

Evaluating the impact of visualization of wildfire hazard upon decision-making under uncertainty

Lisa Cheong, Susanne Bleisch, Allison Kealy, Kevin Tolhurst, Tom Wilkening & Matt Duckham

To cite this article: Lisa Cheong, Susanne Bleisch, Allison Kealy, Kevin Tolhurst, Tom Wilkening & Matt Duckham (2016): Evaluating the impact of visualization of wildfire hazard upon decision-making under uncertainty, International Journal of Geographical Information Science, DOI: [10.1080/13658816.2015.1131829](https://doi.org/10.1080/13658816.2015.1131829)

To link to this article: <http://dx.doi.org/10.1080/13658816.2015.1131829>



Published online: 20 Jan 2016.



Submit your article to this journal [↗](#)



Article views: 26



View related articles [↗](#)



View Crossmark data [↗](#)



Evaluating the impact of visualization of wildfire hazard upon decision-making under uncertainty

Lisa Cheong ^a, Susanne Bleisch^b, Allison Kealy^a, Kevin Tolhurst^c, Tom Wilkenning^d and Matt Duckham^e

^aDepartment of Infrastructure Engineering, University of Melbourne, Melbourne, VIC, Australia; ^bSchool of Architecture, Civil Engineering and Geomatics, FHNW University of Applied Sciences and Arts Northwestern Switzerland, Muttenz, Switzerland; ^cDepartment of Forestry and Ecosystem Science, University of Melbourne, Melbourne, VIC, Australia; ^dDepartment of Economics, University of Melbourne, Melbourne, VIC, Australia; ^eSchool of Mathematical and Geospatial Sciences, RMIT University, Melbourne, Australia

ABSTRACT

The paper investigates whether the methods chosen for representing uncertain geographic information aid or impair decision-making in the context of wildfire hazard. Through a series of three human subject experiments, utilizing 180 subjects and employing increasingly difficult tasks, this research evaluates the effect of five different visualizations and a text-based representation on decision-making under uncertainty. Our quantitative experiments focus specifically on the task of decision-making under uncertainty, rather than the task of reading levels of uncertainty from the map. To guard against the potential for generosity and risk seeking in decision-making under uncertainty, the experimental design uses performance-based incentives. The experiments showed that the choice of representation makes little difference to performance in cases where subjects are allowed the time and focus to consider their decisions. However, with the increasing difficulty of time pressure, subjects performed best using a spectral color hue-based representation, rather than more carefully designed cartographic representations. Text-based and simplified boundary encodings were among the worst performers. The results have implications for the performance of decision-making under uncertainty using static maps, especially in the stressful environments surrounding an emergency.

ARTICLE HISTORY

Received 15 June 2015

Accepted 10 December 2015

KEYWORDS

Uncertainty; visualization; decision-making; wildfire hazard; evaluation

1. Introduction

When designing information displays to support spatial decision-making, an important question is how to convey the inherent uncertainty connected with any geographic information. This problem is especially pertinent in emergency response situations, where decisions based on uncertain and complex information may have lasting and significant consequences.

Visualization researchers have already proposed a variety of techniques for representing the uncertainty associated with mapped data (see [Section 2](#)). However, this paper

extends existing research in two important aspects. First, existing research has focused primarily on *map-reading* tasks, such as whether humans can correctly identify the levels of uncertainty from a map. In this work, the aim is to go a step further and focus on the *decision-making* task: to what extent the representation of uncertainty affects the actual decision taken. Unlike the map-reading task, in which a person may unambiguously be judged to be right or wrong in their performance, the decision-making is significantly more complex. Under uncertainty, it is possible to make the 'right' decision (i.e., the most rational decision on the basis of available evidence) and still achieve the 'wrong' outcome (because of the uncertainty), or vice versa.

The 'right decision, wrong outcome' problem presents particular challenges to experimental evaluation of the performance of decision-makers. Thus, the second key contribution of this research is to explore the application of incentive-based methodologies to the evaluation of spatial decision-making under uncertainty. Incentive experiments, where a proportion of subjects' participation fee is performance-based, are known to be particularly suited to 'risky' choices where participants need to judge or predict outcomes (Camerer and Hogarth 1999). Although incentive experiments are widely used in experimental economics to assist with understanding human judgments of risk, they have not yet been widely applied to spatial decision-making and geovisualization.

