The Effects of Low versus High Cognitive Load on Judgments of Probability and Verdict: How Inducement of System 1 versus System 2 Processing Impacts the Wells Effect

A Senior Honors Thesis

Presented in Partial Fulfillment of the Requirements for graduation
with research distinction in Psychology in the undergraduate
colleges of The Ohio State University

by

Elizabeth K. Tripp
June 2010

Project Advisor: Hal R. Arkes, Ph.D., Department of Psychology

Running Head: Wells Effect

The Effects of Low versus High Cognitive Load on Judgments of Probability and Verdict: How Inducement of System 1 versus System 2 Processing Impacts the Wells Effect

Elizabeth K. Tripp

Department of Psychology

The Ohio State University

Columbus, OH 43210

Word Count: 6457

Correspondence address:

Elizabeth K. Tripp

64 E. 12th Ave, Apt. A

The Ohio State University

Columbus, OH 43210

Ph. (513) 324-2784

Fax (614) 688-3984

E-mail: tripp.53@gmail.com

Abstract

Humans have been shown to systematically ignore simple base-rate information. This tendency has been demonstrated in a variety of areas; one of the most well documented of these is the field of law. The Wells Effect, a phenomenon in which jurors display an inconsistency between their evaluations of probability and verdict, has perplexed researchers for the better part of two decades. Probability has to do with perceived likelihood of an event occurring, whereas verdict has to do with an actual evaluation of liability; most people concur about probability, whether given simple base rates or testimony from a human witness, but those in the latter group render affirmative verdicts far more often than those in the former. In this experiment, we tried to determine which system of processing, System 1 (immediate, "gut-reaction" mode of processing) or System 2 (measured, more calculated mode), presides over judgments of both probability and liability in a 2x2x2 (base rates versus human testimony, by 80% versus 40% probability, by high versus low cognitive load) between-subjects design. We examined this relationship based on participants' judgments after they read a case summary that employed either base rates or testimony from an expert. The results indicate support for the notion that verdict is a result of System 2 processing, though further clarification through future studies is needed.

Word count: 218 words

Introduction

Since Gary Wells first demonstrated the problem eighteen years ago, the Wells Effect has eluded all attempts to explain it. The phenomenon itself is one of the most puzzling problems in all of judgment and decision-making. The effect has to do with jurors' reluctance to convict a defendant in a liability case when the evidence is presented as "naked" statistical evidence alone (Wells, 1992). Though the statistical probability of liability is rated well above the threshold of culpability (defined as a 51% or greater likelihood that the defendant is liable for the offense in question), the majority of participants who are presented with probabilities in simple base rate form almost invariably decide not to find the defendant liable (Niedermeier, Kerr & Messe, 1999; Sykes & Johnson, 1999; Wells, 1992). The tendency to ignore base rate information, or at least to misuse it, was first noted in Kahneman and Tversky's influential study (1973). Since then, the problem has been identified in a variety of areas; in the case of the Wells Effect, subjects appear to correctly interpret the information, but use it in an irrational manner. That is, when given the probability in base rate form, almost all subjects will agree that the probability of the defendant's liability is well above the threshold of culpability, but only a small number of them will choose to convict, when theoretically all of them should. In terms of the implications for the legal system, the enormity of this problem lies in the obvious fact that this discrepancy probably has, and could continue to lead to many false negative verdicts. As Wells put it, "psychologically, there seems to be a difference between saying that there is an 80% chance that something is true and saying that something is true based on evidence that is 80% reliable" (Wells, 1992, p. 746). That is a very big problem indeed. If the factors underlying this issue were understood, steps could be taken to prevent or counteract it, in the interest of preserving the

way our justice system is intended to operate. The first attempt to explain the reason behind this grave error in judgment was Wells' own.

