Inductive reasoning in the courtroom: Judging guilt based on uncertain evidence

Ann M. Martin (Ann.Martin@unsw.edu.au)
School of Psychology, University of New South Wales
Sydney, 2052, Australia

Brett K. Hayes (B.Hayes@unsw.edu.au)
School of Psychology, University of New South Wales
Sydney, 2052, Australia

Abstract

Most legal systems require jurors to consider all the evidence presented at trial. Hence when there is uncertainty over aspects of evidence this should be factored into juror judgments. Two experiments examined how mock jurors used uncertain information in their ratings of defendant guilt and final verdicts. Participants read scenarios where an eyewitness expressed uncertainty about the identity of a critical piece of evidence (e.g. the object a defendant was holding could have been a knife or a mobile phone). The respective probability of these alternatives was varied, as was their association with the alleged crime. When the probability of the alternatives was varied between subjects (Experiment 1) there was only weak evidence that jurors considered both alternatives. When probability was varied within-subjects (Experiment 2), jurors did consider both alternatives in their guilt judgments. The implications for theories of reasoning with uncertain information and forensic practice are discussed.

Keywords: Inductive reasoning; Probabilistic reasoning; Category-based induction, Forensic judgment

Introduction

The process of deciding on the guilt or innocence of a defendant in a criminal trial could be seen as form of complex inductive inference (Lagnado, 2011). Induction involves drawing probabilistic inferences from given information. When the category membership of an object is known with certainty (e.g., a man at a robbery crime scene was carrying a *knife*) the process of inductive inference is relatively straightforward (e.g., the knife was likely being used as a weapon in the robbery).

However, the evidence presented in criminal cases is typically complex and fraught with uncertainty (e.g., an eyewitness may not be certain about the identity of the object that the defendant was holding). In such cases induction involves the consideration of multiple possible object categories that may have different implications for judgments about the defendant's guilt. For example, if a defendant was thought to be carrying a knife then a juror may judge the defendant as likely to be guilty of committing a crime. But if there is some chance that the object in the defendant's hand was something less incriminating (e.g., a metallic-colored mobile phone) then this may reduce belief in the defendant's guilt.

So just how do jurors respond to such uncertain alternatives when making inferences about guilt or innocence? The legal answer to this question is straightforward and prescriptive. Most criminal jurisdictions specify that jurors should consider *all* evidence presented at trial when determining defendant culpability (e.g., Attorney General's Department of New South Wales, 2007). Laboratory studies of inductive reasoning with uncertain categories however, suggest a more complex answer (see Hayes, Heit & Swendsen, 2010 for a review).

Bayesian approaches to inductive reasoning, such as Anderson's (1991) Rational model, generally agree with the legal ideal, suggesting that reasoners incorporate information about all category alternatives when making inferences (hereafter referred to as *multiple-category reasoning*). To illustrate, let us assume that an eyewitness believes that the probability that the object the defendant was holding was a knife is 0.7 (which we will refer to as the *primary category)*, with a 0.3 probability that the object was instead a metallic mobile phone (*the secondary category*). Further assume that these alternatives are associated with different conditional probabilities of guilt. If the person was holding a knife then the probability of them being guilty is high (e.g., p(guilt| knife) = 0.9), but if they were holding a mobile phone the probability of guilt will be much lower (e.g., p(guilt| phone) = 0.2). Applying Bayes' theorem, the Rational model combines the probabilities from the primary and secondary categories to give an estimate of the probability that the defendant was guilty, given they were seen with a metallic object in their hand (p(guilt| metallic object) = (0.7*0.9) + (0.3*0.2) = 0.69). Note that this probability estimate is considerably lower than would be the case if the uncertainty over object identity was ignored. If the juror assumed that the object in the defendant's hand *was* a knife then the guilt estimate would be 0.9.