In Section 2, a review of the relevant literature in representing uncertainty and evaluating those representations is undertaken. Section 3 then presents a human subject experiment to test the effects of six different representations of uncertainty upon a spatial decision-making task, set in the context of an emergency wildfire scenario. Building on the results of this primary experiment, Section 4 examines whether a more challenging environment (specifically under time pressure or with distractions) can affect decision-making performance. The discussion and conclusions in Section 5 reflect on the implications for spatial decision-making under uncertainty in an emergency context.

2. Background

Uncertainty is an endemic feature of geographic information. This paper focuses specifically on *outcome* uncertainty: uncertainty in the mind of a decision-maker about which locations may be impacted by a near-future event (in our case, the likelihood that a wildfire may burn different locations in a region). There already exists in the literature a substantial body of work investigating cartographic mechanisms for communicating information about uncertainty in geographic information more generally, many familiar from conventional maps.

For example, MacEachren (1992) explored the use of color hue for representing uncertainty, suggesting that this technique is best used for novice users. The evidence for the usability of color value (lightness) is conflicting, with some experiments indicating value is not effective (Schweizer and Goodchild 1992) and others indicating the converse (Leitner and Buttenfield 2000, Aerts *et al.* 2003). Where it is effective, darker values are associated with more certainty, and lighter values with more uncertainty (MacEachren 1992, Buttenfield 1993, McGranahan 1993, Van Der Wel *et al.* 1994). Saturation has been found not to be particularly effective or suitable for representing uncertainty (Buttenfield 1993, MacEachren *et al.* 2012). The shape, size, and orientation of symbols have been used to represent information about uncertainty in maps, as well as the arrangement (pattern) of groups of symbols. Howard and MacEachren (1996)

found linear patterns overlaid on top of standard maps to be an effective way of communicating levels of uncertainty. The use of glyphs to represent uncertainty has also been studied (Pang *et al.* 1997, Cliburn *et al.* 2002). The findings indicate that glyphs can be successful in the context, although more suited to use by experts as they can be visually overwhelming (Pang *et al.* 1997). Uncertainty overlays have also been successfully tested, using contour lines of different colors to depict the uncertainties associated with different variables (Pang 2001).

Fog, focus, and crispness have been applied to blur or obscure displayed information in order to prevent unwarranted certainty being ascribed to these features (MacEachren 1992). For example, fog can be used as an 'atmosphere' through which the map is viewed; the thicker the fog, the more uncertain the underlying information.

Finally, it is worth noting that attempts to represent uncertainty in geographic information are also common in the wider geovisualization literature. For example, blinking and flickering (Fisher 1993, Evans 1997), movie animations (Blenkinsop *et al.* 2000), and even sound (Fisher 1994, Krygier 1994, Lodha *et al.* 1996, Bearman *et al.* 2013) have been proposed as possible routes to communicating uncertainty in geographic information. However, in this paper, our exploration is restricted to static cartographic techniques, such as might be used on a paper map. Static paper maps are still overwhelmingly the default form of communication of geographic information to responders in emergency scenarios (Cai *et al.* 2006, Klimenko *et al.* 2007, Zlatanova and Li 2008, Brooks and Swaminathan 2010).

2.1. Map reading versus decision-making

Map-based tasks in the literature can be classified along a continuum, where at one end sits *map reading* (i.e., extracting encoded information from a map) and at the other end sits *decision-making* (choosing which action to take based on that encoded information). A common theme throughout the studies into the map-based representation of uncertainty discussed previously is that these sit further toward the *map reading* end of the spectrum: the task of interpreting the level of uncertainty from the map.

Map reading is usually assessed using basic value retrieval tasks (Kinkeldey *et al.* 2014). While such studies still require users to make a decision, these decisions are generally limited to interpreting the information contained in the map. For example, in a typical study by Aerts *et al.* (2003), participants were asked to identify areas of urban growth under uncertainty. The task required users to match colors from the map to the legend to find an area's associated uncertainty. Even when faced with more complex tasks, most previous studies of cartographic representation of uncertainty reduce to a map-reading task. For example, a study by Hope and Hunter (2007b) presented users with a thematic map of land suitability for airport sites overlaid with glyphs indicating the level of certainty. Users were then asked to choose the optimal region for a new airport based on the combination of suitability and certainty. While this task forced users to combine two types of data to make a higher-level decision, it is arguable that the task still required users to perform only basic map reading in managing uncertainty (i.e., matching the uncertainty representation with the value in the legend).

