likely to shift their behavior to an overly socially-desirable presentation of themselves: “when incentives are low subjects say they would be more risk-preferring and generous than they actually are when incentives are increased.”\(^{11}\) Additionally, there may be less heterogeneity among the respondents (e.g., they may be more likely to conform to the social norm or to answers expected by their peers). In general, answers under hypothetical conditions may be inconsistent, erratic, and, in many cases, untrustworthy.\(^{12}\) Finally, participants may discount things that would be important in real-world contexts (e.g., budget constraints). To correct the hypothetical bias, experimenters normally incorporate incentives that parallel their real-world counterparts, and motivate the respondents by linking their decisions to a reward amount.

The phenomenon of hypothetical bias has been described in the context of jury simulation studies as well.\(^{13}\) To date, two streams of research have attempted to overcome the

---


\(^{13}\) Ronald Dillehay & Michael Nietzel, Constructing a Science of Jury Behavior, in Review of Personality and Social Psychology 253-54 (L. Wheeler ed., 1980) (discussing the problems with common jury research methods and stating that “We are skeptical that the prototypical jury analogue is capable of capturing the complex socialization processes which are produced by the jury experience.”); Shari Diamond, Illuminations and Shadows from Jury Simulations, 21 L. & Human Behavior 561, 566 (1997) (While concluding after reviewing the results of more detailed modern jury studies that “we can take some comfort that our efforts to invest more resources in the jury
hypothetical bias in jury simulations. The standard practice, favored by research psychologists and jury consultants, is to make the experimental conditions as real as possible (e.g., by using live actors, formal settings, etc.) while allowing the consequences of the participants’ decision to remain hypothetical.14 The idea is that more realistic surroundings will elicit more realistic responses by encouraging participants to forget (or at least downplay) the fact that the consequences are not real. However, there is no guarantee that subjects will behave more similarly to real life due to the increased realism given that there is no incentive to compel such behavior.15

To provide a more rigorous approach to correct the hypothetical bias, some researchers have adopted an incentive approach that rewards individuals personally based on the outcome of their decision.16 Guarnaschelli, McKelvey and Palfrey, for example, conducted an experimental study of the rules underlying jury voting behavior using opaque jars of colored balls representing the state of the world: blue (which was analogized to innocent) and red (analogized to guilty). One would obtain a “signal” as to the overall composition of the jar by individually viewing the color of one ball from the jar and vote accordingly. The incentive to vote correctly was provided

---

14 See Finkel, supra note 2, at 59-60.


by the fact that, for each round of voting, all jurors were paid fifty cents for a correct decision on the composition of the jar and five cents for an incorrect one.

We argue, however, based on the social literature, that straight application of personal incentives in jury simulation study may produce misleading results. This is because the missing incentive in jury simulation is not personal utility (as in the contexts commonly studied in the hypothetical bias literature), but rather social utility. Real jurors do not gain money by delivering a well-thought-out verdict, nor do they lose money even if they deliver a random verdict. Real jurors derive significant utility knowing that, if they make the right decision, both the plaintiff and defendant will receive what they truly deserve, which, obviously, is absent from conventional jury simulations.

Social utility literature argues such “non-self” utility is nontrivial, prevalent, and critical in shaping up individuals’ decisions. For example, subject systematically deviate in dictator and ultimatum games from normative predictions that are based on pure self-interest. Results from bilateral bargaining experiments (e.g., Roth\textsuperscript{17}; Camerer and Thaler\textsuperscript{18}) show that a fraction of the subjects care about more than just material payoffs to themselves. The empirical results of Kahneman, Knetsch, and Thaler\textsuperscript{19} are particularly relevant to jury simulations. They devised an


environment wherein an individual may choose one of two persons to share a certain amount of money, knowing that one had made an unequal offer in a previously played ultimatum game while the other made an equal offer. The payoff is such that they obtain more money if they choose to share with the first person (unequal offer). Nevertheless, a majority chose to share with the second person even though it means they also receive less money. This result demonstrates individuals are willing to sacrifice their personal gain to punish past unfair behavior, even when such behavior was not directed towards them.

Several theories have been proposed in the literature to account for behaviors driven by social utility. Rabin proposed a fairness equilibrium to account for the fact that people wish to reward those who are kind to them and penalize those who are malevolent. Fehr and Schmidt model the fairness as self-centered inequity aversion. They demonstrate that a wide variety of observations could be organized by their theory if different numbers of inequity aversion individuals are present in a given situation. Bolton and Ockenfels have shown the large set of seemingly disparate behavioral regularities that deviate from self-interest could be organized by an elegant theory based on equity, reciprocity and competition.

Due to the biases inherent in an incentive mechanism that rewards individuals directly based on their decisions, it is critical to devise a mechanism that simulates the social utility

---


experienced by real jurors if an understanding of the true behaviors of individuals in jury simulations is to be achieved.

3. An Incentive Aligned Mechanism for Social Utility

We propose here a mechanism that will mimic the social utility a real juror experiences. This mechanism is described in the context of a real-world jury problem that permits a demonstration of the magnitude of hypothetical bias addressed. We use an intellectual property infringement case in which a juror is required to assess the appropriate amount of money (damages) to award to the plaintiff (i.e. subsequent to a determination of liability). The participants are asked to apply unique and complex law and facts learned substantially from the information presented during the study. Social utility is mimicked by establishing that a certain amount of money is at issue in the case and must be divided between two parties (the plaintiff receives what is required to compensate for the defendant’s wrongful act, and the defendant keeps the rest). There is a legally correct division of money (the “reference decision”), and any deviation from that will unfairly punish one party while rewarding the other party undeservedly.

The mechanism we propose — explained to the mock jurors before a jury simulation study and imposed after they complete their verdicts — involves the following steps: (1) experimenters identify two parties a priori, such that one party is generally considered to be deserving of a monetary gift (such as a well-respected charity like UNICEF), and the other party is generally considered to be undeserving of such largesse (such as a passerby in the street); (2) upon completion of the experiments, the experimenters randomly pick a mock juror and compare

---

23 These two parties do not correspond to the two parties involved in litigation (defendant and plaintiff), but rather, they represent two states of outcome (a desirable outcome and an undesirable outcome).
his or her verdict to the reference decision (e.g., legally correct decision), and the percentage of deviation from the reference decision is calculated (e.g., 20%); (3) experimenters then divide a certain amount of money (e.g., $200) such that the undesirable party (e.g., passerby) receives the total amount multiplied by the deviation (e.g., $200*20%=$40), and the desirable party (e.g., UNICEF) receives the rest ($140). This mechanism thus rewards a mock juror with social utility, the magnitude of which (the amount of money the desirable party receives) is dependant on the mock juror’s ability to deliberate the case in a legally correct manner. The mechanism is flexible and can be used to accommodate specific cases by changing the following three elements, (1) the two parties used, (2) the total amount of money to be divided, and (3) the relationship between the deviation and division of money.

4. Experiment One

To test the effectiveness of our proposed incentive structure, we conducted an experiment that implements the proposed incentive structure and compares the results with a control group under a conventional hypothetical condition.

4.1. Experimental Design

We discuss here two important aspects of the experiment aimed to achieve higher external validity. Both aspects, the construction of the experimental context (legal case) and

Of course, a person who does not care about social utility will behave no differently when this mechanism is used, but this is quite realistic, as such a person would not take social utility into consideration if that person were a real juror.
individual decision making in a group environment, are related to the realism of jury deliberation.

We undertook to create a very realistic and complex legal case, loosely based on the facts of a real litigation that took place in the U.S. District Court for the District of Massachusetts (see Appendix 1).\textsuperscript{25} This allows us to simulate the conditions of a typical jury study to ensure that any observed effects could be generalized to such studies. It also ensures that subjects were not tested simply on abstract math skills by requiring subjects to grasp legal and factual aspects of the case as real jurors do.