In Wells' studies (1992), participants read a case study in which Mrs. Prob was suing the Blue Bus Company for running over her dog. The plaintiff was walking her dog down a county road when a bus hit it; unfortunately, Mrs. Prob is colorblind and cannot identify the vehicle. A county transportation official testified that only two bus companies use that particular county road: the Blue Bus Company and the Grey Bus Company. Several conditions were used in this study. In the tire-tracks version (base rates only), a second county transportation official testified that 80% of all the buses were from the Blue Bus Company, and 80% of the traffic on the road comes from them. Accordingly, the Grey Bus Company owned 20% of all the buses and they were responsible for 20% of the traffic on that road. In the weigh-attendant version, a weigh attendant testified that 10 minutes prior to the accident at his weigh station, which is 10 minutes down the road from where the accident occurred, he wrote down in his official logbook that a blue bus was weighed. Another official testified that this weigh attendant's logbook notes were accurate 80% of the time and wrong 20% of the time, giving the same 80% probability that the Blue Bus Company is liable as in the tire tracks version. After reading the case summary, the mock jurors were asked to make an assessment of the probability that the Blue Bus Company was liable and to render a verdict. Wells found that people who heard the evidence from the weight attendant used the information far more appropriately than those in the base rate condition; they were far more likely to find the Blue Bus Company liable. Based on these findings, Wells suggested that the reluctance to convict came from jurors' use of "fact-to-evidence reasoning": for evidence to have a significant impact on verdict, the juror's own belief about the facts of the case must affect their belief about the evidence. That is, if your beliefs

about the facts of the case do not require you to disbelieve the evidence, then the evidence will not ultimately influence the way that you see the facts of the case (Wells, 1992). This theory, however, has been shown to be inaccurate (Niedermeier et al., 1999).

Several studies have been conducted in an attempt to parse out the flaws in human judgment and decision-making that lead to the Wells Effect. Niedermeier et al. (1999) had a small amount of success with one of their theories. They replicated Wells' original study, systematically disproved his fact-to-evidence theory, and ruled out their own human factor and general conclusion hypotheses. The first hypothesis (the human factor hypothesis) is that humans in general prefer to base their own beliefs on the beliefs of another person, an explanation that Wells had briefly mentioned in his own study (Niedermeier et al., 1999; Wells, 1992). The general conclusion hypothesis posited that perhaps humans prefer to simply have someone give them a "yes or no" opinion with regard to the liability of a party. They tested these hypotheses by adding a third condition in which a non-human party (a computer program) provided a conclusion about the liability of the Blue Bus Company. The results revealed that the participants were just as willing to accept the determination of the computer program as the testimony of a human witness, eliminating the human factor hypothesis. The general conclusion hypothesis was also rendered inapplicable; the results of the second experiment showed no significant difference in verdict judgments, whether the participants read testimony from witnesses who specifically stated their conclusion or not. One of their tested explanations did garner a small amount of support, though. The ease-of-simulation hypothesis postulated that the degree to which a situation allows a juror to imagine a scenario in which the defendant is innocent has the greatest correlation to their actual verdict. By including a condition in which the previously used tire-tracks version provided only a partial match instead of a complete

match, they examined how easy it was for jurors to imagine that the Grey Bus Company might be liable. This explanation had the most consistent support throughout the study, but direct measures of imagination only partially mediated verdict decisions (Niedermeier et al., 1999).

The trend of failed explanations continues with the third and final study of the Wells Effect. The main hypothesis, that surprise upon learning that the *less* probable situation occurred, is the best predictor of liability judgment did collect some positive evidence (Sykes & Johnson, 1999). They used two different "belief" conditions in addition to the base-rate and weigh-attendant conditions, one in which a witness testifies to the accuracy of the test and expresses only their belief in its precision. In the other version, the witness testifies to his belief that, based on the test results, the Blue Bus Company did in fact kill the dog. In the latter condition, surprise was indeed a mediating construct; in the condition where only probability statistics were offered, however, surprise had no mediating effects. This study once again offered no absolute evidence of a plausible explanation for the Wells Effect. Sykes and Johnson (1999) did, however, suggest that a possible key to understanding the phenomenon lies in the discrepancy between the two systems of human information processing.

Dual-Systems Theory: The Key to the Wells Effect?