Unfortunately (from a legal point of view), empirical studies have so far found little evidence of multiple-category reasoning. Malt, Ross and Murphy (1995) for example, presented vignettes in which the category membership of a target character was uncertain, and asked participants to make various inferences about the targets. For each scenario two possible category identities were suggested, a more likely primary category and a less likely but plausible secondary category. The primary category was held constant across conditions (e.g., the vignette always made it clear that the target character was most likely a realtor). The secondary category however was varied; in one condition the target was most likely a *cable repairman*, in

another it was most likely a *burglar*. These alternatives have different implications for inferences such as "how likely is it that the man will pay attention to the sturdiness of the doors on the house?" Such behavior seems more likely if the target was a burglar than if they were a realtor or cable repairman. If people do consider multiple categories in induction, then their inferences should differ across the conditions with different secondary categories. Malt et al. (1995) however, found that participants tended to ignore the secondary category when making inductive inferences. Predictions were predominantly based on consideration of the primary category alone (also see Ross & Murphy, 1996).

Such "single-category reasoning" seems pervasive in non-forensic domains, having been demonstrated with a wide variety of artificial and natural categories (see Murphy & Ross, 2007, 2010 for reviews). Murphy and Ross (2007) suggest that single-category reasoning can be viewed as a cognitive heuristic that reduces the complexity of deriving inductive predictions from multiple uncertain alternatives.

The pervasive nature of single-category reasoning in previous studies leads to a negative prognosis for forensic reasoning, suggesting that jurors are also likely to use the single-category heuristic. Such a prediction is consistent with reports of juror "satisficing" where the juror focuses on aspects of trial evidence that are consistent with a single coherent story, ignoring contradictory evidence (Kuhn, Weinstock, & Flaton, 1994; Pennington & Hastie, 1992).

On the other hand, multiple-category reasoning may be more common in forensic situations because of their particular motivational demands. Motivational factors play an important role in determining the depth and complexity of reasoning (Gilovich & Griffin, 2010; Kunda, 1990). Forensic judgments like guilt or innocence are widely recognised as having profound consequences for a defendant and for the wider community (Bornstein & Greene, 2011). Hence jurors may be more likely to consider uncertain alternatives when making highly consequential judgments. Some support for this prediction comes from Hayes and Newell (2009) who found that multiple-category reasoning was more likely when neglect of the secondary category could lead to a highly negative outcome (e.g. when the primary category was a common but easily treatable disease and the secondary category was a rare but potentially terminal disease).

The main aim of the current studies therefore was to examine whether mock jurors would show multiple-category reasoning in cases where there was uncertainty about the identity of forensically relevant evidence in a criminal trial.

Experiment 1

Experiments 1 and 2 were patterned after those of Malt et al. (1995) and Ross and Murphy (1996), using a design where probability estimates of guilt were compared across conditions in which the primary category was held constant and the secondary category varied (see Table 1). Participants in all conditions were shown written vignettes which described criminal cases where there was uncertainty about the identity of a critical piece of evidence. An eyewitness testified that they observed the defendant carrying an object, which they believed to be a particular item (the primary category). However, they acknowledged that there was a lower probability the object may have been something else (the secondary category). In two Primary Related conditions (Comparison 1 in Table 1), the primary category consistently implicated the defendant in the crime (e.g., in a robbery the primary category was "knife"). However, the less likely secondary category varied such that it indicated that the defendant was guilty in one condition only (e.g., a stolen watch in one condition and a mobile phone in the other).

After reading the vignettes participants judged which object they thought the defendant was actually carrying (object categorization) and made inferences about the defendant's guilt. Whether participants used single- or multiple-category reasoning to arrive at these inferences could be determined by comparing guilt estimates in the conditions where the secondary categories were varied (see Comparisons 1 & 2 in Table 1).