The hypothetical case involved a patent on a method of producing in tissue culture cells large quantities of erythropoietin (a protein that stimulates red blood cell production). The patent’s owner, called Genomik in the experiment, sued a competitor, called CellPro in the experiment, for making use of erythropoietin production methods that allegedly infringed Genomik’s patent. The study participants were given detailed facts surrounding the development and sale of the two companies’ erythropoietin products as well as the law regarding patent infringement and damages.\textsuperscript{26} The participants were informed that liability phase of the case was already decided, with CellPro found to have infringed Genmomik’s patent, and the only remaining task was to decide the damages CellPro owed for the infringement over the years.

\textsuperscript{25} Amgen, Inc. v. Hoechst Marion Roussel, 126 F. Supp. 2d. 69 (D. Mass. 2001). Note, however, the alleged acts of infringement described in our experiment were entirely fictional and should not be attributed to the either of the parties involved in that case. Because the case was more complicated and ended in a bench trial rather than a jury trial, the outcome is not directly comparable to our experimental results.

infringement was divided into six specific time periods, each of which required an answer.

Determining damages in a case like this is often quite complex, hinging on such facts as a party’s actual sales and whether the plaintiff would have been able to make additional sales if the defendant had not been on the market. The facts were written such that a single legally correct\textsuperscript{27} answer existed for each of the six time periods, but could be derived only with significant mental effort. We summarize the key elements in each problem and the correct and most likely incorrect answers in Table 1.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
\textbf{Time Period} & \textbf{Correct Answer} & \textbf{Most Likely Incorrect Answer} \\
\hline
Period 1 & $100,000$ & $50,000$ \\
Period 2 & $150,000$ & $75,000$ \\
Period 3 & $200,000$ & $100,000$ \\
Period 4 & $250,000$ & $125,000$ \\
Period 5 & $300,000$ & $150,000$ \\
Period 6 & $350,000$ & $175,000$ \\
\hline
\end{tabular}
\end{table}

Actual jurors make decisions in a group environment wherein they are exposed to each other’s opinions while arriving at their own final determination. This is a key element of juror deliberation and we wanted to replicate it in the experiment in a carefully controlled manner. One approach would be to assign individuals to a group and allow them to talk within their group before final deliberation. However, this has two major limitations: (1) it is hard to control the discussion and dynamics within each group, and (2) the group effectively becomes the unit of analysis, which makes it prohibitively expensive to obtain multiple data points in the experiment. Therefore, we designed an alternative approach that captures the essence of group decision-making (exposure to other participants’ opinions during decision making) in a well-controlled environment.

\textsuperscript{27} The legally correct damages amount could be determined by applying the law to the facts contained in the test instrument.
This new design involves providing study participants with the additional information to which one would be exposed in a real jury, specifically, multiple (different) answers to a given question and the percentage of people in the group that adhere to each. This design allows us to precisely control the amount of information each participant is exposed to. In the experiment, we provide two to three possible answers for each of these questions, each answer being associated with a percentage of prior jurors who have selected it. The various answers were obtained from the most common calculations observed in a pre-test (e.g., calculations producing the correct answer and the most common mistakes, see Table 1).

However, the choice of percentage presents a dilemma regarding whether the correct answer should always be assigned to the highest percentage (i.e., treating it as the majority opinion). If we do, the simple, effortless, and commonly adopted strategy of agreeing-with-the-majority will necessarily lead to the correct answer and confound the hypothetical bias to be tested. Interestingly, it is hard to predict, a priori, who is more likely to adopt this strategy, a subject in the hypothetical condition or a real juror. Nevertheless, one can assume that the outcome (how many individuals reach the correct answer) could be contaminated. Conversely, the agreeing-with-the-majority would not effect the outcome if the correct answer is not presented as the majority opinion (as the previously-described strategy will necessarily lead to incorrect answer). However, subjects may become suspicious of the overall test design if “majority” answers are never correct.

To address this issue, we randomly selected half of the questions from the case and made the majority answer correct (1, 2, and 4), while half of the questions were written with an incorrect majority (3, 5, and 6). We incurred a cost with this solution, as we could no longer obtain meaningful interpretations from subjects’ answers to the majority-correct questions due to
the confounding effect. Since these are the only responses likely to demonstrate the effects of the incentive system, we focus our analysis and discussion on majority-incorrect questions (but for completeness, we have included the analysis for Questions 1, 2, and 4 in Appendix 3)\(^2^8\).

4.2. Experimental Procedure

Subjects were undergraduate business students at a large U.S. university. They randomly signed up for one of two sessions. A total of 25 subjects showed up for the first session and were used for the hypothetical condition. After signing an informed consent form, each subject was given a handout consisting of general instructions, a summary of facts and law involved in the case, and a special verdict form (See Appendix 1). The subjects were permitted to read and complete the special verdict form at their leisure. It took about 20 to 40 minutes for all subjects to complete the study.

A total of 28 subjects showed up for the second session, which was used for the incentive condition. Before the start of the session, one subject was asked to go outside the classroom and find a student at random walking by the building. That subject was to tell the potential recruit that he or she would receive a minimum of $10 for coming to the classroom and remaining for about 45 minutes to an hour (during which time that person could quietly do as he or she wished). The recruit was also informed that there was a possibility that he or she would receive

\(^{28}\) A subject has the option to accept majority opinion even if it is different from his/her own stated answer on the verdict form (in “Mark Your Vote”, option 3). In the experiments described below, however, only a negligible number of such instances exists and there are no significant difference between the hypothetical and incentive conditions. Thus, the data analysis presented in this paper only uses the actual answers each subject stated regardless whether they chose to defer to majority opinion or not. This helps to capture individual decision outcome, but the results are substantially unchanged if we replace these answers with majority opinion.
substantially more money — up to $100 — depending on the outcome of the experiment going on in the classroom. A female undergraduate student was eventually recruited and introduced to all subjects in the classroom; none of the subjects was acquainted with the recruit.

The same experimental handouts as used with the hypothetical group were then distributed to the subjects. They were asked to read the general instructions carefully, scan the summary and special verdict form, and then stop. The experimenter then put $100 in cash on the table in front of the class and described on the blackboard, lecture style, how their answer will be used to divide this $100: Two parties were designated, the first being the recruit in the room, the second being the Make-A-Wish foundation (introduced by showing the subjects the Foundation’s website on a large projection screen which detailed its mission). One of the subjects would be randomly selected at the end of the experiment, and one out of the six answers on that subject’s verdict form would be randomly chosen. The subject’s answer for that selected question would be used as the basis for dividing the $100 between the two parties (unless the subject marked an option that stated “I will agree with the majority even though my answer is different,” in which case the majority answer would be used instead).

Subjects were shown by example how the experimenter would calculate the percentage of deviation (if any) between the subject’s answer and the correct answer: If the subject’s answer is $2500, and the correct answer is $2000, the deviation is (2500-2000)/2000=25%. The result would be the same if the subject’s answer were 25% below the correct answer. The deviation is bounded by 100% more than the correct answer; any answers above that will be treated as 100% deviation (e.g., a $10000 answer has an effective deviation of 100% even though (10000-

29 The Make-A-Wish Foundation is a well-known national charity that grants the last wishes of terminally ill children. The authors are not aware of any political or religious affiliation on the part of the Foundation.
2000)/2000=400%). The product between the $100 cash and the percentage of deviation would be the amount of money presented to the student recruit, and the remainder would be sent to the Make-A-Wish foundation. In the preceding example, $25 would go to student recruit and $75 to the Make-A-Wish foundation. Additional examples were given where the subjects were asked to figure out the resulting division to insure they understood the incentive.

4.3. Results

We predicted that subjects under the incentive conditions would behave differently, even though they do not benefit personally from the outcome. A direct prediction of the result of this enhanced effort is that more subjects are likely to obtain the correct answer. To verify this prediction, we examined the three questions that were free of the confounding effects discussed above (i.e., the majority answers were not correct), questions 3, 5, and 6 (results for Questions 1, 2, and 4 have been included in Appendix 3 for completeness).

We first looked at the performance across each question. To examine the performance across question, we counted the number of subjects with the same answers, and tabulated the numbers into four groups corresponding to either one of the three presented answers or all other answers that did not match any of the presented ones (See Table 2).