The theory of dual systems seems to hold the answer to the Wells Effect, in that liability is likely guided by either System 1 or System 2 processing, and verdict is guided by the other. System 1 is thought of as an associative mode of processing; it is an automatic reaction to new information. System 2 is characterized by logic and analysis, and as such takes more time to "use" than System 1 (Sykes & Johnson, 1999). Granted, recently there have been some challenges made to the dual-system theory of information processing; one in particular asserts that a "binary opposition…cannot substitute for theories of cognitive processes" (Gigerenzer,

1999). The most recent anti-dual systems theory article goes so far as to break down every aspect of the theory in an attempt to show that it is unsatisfactory, and quite possibly does not even exist, but says that it may perhaps be used adequately as a theoretical framework (Keren & Schul, 2009). These arguments are very persuasive, and backed by a lot of analysis and perfectly logical reasoning. However, it seems fairly clear from the plethora of peer-reviewed articles on the subject, and its longevity in the field, that dual-systems theory is quite useful in describing the distinctions in modes of information processing. Without distinct data in favor of quashing the theory altogether, this fledgling researcher is reluctant to reject such a powerful theory outright. This intellectual clash notwithstanding, in the case of the Wells Effect, nearly every other possibility that is even vaguely plausible has been ruled out. Twenty-odd unpublished studies recently performed by Arkes, Shoots-Reinhard, and Mayes (2009) have done much to narrow down the best possibilities, and it seems that we may finally lay our finger on a definitive answer. Though it has been ten years since the last study on the effect was published, in this study we intended to both identify the direction of the causal arrow from probability to liability, and find what affects how humans perceive the direction of that arrow. In this experiment, we varied the probability statistics, the manner in which the probability information was delivered, and primed the subjects for either System 1 or System 2 processing in order to determine which system rules which part of the process of liability judgment.

The Current Study

We primed the subjects for System 1 or 2 processing by placing them under either a high or a low cognitive load. Theoretically, a high cognitive load would impair a subject's ability to use System 2 processing, since that load would occupy most of his or her mental faculties; it "knocks out" that System, so to speak. We then examined the data for variations in the

dependent variables (probability and verdict judgments) that fluctuated across the low-high load manipulation.

Possibility #1: Imposing a high cognitive load does not influence the probability judgment that the Blue Bus Company is responsible for the death of Mrs. Prob's dog; the probability estimates made by the weight attendant and tire tracks groups remain equal as they have always been in prior research. However the traditionally higher proportion of liability judgments in the weigh attendant group compared to the tire tracks group disappears; the proportion of liability verdicts would now be equal. We would then conclude that probability is normally processed by System 1 and verdict by System 2. "Knocking out" System 2 with a high cognitive load results in probability being the major input to verdict. With no System 2 processing available, verdict would have no other inputs to render the proportion of liability verdicts different in the weigh attendant and tire tracks groups.

Possibility #2: Imposing a high cognitive load does not influence judgments of liability; the proportion of liability judgments would be higher in the weigh attendant group than in the tire groups as they have always been in prior research. However the traditionally equal probability estimates made by the weigh attendant and tire tracks groups would now differ; the estimate made by the weigh attendant group would be higher. We would then conclude that verdict is normally processed by System 1 and probability by System 2. "Knocking out" System 2 with high cognitive load results in verdict being the major input to probability. With no System 2 processing available, probability would have no other inputs other than the normally divergent verdicts in the weigh attendant and tire tracks groups.

We hypothesized that possibility #1 was the correct one. This is first because the prior research indicates that probability estimates between the two groups are typically equal, which

suggests that this number is particularly salient, or easy to grasp. This would indicate that probability is under the influence of System 1, as it is an associative mode of processing. After all, the subjects are given the probability outright in the case summary. Where the groups differ greatly is in their judgments of liability. The subjects are offered no definitive answer as to whether the Blue Bus Company is liable, other than their knowledge of preponderance of evidence and the probabilities given in the testimony. Unlike their probability estimate, which they could simply take directly out of the evidence provided with very little effort, a bit of mental power is required in order to determine whether or not the Blue Bus Company should in fact be found liable. This mental effort would lie in System 2's jurisdiction, which is why we chose to assume that judgments of verdict are derived from this mode of processing.