Table 1: Summary of the experimental design (with examples of critical object alternatives from the Robbery vignette)

	Secondary Category	
	Related (e.g., stolen watch)	**Unrelated** (e.g., mobile phone)
Comparison 1 **Primary Category Related** (e.g., knife)	Primary = *knife* Secondary = *stolen watch*	Primary = *knife* Secondary = *mobile phone*
Comparison 2 **Primary Category Unrelated** (e.g., keys)	Primary = *keys* Secondary = *stolen watch*	Primary = *keys* Secondary = *mobile phone*

A subsidiary aim of Experiment 1 was to re-examine the Ross and Murphy (1996) finding that multiple-category reasoning is more likely when the secondary category is closely linked to the prediction being made and the primary category is not. In a variant on the realtor/cable guy/burglar task, Ross and Murphy (1996) asked participants to make predictions that were more strongly associated with the secondary category of burglar but not with the primary category of realtor (e.g., the prediction "how likely is it that the man will try to find out if the householder keeps her windows locked?"). Participants were more likely to consider both categories when making such predictions.

Hence in the current study, we also ran a Primary unrelated condition where the primary category was not strongly associated with culpability and the association of the secondary category with culpability was varied. In the Primary unrelated, Secondary related condition the

secondary category was associated with guilt (e.g., a stolen watch), whereas in the Primary unrelated, Secondary unrelated condition, neither category was associated with the crime. Consideration of the secondary category should lead to higher guilt estimates in the Primary unrelated, Secondary related condition that in the Primary unrelated, Secondary unrelated group.

Method

Participants Ninety-eight undergraduate students participated for course credit. The majority were female (n = 72), and the mean age was 19.95 years (SD = 1.84). All were Australian citizens aged 18 years or older, in accordance with Australian juror selection criteria.

Design and Materials The experiment followed a 2 × 2 factorial design with the first factor being whether the *primary* category related to the crime and the second factor being whether the *secondary* category related to the crime (see Table 1). In Experiment 1 both factors were manipulated between subjects, with approximately equal numbers allocated to each experimental condition.

These factors were operationalized using written vignettes presented as brief criminal trial summaries. Each vignette was approximately 290 words in length and described a case in which an eyewitness reported seeing the defendant at the crime scene carrying an object whose identity was critical for evaluating defendant guilt (see Appendix for an example). Two category possibilities were provided for this object: a primary category, which was described as the most likely identity of the critical object (with an explicit likelihood of 70%), and a secondary category, which was described as having "a small chance" of being the identity of the object. Two vignettes with this structure were developed. One described a criminal trial for assault and robbery, and the other described a trial for arson.

Our assumptions about the relatedness of the various critical objects with the crime were confirmed in a pilot study. Fifty nine participants who did not take part in the main experiments read versions of the vignettes in which the eyewitness provided only one category for the critical object. Participants were then asked to rate the likelihood that the defendant was guilty on a 100-point scale (1= not at all likely, 100 = very likely). Defendants seen carrying "related objects" (*knife, stolen watch*) were rated as more likely to be guilty (M = 62.85) than those carrying "unrelated objects" (*mobile phone, keys*), (M = 35.37), $F(1, 55) = 27.24$, $p < .001$.

Procedure Participants were told that they were to play the role of a juror in determining the guilt of a defendant in a criminal trial. Each vignette was then presented on a computer screen with the eyewitness evidence about the critical object alternatives written in bold type. After reading the vignette participants clicked an on-screen button which started a series of questions (with the vignette no longer visible). The first was a categorization question that asked participants to rate the percentage likelihood that the item the defendant was carrying was (a) the primary category, (b) the secondary category, or (c) some other item, with the restriction that the three estimates must sum to 100 percent (see Murphy & Ross, 2010 for a similar procedure). This was followed by a filler question asking participants to recall the general location of the crime as described in the vignette. The final two questions required participants to infer defendant culpability based on the trial evidence. Participants first estimated the likelihood that the defendant was guilty on a 100-point scale (1=not at all likely, 100=very likely). They then rendered a binary verdict ('guilty' or 'not guilty') by clicking on one of two forced choice buttons.