As predicted, subjects in the incentive condition do better for questions 5 and 6. In question 5, 64% of subjects in incentive conditions obtained the correct answer, compared to 36% in the hypothetical condition, and the difference is statistically significant based on chi-square test.
For question 6, 61% of subjects in the incentive condition were correct, compared to 44% in the hypothetical condition, and this difference is significant (p=0.07). These findings provide strong evidence that individuals indeed act differently under the proposed incentive structure as compared to the pure hypothetical condition. An interesting observation is that for question 5, but not for question 6, the number of subjects that have the same opinion as majority is significantly higher in the hypothetical condition compared to the incentive condition (p=0.04).

It should be noted that the result for question 3 is not informative due to the construction of the instrument. Following the experiment, we interviewed subjects and discovered that the question was ambiguous. We determined that an answer we designated as wrong — the majority answer — could in fact be considered legally correct if the question were read a certain way (i.e., subjects who put in the most effort may have chosen this answer). As interpreting the results depends on having a single, unambiguously correct answer, we believe that the responses for this question are unclear at best.

We next examined performance at the individual level. To do this, we counted the number of the questions in which a subject provided the correct answer (0, 1, 2, or 3). Given our assessment that question 3 was ambiguous, we also examined the result for questions 5 and 6 only (0, 1 or 2). Both results are included in Table 3. Clearly, subjects in the incentive condition

---

30 We used the percentage in the hypothetical condition as the expected percentage for the incentive-aligned condition.
performed better in each individual’s total correct answers, and the difference is statistically significant based on chi-square test (p=0.06). The same observation holds when question 3 was removed (p=0.00). These results suggest that the incentive-aligned group appeared to have a better grasp of the facts and law presented in the summary and worked harder to obtain the correct result. This higher performance occurred without the incentive of any personal gain, as in the case of an actual jury.

Lastly, we examined the average answers of question 5 and question 6. The average is a useful measure because it allows us to determine whether the group as a whole has moved toward a more accurate result, or if the numbers of correct or near correct answers are washed out by the magnitude of incorrect ones. If one were attempting to obtain an overall picture of the group response — such as determining a likely jury verdict for a given group of individuals — one would wish to know whether the results could be different in the incentive-aligned group. Because the amount of deviation is not necessarily a function of effort, it is not as easy to predict the outcome. It may depend on the specific question asked.

We plotted (see Figure 1), for questions 5 and 6, the correct answer, the average of those
in the incentive condition and those in the hypothetical condition.\textsuperscript{31} As suspected, results for questions 5 and 6 are not the same. The average in the incentive condition is statistically smaller than the correct answer (p=0.02), and it is marginally statistically higher than the average in the hypothetical condition (p=0.08). For question 6, however, the averages in the incentive and hypothetical conditions are not statistically different (p=0.31).

While experiment 1 provides strong support for our prediction and the validity of the incentive structure, it does have a major limitation. While the Make-A-Wish Foundation appears to be an appropriate party to inspire largess because it is almost universally acceptable to potential subjects and accessible to future experimenters and practitioners, the other party we used, a randomly selected passerby, may not be consistent. Clearly, there could be a huge variation in terms of what kind of person that is selected, and some may confound the results. For example, although we do not believe it occurred in experiment 1, it is possible that an unusually physically desirable or needy passerby may make subjects indifferent between dividing money between the Foundation and the passerby. Therefore, we decided to modify this and other aspects of the experiment in the next round of studies.

5. Experiment Two

In experiment 2, we addressed the following issues: First, we wanted to replicate the positive results from experiment 1. Second, we wanted to remedy the ambiguous nature of Question 3 in experiment 1. Third, we wanted to use a more objective party than the random

\textsuperscript{31} We omitted Question 3 here for two reasons, first, the question is ambiguous as discussed above, and second, we plotted the results for both experiment 1 and experiment 2 in the same graph to ease of comparison and Question 3 in experiment 2 has a different correct answer.
passerby, one that would not be affected by the subjective nature of the selection process.

Finally, we wanted to investigate the underlying reasons for the differences between the hypothetical and incentive-aligned group. Since the incentive mechanism used awards social utility, it is conceivable that it will make a bigger difference for those subjects who care more about social welfare. The analysis here focuses on Questions 3, 5, and 6, but the results for the other three have been included in Appendix 3 for completeness.

5.1. Experimental Design and Procedure

The instrument used in experiment 2 was the same as in experiment 1, except for the following: (1) we changed question 3 to eliminate the ambiguity. The legal/factual issue is that a reasonable probability of lost sales cannot be defeated by the mere possibility of a different outcome (compare to Table 1). The correct answer is $1,800,000, and two likely incorrect answers are $990,000(accord improper weight to speculative argument) and $180,000(award royalty for all sales); (2) we used a new implementation of the incentive structure; and (3) we added a brief survey to understand subjects’ preference of socially desirable behaviors (Appendix 2).

Subjects were undergraduate business students recruited from a large U.S. university approximately two months after the completion of experiment 1 (none of them participated in Experiment 1). They randomly signed up for one of two sessions, each of which was further split into two groups. To ensure that there was no difference between the two groups in each session, we first placed all subjects in a single room and randomly selected half of the subjects and moved them to a different room. One room was used for the hypothetical condition and one
for the incentive condition. There were a total of 33 subjects in the hypothetical condition and 32 subjects in the incentive condition.

For the incentive condition, the subjects were informed as to how $100 would be divided between two parties as in experiment 1 except that the counter party to the Make-A-Wish foundation would no longer be a randomly selected passerby. The subjects were told instead that any deviation of the randomly selected result from the correct answer would result in the corresponding money being completely wasted. Specifically, the money would be used to purchase expensive coffee (or coffee beans) from a Starbucks café on campus, which would then be promptly dumped down a drain (coffee) or into a trashcan (coffee bean). The subjects were told that this would be done immediately following the experiment, and they were invited to stay and watch. In the experiment, the randomly selected answer in session 1 resulted in an $83 deviation. Two subjects were paid $5 to go to the café afterwards and purchased $83 worth of coffee beans that was then dumped into the trashcan in the classroom upon their return. The random answer in session 2 was 100% correct, thus all $100 went to Make-A-Wish foundation.

After each subject completed his or her deliberation, the subject was asked to take a short survey after turning in the verdict form. The short survey was intended to measure social desirability by employing a summated ratings scale purporting to measure the degree to which people describe themselves in socially acceptable terms to gain the approval of others.\footnote{See Gordon Bunner, et al., Marketing Scales Handbook: A Compilation of Multi-Item Measures (2001).} It was originally developed by Crowne and Marlowe\footnote{Douglas P. Crowne & David Marlowe, A New Scale of Social Desirability Independent of Psychopathology, 24 J. Consulting Psych. 349 (1960).} and has been used extensively in the behavioral
5.2. Results

Similar to experiment 1, we first examined the aggregate outcome for questions 3, 5 and 6 by problem (Table 4). Consistent with results in experiment 1, we observe 59% correct answers in the incentive condition for question 5, versus 36% in the hypothetical condition. This difference is statistically significant (p=0.01). The number of people who selected the majority answer was statistically smaller in the incentive-aligned condition as compared to the hypothetical condition (p=0.01). For question 6, it is 56% versus 42%, however, unlike experiment 1, the difference was not statistically significant (p=0.11). On the other hand, the number of people who selected the majority answer is statistically smaller in the incentive-aligned condition as compared to the hypothetical condition (p=0.02). Interestingly, the percentage of subjects who obtained the correct answers under either condition was very close in both experiments. For question 5 in experiment 1, the percentage of correct answers was 64%

and 36% for incentive and hypothetical conditions, respectively. For question 6 in experiment 1, it’s 61% and 44%, respectively. This high consistency provides another piece of evidence for the robustness of our hypothesis and experiment. An interesting observation is that this consistency also holds true for those who have selected the majority opinion (which is wrong in these questions), but only for those in the incentive condition. To a large extent, this supports the notion that results in the incentive condition are more consistent across different runs of the same experiment (with different subjects), while the hypothetical condition subjects are driven by various home grown preferences and their actions may be less consistent when they did not obtain correct answers.