Methods

Participants. Participants were 152 introductory psychology students who received class credit for taking part in the study.

Design and procedure. Each participant was randomly assigned to one of sixteen conditions. They were set up on a computer fitted with MediaLab, the program on which the study was run. Then all subjects, whether in a low or high cognitive load condition, were given a short definition of the term "preponderance of evidence", and then immediately given a simple quiz to ensure that they understood what "preponderance of evidence" means. The importance of this was explained in the introduction:

> In this experiment, you will be asked to assume the role of juror in a civil lawsuit. If you think that the preponderance of the evidence—that is, more than 50% of the evidence--favors one of the two sides in this case, you should rule in favor of that side. You will be

provided with information about the case, and you will be asked to make a decision in this case just as jurors are typically asked to do.

Immediately after this paragraph, the quiz was administered. The quiz was worded as follows:

Before we begin, we need to be sure you know what "preponderance of the evidence" means, because if the preponderance of the evidence favors one side in this dispute, you should rule in favor of that side.

"Preponderance of the evidence" means: (choose one)

1. "100% of the evidence favors one side;"
2. More than 95% of the evidence favors one side;"
3. More than 75% of the evidence favors one side;"
4. More than 50% of the evidence favors one side;"
5. More than 25% of the evidence favors one side;"
6. More than 10% of the evidence favors one side;"
7. "Don't know."

Half of the participants were then placed under a cognitive load in order to undermine their System 2 processing. A variation on Shiv and Fedorikhin's (1999) method was used: the subjects were provided with a two-color (low load condition) or seven-color (high load condition) sequence to remember throughout the experiment. Participants were then asked to read a civil trial summary. Immediately after this, the sequence of colors they memorized before they read the case summary was presented on the screen for a period of two seconds. Then, the participants were asked to reach a verdict, and estimate the probability that the defendant was liable. They then reported the colors they memorized at the beginning of the experiment. After these questions, they were given a five-question memory quiz to check that they had been paying

attention to the data provided in the case summary. All data were automatically recorded on MediaLab data collection files.

The first version of the case is a replication of Wells' 1992 tire-tracks version. The case reads:

> Mrs. Prob is suing the Blue Bus Company for having caused the death of her dog. At the trial, the following evidence was given:
>
> > Mrs. Prob testified that she was walking her dog along County Road #37 when she heard a large vehicle behind her. She turned around and saw a bus serving recklessly down the road. She jumped out of the way but the bus swerved and hit her dog. Unfortunately, Mrs. Prob is color blind and thus does not know the color of the bus.
> >
> > A county transportation official, Mr. Jones, took the stand, was sworn in as a witness, and testified that there are only two bus companies in the county: the Blue Bus Company and the Grey Bus Company. Each company uses the road to run empty buses back to their stations after dropping off passengers. Therefore, one of the two bus companies had to be responsible for the death of Mrs. Prob's dog.
> >
> > A second county transportation official, Mr. Smith, who was the transportation department's chief investigator, took the stand, was sworn in as a witness, and reported that he examined the dead dog and took prints off the tire tracks. These prints were then transferred onto paper and compared with all 10 of the 10 buses owned by the Blue Bus Company and all 10 of the 10 buses owned

by the Grey Bus Company. The tracks matched 80% of the Blue Bus Company's buses and 20% of the Grey Bus Company's buses.

Mrs. Prob's attorney argued that the jury must find the Blue Bus Company liable for damages because in all likelihood, it was the Blue Bus Company that killed Mrs. Prob's dog.

The second version of the case is almost identical to the above; only the numbers in the third paragraph have been changed. This version reads:

A second county transportation official, Mr. Smith, who was the transportation department's chief investigator, took the stand, was sworn in as a witness, and reported that he examined the dead dog and took prints of the tire tracks. These prints were then transferred onto paper and compared with all 10 of the 10 buses owned by the Blue Bus Company and all 10 of the 10 buses owned by the Grey Bus Company. The tracks matched 40% of the Blue Bus Company's buses and 60% of the Grey Bus Company's buses.