Previous research (e.g., Harris & Hahn, 2009) indicates that mock jurors often give less weight to evidence that they perceive to be inconsistent. In our experimental vignettes all eyewitnesses expressed some inconsistency about the identity of the critical object. However, it is possible that the effects of this inconsistency on juror confidence in eyewitness evidence may have differed across experimental conditions (e.g., jurors may give more weight to evidence when both objects are positively related to the crime, than when one is related and the other is not). To check on this possibility all participants were also asked to rate their confidence in the reliability of the eyewitness on a seven-point scale (1= 'very low confidence', 7= 'very high confidence').

Within each condition participants completed both robbery and arson vignettes with order of vignette administration counterbalanced.

Results and Discussion

Preliminary Analyses. The experimental predictions are based on the assumption that participants believe that the critical object was more likely to belong to the primary than the secondary category. The object categorization data identified 12 participants who did not rate the primary category as more likely. Consequently they were excluded from further analyses.

Preliminary analyses confirmed that the percentage likelihood ratings for membership of the critical object in the primary (M = 69%) and secondary categories (M = 26%) were close to those stated or implied in the vignettes. Notably participants in all conditions rated the likelihood of secondary category membership as well above zero (group means ranged from 21% to 30%). Hence participants in all conditions grasped the uncertainty about the category membership of the critical object.

Preliminary analyses confirmed that there were no significant differences between the robbery and arson vignettes in overall guilt estimates or ratings of reliability of eyewitness evidence. Consequently, in both experiments analyses of guilt judgments were collapsed across vignettes.

A two-way analysis of variance revealed a significant main effect of secondary category association on ratings of eyewitness reliability, $F(1,82) = 6.36$, $p = .01$. Witnesses in

the secondary related conditions were rated as more reliable (M = 4.1) than those in the secondary unrelated conditions (M = 3.52). Hence, eyewitness reliability ratings were included as a covariate in analyses of guilt judgments.

Guilt Likelihood Ratings and Guilt Verdicts. Both variates were analyzed in 2(Primary category status) × 2(Secondary category status) analyses of covariance with eyewitness reliability entered as a covariate. Figure 1 shows the adjusted guilt likelihood ratings for each condition. Not surprisingly the defendant was rated as more likely to be guilty when the primary category for the critical object was closely linked to the crime than when it was unrelated, $F(1,81) = 32.97$, $p < .001$. The more critical question was whether the status of the secondary category affected guilt estimates. There was no main effect of secondary category status on guilt likelihood ratings, $F(1,81) = 0.02$, $p = .89$, but there was a significant interaction between the status of the primary and secondary categories, $F(1,81) = 5.05$, $p = .027$. However, when Tukey's HSD tests were applied, there was no significant effect of secondary category association in the critical comparison between the two Primary Related conditions ($q = 2.45$, $p>.05$) nor between the two Primary Unrelated conditions ($q = -2.16$, $p > .05$).

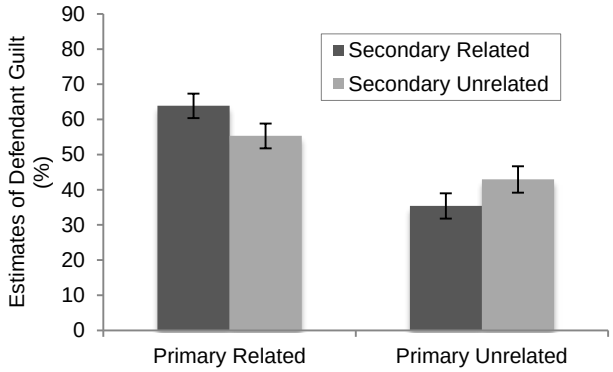

Figure 1. Mean likelihood ratings of defendant guilt (with standard error bars).

The proportion of guilty verdicts was higher when the primary category was related to the crime (M = 0.6) than when it was unrelated (M = 0.21), $F(1,81) = 28.80$, $p < .001$. However, there was no main effect or interaction involving the status of the secondary category (p's >.15). In other words there was no evidence that participants considered the secondary category when deciding on guilt verdicts.