As predicted, after we revised question 3 based on subject feedback, we found the percentage of subjects who obtained the correct answer to be significantly higher in the incentive condition (47%) as compared to that in the hypothetical condition (30%) (p=0.04). This result confirms the observations for questions 5 and 6 from both experiments.

When we examined the data at the individual level, consistent with Experiment 1, we found significant improvement of performance (p=0.00).

------------------------
Insert Table 5 Here
------------------------

The averages of the answers for each question are plotted in the same figure as those from experiment 1, for ease of cross-experiment comparison (Figure 1). Remarkably, the averages in the incentive condition are very similar between the two experiments for questions 5 and 6, respectively. Again, this provides additional evidence of the robustness of our incentive structure. For question 5, the average in the incentive condition is significantly lower than the
correct answer (p=0.04), but significantly higher than the average in the hypothetical condition (p=0.00). This result is consistent with what we have found in experiment 1. For question 6, the averages in the incentive and hypothetical are not statistically different from each other (p=0.68). For question 3, the average in the incentive condition is significantly lower than the correct answer (p=0.00), but significantly higher than the average in the hypothetical condition (p=0.02).

Finally, we investigated our hypothesis that the individuals on whom the incentive structure has the greatest effect are those who care a great deal about third-party utility and doing the right thing for others. We calculated the total scores (out of a possible 33) on the social desirability scale for each subject and plotted the distributions of these scores for both hypothetical and incentive conditions (Figure 2).

------------------------
Insert Figure 2 Here
------------------------

The average scores across subjects are 13.9 with standard deviation of 5.7 in the hypothetical condition, and 14.5 for the incentive condition with standard deviation of 5.5. Although it may appear that the incentive condition has more subjects who exhibit social desirability, the average of the scores are not statistically different (p=0.35). The measures of social desirability also served to provides evidence that samples for both conditions are equivalent. Since the distribution of scores in both conditions is close to normal with symmetric tails on both sides, we chose to divide the subjects into three groups, those (roughly) within one standard deviation of the average, those (roughly) above one standard deviation from the average, and those (roughly) below one standard deviation from the average. We then calculated the individual level
performance (how many questions a subject has answered correctly) for subjects within each of the three groups, for both hypothetical condition and incentive condition (Table 6).

------------------------
Insert Table 6 Here
------------------------

While we observe the average performance in the incentive condition is better than that in the hypothetical condition in all three groups, the only group where such an increase is significant is for subjects in the high social desirability group (1.71 for the incentive condition, compared to 0.75 in the hypothetical condition). This result supports our hypothesis that, if the incentive condition makes people work harder leading to a higher probability of correct answer, the difference with the hypothetical groups is expected to be most significant for individuals with high social desirability values.

6. Discussion and Conclusion

The economic behavior of participants in studies intended to simulate contexts wherein a person other than the participant is affected by the outcome is a relatively unexplored field. The experiments herein demonstrate that such behavior can be significantly influenced by the presence of an incentive alignment mechanism. Such a mechanism can cause participants to perform more accurately, suggesting that they commit more time and effort. The fact that the incentives are similar in nature to those experienced in real world contexts creates a strong presumption that their use can make simulations more realistic and useful.

The results are likely to have particular importance in the context of jury simulations. Americans rely on the results of such studies to determine how to reform the judicial system.
Questions such as what types of trial procedures may evoke prejudice or to what degree to jurors benefit from note taking are paramount for both real and perceived equity. They are substantially informed by simulations, and it is essential that the accuracy of this work be raised to the highest level possible. Moreover, the business world is dramatically affected by litigation; knowing when to settle or pursue one’s case in court often hinges on predictable jury information, a product of simulations. We believe that the use of our incentive alignment mechanism may be essential to obtaining the most accurate results from studies concerning groups like juries. Although ignoring the hypothetical bias occurring in the absence of the mechanism may not be significant in every case, it is impossible to know for sure. And in general, the use of such incentives can be a no-lose situation.

One caution must be imparted regarding the meaning and transferability of the results. Although the incentive-alignment mechanism is in the same category as the incentive experienced by an actual jury, and we have demonstrated that the mechanism does indeed increase subject performance, we could not state that this performance is actually closer than the hypothetical group to that of an actual jury. To make such a determination, one would be required to either (1) conduct several studies in the context of on-going litigations in the hope that one of the cases studied happened to be one of the very few litigations that ended in a jury trial as opposed to settlement or (2) reverse-engineer a realistic jury study starting with a decided case. The former would be prohibitively expensive and the latter would be exceedingly difficult to conduct in an objective manner, knowing the outcome of the litigation.

It is logical to at least make the assumption that real juries are more accurate than hypothetical juries, but they may be less accurate than subjects using our mechanism, particularly if the incentive is too strong. Unfortunately, this complication is highly fact specific,
and would be impossible to prove conclusively without a real jury decision for comparison in each case. As described above, this is generally impossible to achieve. To ensure that such an aberration does not occur, we suggest that experimenters choose an incentive that is less strong than that experienced by an actual jury (though necessarily stronger than the hypothetical).

We hope that this work will be carried on and expanded upon in many contexts. Potential routes for additional studies include the use of our incentive-alignment mechanism to uncover more accurate assessments of jury bias and new mechanisms to defeat hypothetical bias.
References


Figure 1. Average Answers for Each Question

![Graph showing average answers for each question with different conditions.](image)

- Correct
- Exp 2 -- Hypo
- Exp 2 -- Incentive
- Exp 1 -- Hypo
- Exp 1 -- Incentive

Figure 2. Distribution of Socially Desirable Answers for Each Subject

![Bar chart showing the distribution of socially desirable answers.](image)

- Hypothetical
- Incentive
### Table 1. Experiment 1 Problems

<table>
<thead>
<tr>
<th>Prob.</th>
<th>Legal/Factual Issue</th>
<th>Correct Answer</th>
<th>Answers from Common Mistakes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>If the plaintiff can make no sales, a royalty is the sole measure of damages</td>
<td>$9000</td>
<td>$90,000 (award plaintiff lost profits damages)</td>
</tr>
<tr>
<td>2</td>
<td>Plaintiff can obtain lost profits for defendant’s sales within plaintiff’s excess sales capacity; royalty for the rest</td>
<td>$495,000</td>
<td>$900,000 (award plaintiff lost profits for all sales) $90,000 (award royalty for all sales)</td>
</tr>
<tr>
<td>3</td>
<td>For lost profits damages, plaintiff must demonstrate by a preponderance of the evidence a reasonable probability of lost sales</td>
<td>$990,000</td>
<td>$1,800,000 (award plaintiff lost profits for all sales) $180,000 (award royalty for all sales)</td>
</tr>
<tr>
<td>4</td>
<td>Evidence sufficient to justify lost profits must be relevant</td>
<td>$1,350,000</td>
<td>$945,000 (misread facts and award royalty for 500 units) $1,800,000 (award plaintiff lost profits for irrelevant speculation)</td>
</tr>
<tr>
<td>5</td>
<td>A reasonable probability of lost sales cannot be defeated by the mere possibility of a different outcome</td>
<td>$1,800,000</td>
<td>$900,000 (accord improper weight to speculative argument) $2,700,000 (award plaintiff lost profits based on excess capacity instead of lost sales)</td>
</tr>
<tr>
<td>6</td>
<td>One cannot obtain any damages after the expiration of a patent, even if the defendant was unaware of the expiration</td>
<td>$1,080,000</td>
<td>$1,800,000 (award lost profits for period after patent expires) $990,000 (award lost profits for half sales instead of sales occurring in half year)</td>
</tr>
</tbody>
</table>
Table 2. Cross Question Results in Experiment 1

<table>
<thead>
<tr>
<th>Question 3</th>
<th>Question 5</th>
<th>Question 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>H. Incentive</td>
<td>H. Incentive</td>
<td>H. Incentive</td>
</tr>
<tr>
<td>Majority Opinion</td>
<td>9 (36%)</td>
<td>10 (36%)</td>
</tr>
<tr>
<td>2nd Common Opinion*</td>
<td>10 (40%)</td>
<td>10 (36%)</td>
</tr>
<tr>
<td>3rd Common Opinion</td>
<td>3 (12%)</td>
<td>2 (7%)</td>
</tr>
<tr>
<td>Other Opinions</td>
<td>3 (12%)</td>
<td>6 (21%)</td>
</tr>
<tr>
<td>Total</td>
<td>25 (100%)</td>
<td>28 (100%)</td>
</tr>
</tbody>
</table>

* The 2nd common opinion is the correct answer for all three questions.