The third version of the case is a replication of Wells' 1992 weight-attendant version. The third version reads:

Mrs. Prob is suing the Blue Bus Company for having caused the death of her dog. At the trial, the following evidence was given:

Mrs. Prob testified that she was walking her dog along County Road #37 when she heard a large vehicle behind her. She turned around and saw a bus serving recklessly down the road. She jumped out of the way but the bus swerved and hit her dog. The incident occurred at 11:40AM. Unfortunately, Mrs. Prob is color blind and thus does not know the color of the bus.

A county transportation official, Mr. Jones, took the stand, was sworn in as a witness, and testified that there are only two bus companies in the county: the Blue Bus Company and the Grey Bus Company. Each company uses the road to run empty buses back to their stations after dropping off passengers. Therefore, one of the two bus companies had to be responsible for the death of Mrs. Prob's dog.

A second county transportation official took the stand, was sworn in as a witness, and reported that he was on duty as the weigh attendant the day of the bus-dog accident. He explained that all vehicles with more than two axles (such as buses) must enter a weigh station and drive slowly over a set of scales. As they drive over, the weigh attendant notes their weight and jots down a two-word description of the vehicle in the log book. In the weigh attendant's log book for the day in question, he had entered "blue bus, 11:30AM" along with a weight. The dog was hit at 11:40AM and the distance from the weigh station to the point where Mrs. Prob's dog was killed is about a 10 minute drive.

The defense attorney for the Blue Bus Company recalled the weight station attendant and entered evidence showing that his previous log book entries were correct only 80% of the time and wrong 20% of the time. This was proven by records seized after the alleged incident. These records showed that 20% of the time in which a blue bus was weighed the attendant wrote down "grey bus" and 20% of the time that a grey bus was weighed the attendant wrote down "blue bus".

> Mrs. Prob's attorney argued that the jury must find the Blue Bus Company liable for damages because in all likelihood, it was the Blue Bus Company that killed Mrs. Prob's dog.

The fourth version is almost identical to the third, only with the weigh attendant's reliability changed to 40% of the time right, 60% of the time wrong. In this case, the fourth paragraph reads:

> The defense attorney for the Blue Bus Company recalled the weight station attendant and entered evidence showing that his previous log book entries were correct only 40% of the time and wrong 60% of the time. This was proven by records seized after the alleged incident. These records showed that 60% of the time in which a blue bus was weighed the attendant wrote down "grey bus" and 60% of the time that a grey bus was weighed the attendant wrote down "blue bus".

After reading one of these four versions, participants individually filled out dependent measure sheets. First, they estimated the probability in percentage that the Blue Bus Company was responsible for the dog's death. Next they indicated by filling in a box labeled *yes* or *no* whether they would rule against the Blue Bus Company and make them pay damages to Mrs. Prob. These two questions' order was reversed for other subject to eliminate the possibility of order effects. After this, subjects had to recall the two- or seven-color sequence they were asked to memorize, and then take the five-question memory quiz, which was written as follows:

1. Whose dog was killed by a bus?
    a. Mrs. Prob
    b. Mrs. Pibs

c. Mr. Pro

d. Mrs. Prim

e. Mrs. Rob

2. What Bus Company was being sued?

a. Green

b. Black

c. Red

d. Blue

e. Grey

3. What color were the other company's buses?

a. Green

b. Blue

c. Black

d. Grey

e. Red

4. Why couldn't the color of the bus be exactly identified?

a. The owner of the dog was colorblind

b. It was dark

c. The owner of the dog did not witness the accident

d. The two colors were too close to one another to be distinguished

e. The owner of the dog was not paying attention

5. What was the testimony of the expert witnesses?

a. Mrs. Prob was not an accurate eyewitness.

b. A picture of a bus was taken near the accident and that it was more likely to be a blue than grey bus.

c. A picture of a bus was taken while the accident was occurring and that it was more likely to be a blue bus than a grey bus.

d. Tire tracks were taken from the dog and were more likely to be from a blue bus than a grey bus.