In sum, this study found minimal evidence for multiple-category reasoning in a forensic context. Note however, that there was considerable variance in guilt judgments *within* conditions. For binary judgments for example, the mean standard error across experimental conditions was 0.073 (in a scale ranging from 0.0-1.0). This high level of variability is consistent with previous findings that individual jurors presented with the same evidence often give very different absolute estimates of guilt (e.g., Solana, García, & Tamayo, 1998). Such levels of variability are likely to have reduced the sensitivity of our tests of multiple-category reasoning. Experiment 2 addressed this issue by manipulating the status of the secondary category *within* subjects.

Experiment 2

Experiment 2 aimed to provide a more sensitive test of whether mock jurors use multiple-category reasoning when faced with uncertain alternatives. The design was similar to the first study except that all participants completed two versions of each trial vignette; one where the secondary category was related to the crime, and one where it was unrelated. Multiple-category reasoning would be indicated if different judgments about guilt are given in these conditions.

Method

Participants Forty-two undergraduate students participated for course credit. The majority were female (n = 29) and the mean age was 20.69 years (SD = 4.98).

Design and Procedure The design of Experiment 2 followed the description given in Table 1. Unlike the previous study however, the status of the secondary category (related or unrelated to the crime) was manipulated within subjects. All participants completed two versions of each of the robbery and arson vignettes; one with a crime-related secondary category, and one with an unrelated secondary category. The order of presentation of these alternate versions was counterbalanced across participants. To reduce sequencing effects alternate versions of the same vignette were never presented consecutively. Before completing the study participants were warned that they would sometimes be reading summaries with similar details and were instructed to "do your best to evaluate the trial summaries independently".

As in the previous study the status of the primary category was manipulated between subjects. Equal numbers were randomly allocated to conditions in which the primary categories were related to the crime or were unrelated. In all other respects the procedure was identical to Experiment 1.

Results

Preliminary Analyses Eleven participants failed to assign the highest categorization rating to the primary category in at least one scenario, and were excluded from further analyses. Preliminary analyses again found that for the remaining participants, ratings of the likelihood that the critical object belonged to the primary and secondary categories closely matched the probabilities stated or implied in the vignettes. Notably, there was no evidence of sequencing effects on guilt judgments for the secondary related and secondary unrelated versions of each vignette. The guilt estimates for these alternate versions were unaffected by which version was presented first ($F<2.0$). As in the previous study however, eyewitness reliability was rated higher in secondary related than the secondary

unrelated conditions, $F(1,29) = 4.23$, $p < .05$. Consequently, eyewitness reliability scores were again included as covariates in guilt analyses.

Guilt Likelihood Ratings and Guilt Verdicts Both variates were collapsed across scenarios and the binary verdicts were coded as per Experiment 1. Both data sets were entered into separate 2 (primary category status) × 2 (secondary category status) multivariate analyses of variance, with repeated measures on the second factor and eyewitness reliability scores entered as covariates. Figure 2 shows adjusted mean guilt likelihood ratings. Once again there was a significant main effect of primary category status, $F(1,28) = 27.22$, $p < .001$, with higher guilt estimates when the primary category was incriminating than when it was not. More importantly, in this case there was also a robust effect of secondary category status, $F(1,28) = 36.46$, $p < .001$, but no primary x secondary category interaction, ($F < 1$). When the secondary category was unrelated to the crime, ratings of guilt likelihood were lower than when that category incriminated the defendant. This shows that participants were factoring both primary and secondary categories into their guilt estimates.

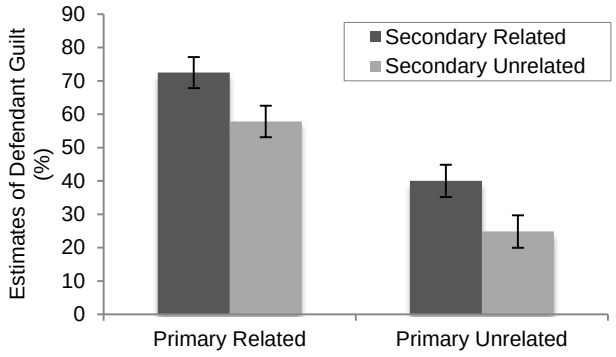

Figure 2. Mean likelihood ratings of defendant guilt (with standard error bars).