Table 3. Individual Performance in Experiment 1

<table>
<thead>
<tr>
<th>Number of Questions Correctly Answered</th>
<th>Questions 3, 5, 6</th>
<th>Questions 5, 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hypothetical</td>
<td>Incentive</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>0</td>
<td>7 (28%)</td>
<td>4 (14%)</td>
</tr>
<tr>
<td>1</td>
<td>9 (36%)</td>
<td>7 (25%)</td>
</tr>
<tr>
<td>2</td>
<td>7 (28%)</td>
<td>13 (46%)</td>
</tr>
<tr>
<td>3</td>
<td>2 (8%)</td>
<td>4 (14%)</td>
</tr>
</tbody>
</table>
Table 4. Cross Question Results in Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>Question 3</th>
<th></th>
<th>Question 5</th>
<th></th>
<th>Question 6</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hypothetical</td>
<td>Incentive</td>
<td>Hypothetical</td>
<td>Incentive</td>
<td>Hypothetical</td>
<td>Incentive</td>
</tr>
<tr>
<td>Majority Opinion</td>
<td>14 (42%)</td>
<td>14 (44%)</td>
<td>17 (52%)</td>
<td>9 (28%)</td>
<td>13 (39%)</td>
<td>6 (19%)</td>
</tr>
<tr>
<td>2nd Common Opinion*</td>
<td>10 (30%)</td>
<td>15 (47%)</td>
<td>12 (36%)</td>
<td>19 (59%)</td>
<td>14 (42%)</td>
<td>18 (56%)</td>
</tr>
<tr>
<td>3rd Common Opinion</td>
<td>4 (12%)</td>
<td>1 (3%)</td>
<td>0 (0%)</td>
<td>3 (9%)</td>
<td>3 (9%)</td>
<td>4 (13%)</td>
</tr>
<tr>
<td>Other Opinions</td>
<td>5 (15%)</td>
<td>2 (6%)</td>
<td>4 (12%)</td>
<td>1 (3%)</td>
<td>3 (9%)</td>
<td>4 (13%)</td>
</tr>
<tr>
<td>Total</td>
<td>33 (100%)</td>
<td>32 (100%)</td>
<td>33 (100%)</td>
<td>32 (100%)</td>
<td>33 (100%)</td>
<td>32 (100%)</td>
</tr>
</tbody>
</table>

* The 2nd common opinion is the correct answer for all three questions.

Table 5. Individual Performance in Experiment 2

<table>
<thead>
<tr>
<th>Number of Questions Correctly Answered</th>
<th>Questions 3, 5, 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hypothetical</td>
</tr>
<tr>
<td>0</td>
<td>9 (27%)</td>
</tr>
<tr>
<td>1</td>
<td>15 (45%)</td>
</tr>
<tr>
<td>2</td>
<td>6 (18%)</td>
</tr>
<tr>
<td>3</td>
<td>3 (9%)</td>
</tr>
</tbody>
</table>
Table 6. Cross Individual Performance Segmented Based on Preference of social desirability in Experiment 2, Questions 3, 5, 6

<table>
<thead>
<tr>
<th>Social Desirability*</th>
<th>Hypothetical**</th>
<th>Incentive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>1.17 (0.97)</td>
<td>1.83 (2.16)</td>
</tr>
<tr>
<td>Medium</td>
<td>1.13 (0.94)</td>
<td>1.5 (0.97)</td>
</tr>
<tr>
<td>High</td>
<td>0.75*** (0.25)</td>
<td>1.71*** (0.9)</td>
</tr>
</tbody>
</table>

* The subjects are grouped into three segments: Low (those below 1 SD of the mean (scores 1-9)), Medium (those within 1 SD of the mean (scores 10-19)), and High (those above 1 SD of the mean (scores 20-30)).

** Average, with standard deviation in parenthesis.

*** The average is significantly higher in the incentive condition compared to the hypothetical condition, among those with high social desirability (p=0.028).
We have included the complete experimental instruction for hypothetical condition in experiment 2 here. Subjects in the incentive condition saw the same instruction except the relationship between their responses and the money was illustrated using examples on the blackboard, followed by questions and answers to ensure everybody understood the link. Experiment 1 is identical to experiment 2 except for question 3 (period 3), which is reproduced here together with the corresponding verdict form.

Experiment 1. Question 3 (Period 3):

3. **June 1, 1999-May 31, 2000**

During this time period, Genomik increased its manufacturing capacity to 3,000 injections, but sold only 1,000 injections. Genomik’s price was $1000 per injection for with a profit per sale of $900. CellPro increased its sales to 2,000 injections during this time period. The increase was due to an exclusive agreement CellPro made to provide a government clinical trial with 1,000 injections. All of CellPro’s sales collected a $900 profit per sale.

Genomik demonstrates that it had the capacity to produce 2,000 more injections than it actually made, and argues that it should obtain lost profits on all of the sales that CellPro made during this time period.

CellPro states that half of its sales (1,000 injections) were made only through an exclusive agreement that would not have been extended to Genomik. CellPro produces several doctors who participated in the study who state that they were particularly interested in CellPro’s product.

Experiment 1. Question 3 (Period 3) Verdict Form:

**Time Period 3 – June 1, 1999-May 31, 2000**

**Damages Calculation**

Total damages CellPro owes to Genomik for Time Period 3

**Jury Deliberation**

<table>
<thead>
<tr>
<th>Votes by Prior Jurors</th>
<th>Mark Your Vote</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Damages Amt.</strong></td>
<td><strong>Percentage of Jurors</strong></td>
</tr>
<tr>
<td>$1,800,000</td>
<td>74% (majority)</td>
</tr>
<tr>
<td>$990,000</td>
<td>19%</td>
</tr>
<tr>
<td>$180,000</td>
<td>7%</td>
</tr>
</tbody>
</table>
GENERAL INSTRUCTIONS FOR THE JURY STUDY

Thank you for participating in this study. The purpose of this study is to understand how an individual makes decisions as a juror. In this study, you will act as a juror in a civil (non-criminal) trial. The case you will be deciding involves two firms, CellPro and Genomik. CellPro has made a certain profit from selling a particular product. Genomik believes CellPro’s product sales infringe its patent and has sued CellPro. Genomik demands that CellPro compensate it for the profits it lost due to CellPro’s infringing sales. Your task as a juror is to determine how much money CellPro should pay Genomik. All responses will be anonymous.

You have been presented with a short summary of facts related to this case. The summary provides some background and points out what basic facts are in dispute. The summary also states the law related to the facts presented to you.

Additionally, you have received a Special Verdict Form for use in making your decision. The form is divided into separate time periods, each of which requires a decision. It is your duty to decide the disputed facts based on the evidence introduced in accordance with the principles of law laid down in the summary. Make your decision on the facts of the case only and set aside your personal preference.

After you have made a decision for each time period, you are asked to participate in a mock jury deliberation (discussion and voting) by deciding how you would vote in view of the answers of participants in prior studies.

How your answers will be used:

For the case included in this experiment, there is a solution upon which most legal experts would agree. Your answers to this case will be compared to this result and we will attempt to understand the underlying reason for differences and similarities.

GENOMIK V. CELLPRO

Summary of Facts & Law for Jury Deliberation

A. Background Facts:

Red blood cells (“RBCs”) are a component of human blood. They are essential for carrying oxygen from the lungs throughout the body. A deficiency in RBCs produces a serious condition called anemia, in which the sufferer cannot obtain enough oxygen from the air and is chronically exhausted. Several medical conditions may cause anemia such as cancer, cystic fibrosis, acute blood loss, pregnancy and even aging. It is estimated that 1,600,000 Americans suffer from anemia.