Data Analysis

For the analysis of the probability dependent variable, we used SPSS to perform a 2 (cognitive load: high/low) x 2 (scenario version: tire tracks/weigh attendant) x 2 (accuracy level: 80%/40%) analysis of variance (ANOVA). For the analysis of the verdict dependent variable, we used SPSS to perform a logistic regression with the same three predictors--cognitive load, scenario version, and accuracy level. Only those participants who passed the preponderance of evidence quiz were allowed to move forward with the experiment. Thirty-eight subjects out of 152 were excluded for this reason. Then the data from twenty-two participants who failed the manipulation check (those who made more than one mistake during recall) were thrown out, which left a total of ninety-two sets of data to use.

Results

Probability.

A three-way ANOVA revealed a significant effect of accuracy level. Subjects in the 80/20 groups assigned an average probability of 73.51; those in the 40/60 group assigned an average probability of 44.93. This difference was significant, $F(1,84) = 142.75$, $p < .001$, $\eta^2 = .630$. The subjects were obviously highly responsive to the information provided concerning the accuracy of the tire-matching and weigh attendant testimony.

This main effect was qualified by an accuracy level by group interaction: ($F(1,84) = 4.383$, $p = .039$, $\eta^2 = .050$). When the tire tracks or weigh attendant manifested a low level of accuracy or probability of a match (40%), the estimated probabilities that the Blue Bus Company was responsible for the accident were nearly identical (44.0 for the tire tracks group and 45.7 for the weigh attendant group). However in the 80/20 group, the estimated probabilities that the Blue Bus Company was responsible for the accident were more divergent (77.3 for the tire tracks group and 69.1 for the weigh attendant group).

There was also a three-way interaction: $F(1,84) = 4.22$, $p = .043$, $\eta^2 = .048$. Irrespective of cognitive load, in the 80/20 group the probability assigned by the tire tracks subjects exceeded that assigned by the weigh attendant subjects. This was also the case among the 40/60 subjects with high cognitive load. However with low cognitive load the 40/60 subjects assigned a higher probability estimate to the weigh attendant subjects. See Table 1 in Appendix A (page 23 of this paper) for a full summary of probability estimates.

Verdict.

Using a logistic regression, the results reveal that not surprisingly, accuracy level had a strong effect on verdict judgments ($b = -1.34$, $p < .007$). More accurate weigh attendants or tire track evidence resulted in much higher liability judgments than less accurate evidence (47% vs. 20%). Group had a significant effect ($b = -0.988$, $p < .042$). This effect of group is a replication of the original Wells Effect in that 44% of the weigh attendant subjects voted to find the Blue Bus Company liable, but only 24% of the tire tracks subjects did so. See Table 2 in Appendix A (page 23 of this paper) for a complete summary of liability verdicts.

Discussion

The main objective of this study was to determine which information processing system is responsible for judgments of verdict, and which is responsible for judgments of probability. If verdict is guided by System 2, then in the high cognitive load condition, participants' decisions on verdict should align with their judgment of probability. If the weigh attendant and tire tracks subjects differ in the probabilities they assign, then they must also differ in the same manner in the verdicts they assign. Under high cognitive load there would be no available processing capacity for System 2 to consider other factors besides probability that might make the pattern of verdicts differ from the pattern of probability results. On the other hand, if verdict is the responsibility of System 1 processing, then high cognitive load should have the opposite effect, and make probability judgments fall in line with verdict.