A similar pattern was observed for binary guilt verdicts. The proportion of guilty verdicts was higher when the primary category was related to the crime ($M = 0.63$) than when it was unrelated ($M = 0.13$), $F(1,28) = 22.83$, $p < .001$. Critically there was also a robust effect of secondary category status, $F(1,28) = 8.33$, $p = .01$. The proportion of guilty verdicts was significantly lower when the secondary category was unrelated to the crime ($M = 0.30$) than when it was crime related ($M = 0.46$). There was no primary x secondary category interaction, ($F < 1.0$).

These data indicate that mock jurors were considering both category alternatives when inferring guilt.

General Discussion

The goal of these experiments was to examine whether mock jurors consider more than one alterative scenario when inferring defendant guilt on the basis of uncertain evidence. In our trial vignettes there was always some uncertainty about the category membership of a critical piece of evidence, with a more likely (primary) and a less likely (secondary) category alternative. Mock jurors in both studies acknowledged this uncertainty, recognizing that the object *could* have belonged to the secondary category.

In Experiment 1 we found minimal evidence that jurors factored these uncertain alternatives into their guilt judgments but analyses of the critical comparisons did not find these differences to be reliable. Notably though when within-condition variance was reduced by manipulating secondary category status within-subjects (Experiment 2), robust evidence of multiple-category reasoning was found. When there was a possibility that the defendant was holding an innocuous rather than an incriminating object, participants reduced their ratings of guilt likelihood and were less like to return a verdict of guilty.

This is an important result because most previous work (e.g., Malt et al., 1995; Murphy & Ross, 2007, 2010; Ross & Murphy, 1996) has failed to find evidence that people consider more than one category when making predictions about objects whose category membership is uncertain. Ross and Murphy (1996) reported some evidence of multiple-category reasoning but only when the secondary category was more highly associated with the prediction than the primary category (equivalent to our primary unrelated condition). In Experiment 2 however, we found multiple-category reasoning in both primary related and primary unrelated conditions. This result is particularly interesting because it shows that consideration of the secondary category can lead to either decreases in the probability of a given prediction (in the Primary related conditions) or increases in prediction probability (in the Primary unrelated conditions).

One concern is that our strongest evidence of multiple-category reasoning was found when participants completed both the secondary related and secondary unrelated versions of the vignettes. Under these conditions participants could conceivably compare the structure of the two scenarios and this may have increased their sensitivity to the role of the secondary category in determining guilt. In other words, the strong evidence of multiple-category reasoning may have been an artifact of the repeated measures design. To examine this possibility we looked at participant guilt judgments the first time the vignettes were presented and then the second time the vignettes were presented as two separate sets of data. This order-artifact account predicts that we should only see evidence of multiple-category reasoning the second time that a participant sees the vignettes as it is the exposure to the first version that directs attention to the secondary category. Contrary to this account however, the effect of the secondary category on guilt ratings was robust on the first presentation on both the guilt likelihood ratings, $F(1,26) = 7.70$, $p < 0.05$, and the proportion of guilty verdicts, $F(1,26) = 16.02$, $p < 0.01$. In other words, the first time participants read the vignettes

they factored category uncertainty into their guilt ratings.[1].

So just why did we succeed in finding robust evidence in multiple-category reasoning in Experiment 2 when a majority of previous studies have failed to do so? Further research will be required to give a complete answer to this question. As noted earlier however, the highly consequential nature of forensic decisions may lead participants to be more reflective in their consideration of uncertain alternatives. An analogous finding is that expert clinicians have been shown to consider multiple uncertain categories but only when making clinically relevant predictions (Hayes & Chen, 2008). When required to make predictions about nonclinical materials they ignored category uncertainty.