Problems associated with anemia can be reduced or eliminated if the human body can be coaxed to produce more RBCs. Erythropoietin is a protein that stimulates the production of RBCs. It is produced naturally in the human body, but normally only in very low quantities. Injecting erythropoietin into the blood increases its concentration in the body and causes the production of more RBCs. Supplemental erythropoietin can be an excellent treatment for anemia; however, it is very difficult to isolate in any substantial quantities from human or animal tissues.
In 1992, a method of producing large quantities of erythropoietin in a laboratory setting was discovered. Dr. Smith, a scientist working at a biotechnology company called Genomik, took the human gene (a piece of human DNA) that makes the erythropoietin protein, added a “promoter” to make it function, and inserted it into a hamster cell that can be grown in large quantities on special dishes in a lab (no live hamsters are involved). The cells then produce large quantities of erythropoietin, which can then be collected, purified and injected into patients with anemia. Genomik patented Dr. Smith’s method as well as “erythropoietin protein produced by a method using foreign DNA inserted into cells.” This is the ‘607 patent, which was issued in September 1993, and is wholly owned by Genomik.

CellPro is a biotechnology company that researches novel techniques of genetic manipulation. In January 1994, CellPro scientists discovered a very inexpensive way to make large quantities of proteins using human cells kept in a lab. The technique involved inserting a piece of viral DNA into the cells, which causes specific human genes to “turn on” at a very high level. CellPro realized that this technique could be used to produce erythropoietin. The company decided that, since the technique was so different from Genomik’s, it was unlikely to infringe the ‘607 patent. Moreover, CellPro was aware that Genomik was working on an erythropoietin product of its own, and CellPro wanted to get on the market first. CellPro began making erythropoietin and started selling it on June 1, 1997. It was an instant success and widely prescribed.

Immediately after obtaining the ‘607 patent, Genomik began building a new production facility to make patented erythropoietin obtained using Dr. Smith’s method. But because of construction problems and equipment delays, Genomik did not sell erythropoietin until June 1, 1998. Genomik increased its manufacturing capacity on June 1, 1999.

Genomik sued CellPro for patent infringement on June 1, 2003. Genomik claimed that, despite the fact that CellPro’s method was quite different than Genomik’s, CellPro’s erythropoietin was “erythropoietin protein produced by a method using foreign DNA inserted into cells” as claimed by the ‘607 patent. The first part of the trial has already taken place, and it has been determined that CellPro’s erythropoietin infringes Genomik’s patent.

All that is left is to determine what damages CellPro owes to Genomik. The parties disagree on the proper amount, and you must decide.

B. Applicable Law (Jury Instructions):

The United States Patent
A valid United States patent gives its owner the right to exclude others from, making using, offering to sell, or selling the patented invention within the United States, its territories, and its possessions.
**Patent Litigation**
A company is said to be infringing on claims of a patent when they, without permission from the patent owner, make, use, import, offer to sell or sell the patented invention, as defined by the claims, within the United States before the term of the patent expires. A patent owner who believes someone is infringing on his or her exclusive rights under a patent may bring a lawsuit like this to stop the alleged infringing acts and recover damages. Damages may be measured by the patent owner's lost profits caused by the infringement.

**Damages in General**
You must determine the amount of damages to be awarded to Genomik for the infringement. The amount of those damages must be adequate to compensate Genomik for the infringement. Your damage award should put Genomik in approximately the financial position it would have been in had the infringement not occurred; but in no event may the damage award be less than a reasonable royalty.

I will now explain how you should determine an appropriate damage award.

**Lost Profits**
To recover lost profits for some or all of the infringing sales, Genomik must show that if CellPro had not infringed, Genomik would have made those sales or a portion of them that CellPro made with the infringing product. You must determine what the customers who purchased the infringing product would have done if the infringement had not occurred.

Genomik need not present absolute proof that purchasers of the infringing product would have bought Genomik’s product instead. Genomik need not eliminate every possibility that purchasers might not have bought another product or might not have bought any comparable product at all. Rather, Genomik must show that there is a reasonable probability that Genomik would have made the sales that CellPro made of the infringing product.

You may infer that the Genomik has proven its lost profits if you find that Genomik has proven each of the following factors by a preponderance of the evidence:

1. that there was a demand for the patented product,
2. that there were no acceptable noninfringing substitutes,
3. that Genomik had the manufacturing and marketing capacity to make the infringing sales or a portion thereof actually made by CellPro, and
4. the amount of profit that Genomik would have made had CellPro not infringed.

**Reasonable Royalty**
The patent law specifically provides that the amount of damages that CellPro must pay Genomik for infringing Genomik’s patent may not be less than a reasonable royalty for the use that CellPro made of Genomik’s invention.
If you find that Genomik has not proved its claim for lost profits, or if you find that Genomik has proved its claim for lost profits for only a portion of the infringing sales, you must then determine what a reasonable royalty would be for that portion of the infringer's sales for which you have not awarded lost profit damages.

In this case, the parties have agreed that, if lost profits damages are not awarded for a particular infringement, a reasonable royalty for the use of the technology would be 10% of the defendant’s profits on the sales at issue.

Following are Two Hypothetical Examples for Calculating Damages

**Example 1 – When the patent holder could not make all infringer’s sales**
Imagine the following: Company B has infringed Company A’s patent by selling 100 widgets over a one-year time period. Company B makes $10 in profit for each widget. If Company B had not infringed by selling the 100 widgets, Company A would have had the capacity to make 50 of those 100 sales at $10 in profit per sale.

To determine Company A’s lost profits, one would determine the number of additional sales it could have made if Company B had not infringed, multiplied by A’s profits per sale:

\[ 50 \times \$10 = \$500 \text{ in lost profits} \]

Because only half of Company B’s infringing sales are accounted for by lost profits (50 sales out of 100), the rest MUST be assessed for a reasonable royalty. To determine the reasonably royalty on B’s remaining sales, one simply takes the number of sales multiplied by B’s profits per sale multiplied by the agreed upon royalty rate (10%):

\[ 50 \times \$10 \times 0.10 = \$50 \text{ in reasonable royalty payments} \]

To determine the total damages for that year, one simply adds the above

\[ \$500 + \$50 = \$550 \]

**Example 2 – When Facts Are in Dispute**
Imagine the following: Company B has infringed Company A’s patent by selling 100 widgets over a one-year time period. Company B makes $10 in profit for each widget. Although Company A had the capacity to make all 100 sales if Company B had not infringed, there is still a dispute about how many widgets Company A would have sold.

Company B claims that Company A did not have government regulatory approval for its product until the middle of the year, meaning it could only sell when Company B sold 50 widgets.
Company A claims that it would have requested government regulatory approval earlier. Company A also produces a noted expert who says that it would have received regulatory approval by the beginning of the year.

Because Company A has produced more evidence, one might conclude that it has shown a “reasonable probability” that it would have been able to make all of Company B’s infringing sales. The damages would then be entirely based on lost profits, calculated by simply multiplying the number of sales by Company A’s profits per sale:

\[ 100 \times 10 = 1000 \text{ in lost profits} \]

C. **Facts and Underlying Sales Information:**

CellPro made $1,440,000 in infringing sales of erythropoietin during the period of time in dispute. Genomik claims that it should receive from CellPro the profits Genomik would have made if almost every infringing sale of erythropoietin made by CellPro – 3,100 injections – was made by Genomik instead. CellPro disputes this for various reasons stated below. Use the facts below in view of the law discussed previously to resolve the question of damages in this case.

The dispute is easier to understand if broken up into three one-year time periods:

1. **June 1, 1997-May 31, 1998**
   It is undisputed that CellPro was the only manufacturer of erythropoietin on the market because Genomik could not operate its manufacturing facility. Its product sold for $1000 per injection, and CellPro sold 100 injections that year. CellPro’s profit per sale was $900.

   Genomik argues that lost profits would be appropriate on all of the sales that CellPro made during this time period because CellPro sold its products after Genomik was granted a patent. Genomik points out that patents are supposed to protect important inventions, and it would be unfair to allow a company to copy that invention and get away with paying only a royalty. Genomik contends that it should receive an amount equal to CellPro’s profits during this time period, approximately $90,000.