The theory that verdict is guided by System 2 seems to have significant but not complete support. According to this notion, because the probabilities in both the weigh attendant and tire tracks conditions have traditionally been equal, the rate of liable verdicts should also be equal in the high cognitive load conditions. This is due to the fact that under high cognitive load, there would be no System 2 processing available to incorporate considerations other than the probabilities, which have always been the same in the tire tracks and weigh attendant groups. In the 80% probability condition, this holds true; probability judgments show no significant difference (76% tire tracks vs. 71% weigh attendant), though there is a gap in verdict. Participants in the weigh attendant condition convicted the Blue Bus Company 72.7% of the time, where those in the tire tracks version convicted at 45.5% of the time. This result is not statistically significant, however. In the 40% probability condition, the estimations of probability are the same across groups as well (46.4% weigh attendant vs. 41.6% tire tracks), and

the estimations of verdict are equal, as well (78% weigh attendant vs. 73% tire tracks). Thus under high cognitive load, verdict does generally mimic probability, which gives support to the theory that judgments of verdict are due to System 2 processing.

On the low cognitive load side of this theory, in the 80% probability condition, the results should replicate the usual Wells Effect: the judgments of probability should be equal across conditions, while participants in the weigh attendant version should deem the Blue Bus Company liable more than those in the tire tracks condition. The probability judgments are indeed similar (67% weigh attendant vs. 78% tire tracks), but the verdicts do not differ significantly (46% weigh attendant vs. 29% tire tracks). In the 40% probability condition, probability judgments are similar as well (50% weigh attendant vs. 42% tire tracks). The verdict predictions are complicated in this situation. The weigh attendant group subjects deem the probability to be right at the "conviction" threshold of 50%! The tire tracks subjects are below the threshold on average. Thus the small 8-point difference between the two groups in their probability estimates is at the most crucial point on the scale. As a result the weigh attendant group's proportion liability verdicts is higher (33%) than that of the tire tracks group (0%). This difference was significant by a Fisher Exact Test ($p < .039$).

The theory that verdict is guided by System 1 processing is disconfirmed by a main result of this study. Under high cognitive load in the 80% probability condition, the weigh attendant subjects rendered a liable verdict much more often than did the tire track group's subjects (73% vs. 46%). (As mentioned before, due to the small *N* this difference is not significant, however.) If probability estimates were fueled by System 2 processing, then the probability estimates would also show this same pattern, because under high cognitive load there would be no System 2 processing available to allow any considerations other than verdict to influence probability.

However, the probability estimate of the tire tracks group was actually slightly *higher* than that of the weigh attendant group (76% vs. 71%), the opposite of the verdict results. "Knocking out" System 2 did not make probability estimates track the verdict. This is contrary to what we would expect if probability were processed by System 2. With no System 2 processing available, the probability results should have had the same pattern as the liability results if probability required System 2.

Now we'll consider the 40/60 condition under high cognitive load. The proportion of liability verdicts is very similar in the two groups (27% for tire tracks and 22% for weigh attendant), as are the probability estimates (46.4% for tire tracks vs. 41.6% for weigh attendant). So knocking out System 2 processing did result in the probability estimates duplicating the liability judgments. This portion of the results is consistent with the hypothesis that verdict is due to System 1 processing and probability is due to System 2.

Limitations.

Though the data garner good support for our hypothesis, more studies are needed in order to confirm this. The modest statistical power shows that we need to increase the sample size for our next study, to ensure that every condition has a large enough *N* to be confident that the analyses are relevant to the population and correct. Also, though there was only one statistically significant difference in judgments of verdict between the groups where, according to our hypothesis, the proportions should have been equal, there were some noticeable gaps in the percentages (i.e., in the 80% probability – high cognitive load condition, the difference in affirmative verdict judgments is 72.7% in weigh attendant to 45.5% in tire tracks). A larger study would help to either confirm that there is no statistical difference, or show that there is in fact a significant difference in the proportions of judgments of verdict. There were also an

appreciable number of subjects who failed the preponderance of evidence quiz, something that should have been nearly impossible to fail. Measures may need to be taken to make sure that the definition given to them is even clearer or more salient, so that we can be sure they understand it adequately and increase the retention rate of the study.

Implications.