Overall these studies provide qualified support for the conclusion that jurors can use multiple-category reasoning when making inferences about guilt. Clearly much further work is needed to be able to generalize these results to more realistic trial contexts where, for example, the explicit probabilities of different alternative scenarios are unlikely to be given. Nevertheless our findings suggest that, at least in some contexts, jurors can satisfy the legal imperative to consider multiple uncertain aspects of evidence.

Acknowledgments

This work was supported by Australian Research Council Grant DP0770292 to the second author. We would like to thank Melissa Lim for her assistance with programming.

References

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98,* 409-429.

Attorney General's Department of NSW. (2007). *A guide for jurors*. Sydney, NSW: Court Services, Attorney General's Department of NSW.

Bornstein, B. H., & Greene, E. (2011). Jury decision making: Implications for and from Psychology. *Current Directions in Psychological Science, 20*, 63-67.

Gilovich, T. D., & Griffin, D. W. (2010). Judgment and decision making. In T. Susan & D. T. Gilbert (Eds.), *Handbook of social psychology, Vol 1 (5th ed.)* (pp. 542-588). Hoboken, NJ: John Wiley & Sons Inc.

Lagnado, D. A. (2011). Thinking about evidence. *Proceedings of the British Academy, 171*, 183–223.

Harris, A. J. & Hahn, U. (2009). Bayesian rationality in evaluating multiple testimonies: Incorporating the role of coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1366-1373.

Hayes, B. K**.** & Chen, T-H. J. (2008). Clinical expertise and reasoning with uncertain psychodiagnoses. *Psychonomic Bulletin & Review, 15*, 1002-1007

Hayes, B. K., Heit, E., & Swendsen, H. (2010). Inductive reasoning. *Wiley Interdisciplinary Reviews: Cognitive Science, 1,* 278–292.

Hayes, B. K. & Newell, B. R. (2009). Induction with uncertain categories: When do people consider the category alternatives? *Memory & Cognition, 37*, 730-743.

Kuhn, D., Weinstock, M., & Flaton, R. (1994). How well do jurors reason? Competence dimensions of individual variation in a juror reasoning task. *Psychological Science, 5*, 289-296

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*, 480-498.

Malt, B. C., Ross, B. H., & Murphy, G. L. (1995). Predicting features for members of natural categories when categorization is uncertain. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 646-661.

Murphy, G. L., & Ross, B. H. (2007). Use of single or multiple categories in category-based induction. In A. Feeney & E. Heit (Eds.), *Inductive reasoning: Experimental, developmental, and computational approaches* (pp. 205-225). New York, NY: Cambridge University Press.

Murphy. G. L., & Ross, B. H. (2010). Uncertainty in category-based induction: When do people integrate across categories? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 263-276.

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology, 62*, 189-206.

Ross, B. H. & Murphy, G. L. (1996). Category-based predictions: Influence of uncertainty and feature associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 736-753.

Solana, E. F., García, J., & Tamayo, I. (1998). Some individual differences in perception of the evidence and the verdict choice. *Psychology, Crime and Law, 4*, 361-373.

Appendix

Excerpt from the Robbery Vignette used in Experiments 1-2 showing the primary categories (in bold) and secondary categories (in italics).

Ms. Kelly advised that the defendant Mr Bell was carrying a metallic looking object in his hands. She believed that the object was most likely a **KNIFE/ KEYS**. When asked to estimate her certainty Ms. Kelly advised that she was "reasonably certain, at least 70%". She advised that there was "a small chance" that the object was the *STOLEN WATCH/ a MOBILE PHONE*. In summary, Ms. Kelly testified that there was a small chance that the defendant was carrying the **STOLEN WATCH/ a MOBILE PHONE** but she believed the defendant was most likely carrying a *KNIFE/ KEYS*. Note that Ms. Kelly was sure that the defendant was only carrying one object.

[1] It should be noted that although this first presentation is similar in design to Experiment 1, different results were obtained. There is no clear explanation for this discrepancy. One possibility is that this was due to the new instructions used in Experiment 2, which strongly encouraged participants to attend the details of the vignettes.