   CellPro argues that, because Genomik did not have the ability manufacture erythropoietin during this time period, it is not possible for Genomik to have suffered lost profits on CellPro’s product sales. CellPro argues that it owes only the royalty the parties agreed was reasonable on the profits made from its sales.

2. **June 1, 1998-May 31, 1999**
   Genomik finally arrived on the market with its own product. Both Genomik and CellPro charged $1000 per injection. Genomik sold 1,000 injections during this time period with a profit per sale of $900. CellPro also sold 1,000 injections during this time period, with a $900 profit per sale.
Genomik demonstrates that it had the capacity to produce 500 more injections than they actually produced (excess capacity).

CellPro notes that it sold 500 more injections than Genomik’s excess capacity.

3. June 1, 1999-May 31, 2000
   During this time period, Genomik increased its manufacturing capacity to 3,000 injections, but sold only 1,000 injections. Genomik’s price was $1000 per injection for with a profit per sale of $900. CellPro increased its sales to 2,000 injections during this time period. The increase was due to an agreement CellPro made to provide a government clinical trial with 1,000 injections. All of CellPro’s sales collected a $900 profit per sale.

Genomik demonstrates that it had the capacity to produce 2,000 more injections than it actually made, and argues that it should obtain lost profits on all of the sales that CellPro made during this time period.

CellPro states that half of its sales (1,000 injections) were made through the agreement to provide injections to the government clinical trial. CellPro notes that there is a remote chance that the government may not have been willing to substitute Genomik’s product for CellPro’s in the clinical trial.

4. June 1, 2000-May 31, 2001
   Both Genomik and CellPro were able to increase their regular sales during this period. Genomik sold 1,500 injections (and its capacity stayed the same). Genomik’s price was $1000 per injection for with a profit per sale of $900. CellPro sold 1,500 injections, with each injection selling for $1000 and a profit per sale of $900.

   CellPro argues that it could have sold more than 500 additional injections to government clinical trials, but did not due to concern regarding Genomik’s patent.

   Genomik argues that if CellPro had sold additional injections to government clinical trials, it would have done so at a lower profit.

5. June 1, 2001-May 31, 2002
   During this time period, Genomik increased its manufacturing capacity to 5,000 injections, but sold only 2,000 injections (an excess of 3000 injections). Genomik sold the injections for $1000 per injection with a $900 profit per sale. CellPro also sold 2,000 injections during this time period, also for $1000 per injection with a $900 profit per sale.

   CellPro claims that if its erythropoietin had not been on the market, half of the people who bought its product would have received a blood transfusion instead of switching to Genomik’s product. This transfusion is an experimental procedure that CellPro claims would have been rushed to the market if Genomik’s product was the only one available. In other words, according to CellPro, Genomik only lost profits on half of
the customers who purchased CellPro’s products. The other half should be subject to a reasonable royalty.

Genomik claims that the new blood transfusion methods could not have been used by this time period, because it was unpredictable. Genomik has produced a noted expert witness who states that such new methods are often not approved because they are unpredictable. The expert says that, although there is a possibility that the transfusion could have been available, it is reasonably probable that CellPro’s customers would not have viewed the transfusion as an alternative; they all would have used Genomik’s products if CellPro’s products had not been on the market.

Genomik’s product had become even more successful by the summer of 2002. In fact, during this time period, Genomik sold 3,000 injections, but still had the excess capacity to make 2,000 injections. Genomik was able to sell the injections for its normal $1000 per injection at a profit of $900 per injection. CellPro’s sales remained steady at 2,000 injections, also sold at $1000 per injection with a profit of $900 per injection.

Due to an administrative snafu, Genomik allowed its patent to become abandoned at the end of December 2002, by which time CellPro sold 1,200 of the injections it sold during the total period. From January to May, CellPro sold the remaining 800 injections. CellPro argues that, because Genomik had no patent rights during the last six months of this period, it should owe no damages at all.

Genomik notes that it never informed CellPro that it had abandoned the ‘607 patent rights, and CellPro did not even find out until this litigation.
**SPECIAL VERDICT FORM**

Please answer the questions relating to damages for each of the time periods in this case.

**Time Period 1 – June 1, 1997-May 31, 1998**

**Damages Calculation**

Total damages CellPro owes to Genomik for Time Period 1

**Jury Deliberation**

<table>
<thead>
<tr>
<th>Votes by Prior Jurors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Damages Amt.</strong></td>
</tr>
<tr>
<td>$9,000</td>
</tr>
<tr>
<td>$90,000</td>
</tr>
</tbody>
</table>

**Mark Your Vote**

- I agree with the majority
- I disagree with the majority
- I will agree with the majority even though my answer is different

**Time Period 2 – June 1, 1998-May 31, 1999**

**Damages Calculation**

Total damages CellPro owes to Genomik for Time Period 2

**Jury Deliberation**

<table>
<thead>
<tr>
<th>Votes by Prior Jurors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Damages Amt.</strong></td>
</tr>
<tr>
<td>$495,000</td>
</tr>
<tr>
<td>$900,000</td>
</tr>
<tr>
<td>$90,000</td>
</tr>
</tbody>
</table>

**Mark Your Vote**

- I agree with the majority
- I disagree with the majority
- I will agree with the majority even though my answer is different

**Time Period 3 – June 1, 1999-May 31, 2000**

**Damages Calculation**

Total damages CellPro owes to Genomik for Time Period 3

**Jury Deliberation**

<table>
<thead>
<tr>
<th>Votes by Prior Jurors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Damages Amt.</strong></td>
</tr>
<tr>
<td>$990,000</td>
</tr>
<tr>
<td>$1,800,000</td>
</tr>
<tr>
<td>$180,000</td>
</tr>
</tbody>
</table>

**Mark Your Vote**

- I agree with the majority
- I disagree with the majority
- I will agree with the majority even though my answer is different
**Time Period 4 – June 1, 2000-May 31, 2001**

**Damages Calculation**

Total damages CellPro owes to Genomik for Time Period 4

**Jury Deliberation**

**Votes by Prior Jurors**

<table>
<thead>
<tr>
<th>Damages Amt.</th>
<th>Percentage of Jurors</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1,350,000</td>
<td>73% (majority)</td>
</tr>
<tr>
<td>$945,000</td>
<td>16%</td>
</tr>
<tr>
<td>$1,800,000</td>
<td>11%</td>
</tr>
</tbody>
</table>

**Mark Your Vote**

___ I agree with the majority
___ I disagree with the majority
___ I will agree with the majority even though my answer is different

---

**Time Period 5 – June 1, 2001-May 31, 2002**

**Damages Calculation**

Total damages CellPro owes to Genomik for Time Period 5

**Jury Deliberation**

**Votes by Prior Jurors**

<table>
<thead>
<tr>
<th>Damages Amt.</th>
<th>Percentage of Jurors</th>
</tr>
</thead>
<tbody>
<tr>
<td>$990,000</td>
<td>80% (majority)</td>
</tr>
<tr>
<td>$1,800,000</td>
<td>15%</td>
</tr>
<tr>
<td>$2,700,000</td>
<td>5%</td>
</tr>
</tbody>
</table>

**Mark Your Vote**

___ I agree with the majority
___ I disagree with the majority
___ I will agree with the majority even though my answer is different

---

**Time Period 6 – June 1, 2002-May 31, 2003**

**Damages Calculation**

Total damages CellPro owes to Genomik for Time Period 6

**Jury Deliberation**

**Votes by Prior Jurors**

<table>
<thead>
<tr>
<th>Damages Amt.</th>
<th>Percentage of Jurors</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1,800,000</td>
<td>69% (majority)</td>
</tr>
<tr>
<td>$1,080,000</td>
<td>16%</td>
</tr>
<tr>
<td>$990,000</td>
<td>15%</td>
</tr>
</tbody>
</table>

**Mark Your Vote**

___ I agree with the majority
___ I disagree with the majority
___ I will agree with the majority even though my answer is different

**Total Damages** = ________________

(add all totals from above)
APPENDIX 2. SOCIAL DESIRABILITY SURVEY

Following is a thirty-three-item scale that used to measure subjects’ tendency to conform to socially desirable behaviors. The T or F in parenthesis after each question is the socially desirable answer and is added here for reference, they are not present in the survey that subjects saw.