The implications of this study, in combination with future studies to strengthen the data, could be very far-reaching, and have the potential to have quite a bit of practical use for the legal system. It would be rather redundant to mention once more how large an error in human decision-making this is; obviously, the Wells Effect has the potential to have enormous impact on trial outcomes, which is a bit disconcerting. If we could finally understand what exactly it is that makes jurors make this mistake, a plan could be devised to correct their decision-making process. Judges, lawyers, and jurors themselves could take steps to make sure that nothing but the facts of the case are influencing their verdict decision. Future confirmation of these results pending, we could both gain insight into a major cognitive process, and fix a problem that probably has and could continue to result in unfair verdicts in civil trial cases.

If future studies confirm these results, then further research could focus on devising ways to accurately prime jurors to use better judgment and ignore the pesky human tendencies that cause them to make poor decisions. They could perhaps look into priming jurors before they hear the case, which might stop them from being swayed by any factors irrelevant to the facts of the case. Another possibility is performing some sort of debriefing after the trial has concluded, in an effort to encourage jurors to focus solely on the bare facts, and forget anything that may have purely emotive lure. In any case, there may be multiple ways to effectively remove the bias

the Wells Effect incurs in jurors; if we can make the first step and explain it away, there could be a lot of room made to fix the problem.

Conclusion

There is a healthy amount of support in the data presented to suggest that judgments of verdict are guided by System 2 processing. Further, larger studies are required to confirm this suggestion, but this study may be the beginning of the answer to the Wells Effect. Probabilistic evidence is and will continue to be a common source of information in the legal system. Since we have both replicated the Wells Effect in this study and shown that cognitive load influences the ways in which potential jurors perceive evidence, we have furthered the body of data that proves that measures taken to combat this vulnerability may help lead to a more just, accurate judicial system.

Acknowledgments

I would like to thank Dr. Hal Arkes for his guidance and eternal patience throughout this project, and Brittany Shoots-Reinhard for her advice and help with SPSS coding and MediaLab programming.

Appendix A: Tables

Table 1: Summary of Probability Estimates

Accuracy Level	Cognitive Load	Group	Average Estimate of Probability[c]
40% probability	low	WA[a]	**49.83**
		TT[b]	**41.54**
	high	WA	**41.56**
		TT	**46.36**
80% probability	low	WA	**66.82**
		TT	**78.21**
	high	WA	**71.36**
		TT	**76.36**

Notes: a = Weigh Attendant; b = Tire Tracks
c = Estimations of the probability that the Blue Bus Company is liable

Table 2: Summary of Liability Judgments

Accuracy Level	Cognitive Load	Group	Percent Liable Verdict[c]
40% probability	low	WA[a]	**33.3**
		TT[b]	**0.0**
	high	WA	**22.2**
		TT	**27.3**
80% probability	low	WA	**45.5**
		TT	**28.6**
	high	WA	**72.7**
		TT	**45.5**

Notes: a = Weigh Attendant; b = Tire Tracks
c = The percentage of subjects in each condition who determined that the Blue Bus Company should be found liable

References

Gigerenzer, G. (1998). Surrogates for Theories. *Theory and Psychology. 8(2).* 195-204.

Kahneman, D., & Tversky, A. (1973). On the Psychology of Prediction. *Psychological Review. 80(4).* 237-251.

Keren, G., & Schul, Y. (2009). Two Is Not Always Better Than One: A Critical Evaluation of Two-System Theories. *Perspectives on Psychological Science. 4(6).* 533-550.

Niedermeier, K.E., Kerr, N.L., & Messe, L.A. (1999). Jurors' Use of Naked Statistical Evidence: Exploring Bases and Implications of the Wells Effect. *Journal of Personality and Social Psychology. 76(4).* 533-542.

Shiv, B., & Fedorikhin, A. (2002). Spontaneous versus Controlled Influences of Stimulus-Based Affect on Choice Behavior. *Organizational Behavior and Human Decision Processes. 87(2).* 342-370.

Sykes, D.L., & Johnson, J.T. (1999). Probabilistic Evidence Versus the Representation of an Event: The Curious Case of Mrs. Prob's Dog. *Basic and Applied Social Psychology. 21(3).* 199-212.

Wells, G.L. (1992). Naked Statistical Evidence of Liability: Is Subjective Probability Enough? *Journal of Personality and Social Psychology. 62(5).* 739-752.