Please indicate whether the following statement characterize you (Use T for true and F for false):

_________ Before voting I thoroughly investigate the qualifications of all the candidates. (T)
_________ I never hesitate to go out of my way to help someone in trouble. (T)
_________ It is sometimes hard for me to go on with my work if I am not encouraged. (F)
_________ I have never intensely disliked anyone. (T)
_________ On occasion I have had doubts about my ability to succeed in life. (F)
_________ I sometimes feel resentful when I don’t get my way. (F)
_________ I am always careful about my manner of dress. (T)
_________ My table manners at home are as good as when I eat out in a restaurant. (T)
_________ If I could get into a movie without paying and be sure I was not seen I would probably do it. (F)
_________ On a few occasions, I have given up doing something because I thought too little of my ability. (F)
_________ I like gossip at times. (F)
_________ There have been times when I felt like rebelling against people in authority even though I knew they were right. (F)
_________ No matter who I am talking to, I am always a good listener. (T)
_________ I can remember “playing sick” to get out of something. (F)
_________ There have been occasions when I took advantage of someone. (F)
_________ I am always willing to admit it when I have made a mistake. (T)
_________ I always try to practice what I preach. (T)
_________ I am always courteous, even to people who are disagreeable. (T)
_________ I sometimes try to get even rather than forgive and forget. (F)
_________ When I don’t know something I don’t at all mind admitting it. (T)
_________ I am always courteous, even to people who are disagreeable. (T)
_________ At times I have really insisted on having things my way. (F)
_________ There have been occasions when I felt like smashing things. (F)
_________ I would never think of letting someone else be punished for my wrong-doings. (T)
_________ I never resent being asked to return a favor. (T)
_________ I have never been irked when people expressed ideas very different from my own. (T)
_________ I never make a long trip without checking the safety of my car. (T)
_________ There have been times when I was quite jealous of the good fortune of others. (F)
_________ I have almost never felt the urge to tell someone off. (T)
_________ I am sometimes irritated by people who ask favors of me. (F)
_________ I have never felt that I was punished without cause. (T)
_________ I sometimes think when people have a misfortune they only got what they deserved. (F)
_________ I have never deliberated said something that hurt someone’s feelings. (T)
APPENDIX 3. RESULTS FOR QUESTIONS 1, 2, AND 4

We include the results for questions 1, 2, and 4 for both experiments in this appendix for completeness. Figures 3-1 and 3-2 have been included for easy comparison of the percentages of subjects reaching legally correct answers for all six questions in each experiment, respectively. Figures 3-3 and 3-4 correspond to Figure 1 in the paper, except the results are graphed for each experiment separately (note the correct answer for Question 3 is different in experiment 2 compared to experiment 1). The five tables included in this appendix correspond to Tables 2-6 in the paper, except they are for questions 1, 2, and 4.

With the exception of Question 1 in Experiment 2 (which also affected aggregate outcome, i.e., Figures 3-4 and 3-5), the subjects in incentive aligned condition exhibited the exact same patterns as observed for questions 3, 5, and 6. Although we hesitate to use this result to support our paper due to the confounding effect discussed in the paper, it nevertheless provides additional face validity.
Figure 3-1. Experiment 1 Percentage of Subjects Who Submitted Legally Correct Answers

Figure 3-2. Experiment 2 Percentage of Subjects Who Submitted Legally Correct Answers
Figure 3-3. Average Answers for Each Question in Experiment 1

Figure 3-4. Average Answers for Each Question in Experiment 2
Note the Correct Answer for Question 3 is different in Experiment 2 compared to Experiment 1.
### Table 3-1. Cross Question Results in Experiment 1

<table>
<thead>
<tr>
<th>Question 1</th>
<th>Question 2</th>
<th>Question 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothetical</td>
<td>Incentive</td>
<td>Hypothetical</td>
</tr>
<tr>
<td>Majority Opinion*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 (68%)</td>
<td>24 (86%)</td>
<td>12 (48%)</td>
</tr>
<tr>
<td>2nd Common Opinion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 (24%)</td>
<td>2 (7%)</td>
<td>6 (24%)</td>
</tr>
<tr>
<td>3rd Common Opinion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/A</td>
<td>N/A</td>
<td>2 (8%)</td>
</tr>
<tr>
<td>Other Opinions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 (8%)</td>
<td>2 (7%)</td>
<td>5 (20%)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 (100%)</td>
<td>28 (100%)</td>
<td>25 (100%)</td>
</tr>
</tbody>
</table>

* The majority opinion is the correct answer for all three questions.

### Table 3-2. Individual Performance in Experiment 1

<table>
<thead>
<tr>
<th>Number of Questions Correctly Answered</th>
<th>Questions 1, 2, 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hypothetical</td>
</tr>
<tr>
<td>0</td>
<td>2 (8%)</td>
</tr>
<tr>
<td>1</td>
<td>7 (28%)</td>
</tr>
<tr>
<td>2</td>
<td>10 (40%)</td>
</tr>
<tr>
<td>3</td>
<td>6 (24%)</td>
</tr>
</tbody>
</table>
### Table 3-3. Cross Question Results in Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>Question 1</th>
<th></th>
<th>Question 2</th>
<th></th>
<th>Question 4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hypothetical</td>
<td>Incentive</td>
<td>Hypothetical</td>
<td>Incentive</td>
<td>Hypothetical</td>
<td>Incentive</td>
</tr>
<tr>
<td>Majority Opinion*</td>
<td>29 (88%)</td>
<td>22 (69%)</td>
<td>17 (52%)</td>
<td>23 (72%)</td>
<td>15 (45%)</td>
<td>17 (53%)</td>
</tr>
<tr>
<td>2nd Common Opinion</td>
<td>3 (9%)</td>
<td>8 (25%)</td>
<td>8 (24%)</td>
<td>6 (19%)</td>
<td>10 (30%)</td>
<td>3 (9%)</td>
</tr>
<tr>
<td>3rd Common Opinion</td>
<td>N/A</td>
<td>N/A</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>1 (3%)</td>
<td>6 (19%)</td>
</tr>
<tr>
<td>Other Opinions</td>
<td>1 (3%)</td>
<td>2 (6%)</td>
<td>8 (24%)</td>
<td>3 (9%)</td>
<td>7 (21%)</td>
<td>6 (19%)</td>
</tr>
<tr>
<td>Total</td>
<td>33 (100%)</td>
<td>32 (100%)</td>
<td>33 (100%)</td>
<td>32 (100%)</td>
<td>33 (100%)</td>
<td>32 (100%)</td>
</tr>
</tbody>
</table>

* The majority opinion is the correct answer for all three questions.

### Table 3-4. Individual Performance in Experiment 2

<table>
<thead>
<tr>
<th>Number of Questions Correctly Answered</th>
<th>Questions 1, 2, 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hypothetical</td>
</tr>
<tr>
<td>0</td>
<td>2 (6%)</td>
</tr>
<tr>
<td>1</td>
<td>11 (33%)</td>
</tr>
<tr>
<td>2</td>
<td>10 (30%)</td>
</tr>
<tr>
<td>3</td>
<td>10 (30%)</td>
</tr>
</tbody>
</table>
Table 3-5. Cross Individual Performance Segmented Based on Preference of social desirability in Experiment 2, Questions 1, 2, 4

<table>
<thead>
<tr>
<th>Social Desirability*</th>
<th>Hypothetical**</th>
<th>Incentive**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>2 (0.8)</td>
<td>2.5 (0.3)</td>
</tr>
<tr>
<td>Medium</td>
<td>1.86 (0.94)</td>
<td>1.78 (0.3)</td>
</tr>
<tr>
<td>High</td>
<td>1.5 (1)</td>
<td>2 (1)</td>
</tr>
</tbody>
</table>

* The subjects are grouped into three segments, Low (those below 1 SD of the mean (scores 1-9)), Medium (those within 1 SD of the mean (scores 10-19)), and High (those above 1 SD of the mean (scores 20-30)).

** Average, with standard deviation in parenthesis. None of the differences is significant.