Although there is no crisp distinction between map reading and decision-making, the distinction is still important because different representations may have an effect not

only on a user's ability to interpret from the map the correct level of uncertainty but also on the ultimate action a user selects based on that uncertain information. For example, a user who correctly interprets the level of uncertainty from two maps may still come to different decisions based on that level of uncertainty, depending, for example, on the spatial configuration of uncertainty, such as proximity to more or less uncertainty. Supporting this viewpoint, Mason *et al.* (2014) argued that evaluating the effect of a visualization on a user should include an evaluation of the emotional responses (e.g., trust, confidence, worry, and anxiety). Further, under uncertainty decision-makers must face the 'right decision, wrong outcome' problem: in some cases, a rational decision may still lead to an adverse outcome (and vice versa).

Only a relatively small proportion of work in the literature has examined the effects of different graphical and cartographic representations of uncertainty upon the effectiveness of tasks more strongly associated with the decision-making end of the scale. For example, Finger and Bisantz (2002) used point-based representations of uncertainty (using blurring of icons) to assess effectiveness of users in identifying a radar contact as hostile or friendly. Other such examples are research by Kirschenbaum *et al.* (2013) and Kirschenbaum and Arruda (1994) where representations of uncertainty were tested in submarine localization visualization tasks and research by Andre and Cutler (1998) where display of uncertainty was investigated in navigation tasks. Interestingly, these studies showed that performance was found to be more accurate for novices when uncertainty was displayed spatially.

The work in this paper aligns more closely toward the decision-making end of the task spectrum and investigates cartographic representations of uncertainty and their effect upon improved judgment and decision-making in uncertain situations. Although research into the task of reading uncertain information from a map has an established history, less is known about the effects of uncertainty on spatial decision-making tasks. Focusing our work on the decision-making end of the spectrum is important for at least two reasons. First, from the point of view of emergency response, it is more important to examine if and how the representation of uncertainty affects a user's decision (i.e., which action they choose to make), not only whether that user can correctly interpret the information encoded in a map. In scenarios such as evacuation from a wildfire, numerous factors beyond the simple extraction of information from a graphical representation are important, such as fear, worry, distractions, and risk aversion. Second, only by examining tasks at the decision-making end of the spectrum, beyond only map reading, is it possible to explore the 'right decision, wrong outcome' problem inherent in decision-making under uncertainty.

2.2. Experimental support

Among the many studies of cartographic representations of uncertainty in the literature, only a subset has been empirically evaluated. Some of these empirical approaches adopt a qualitative, questionnaire-based approach common in human factors research. For example, Deitrick and Edsall (2006) performed a study where participants had to interpret uncertainty in land use maps to answer a series of map-reading tasks. Cox *et al.* (2013) examined participants' preferences for different graphical representations of the uncertainty associated with the potential impact of predicted hurricane tracks. A

recent study by Severtson and Myers (2013) assessed via surveys participants' emotional responses to uncertainty displayed on cancer risk maps. While Ash *et al.* (2014) asked participants to answer a series of emotional questions based upon their interpretation of spatially communicated tornado warnings. One difficulty with such studies can occur when a small proportion of those subjects invited actually respond. For example, in Kunz *et al.* (2011), approximately 50% of those subjects who were sent surveys actually responded. Such experimental designs leave open the possibility of self-selection bias (i.e., those people who elected to respond may be systematically different in their responses from those that did not).

Incentive-based experiments are one such way to combat self-selection bias to some degree. Others have relied on quantitative experimental evidence in examining the representation of uncertainty, such as in work on using value, saturation, and texture to communicate uncertainty in maps (Leitner and Buttenfield 2000, Viard *et al.* 2011). However, practical considerations may mean such studies use human subjects drawn from students engaged in the researcher's own courses, which introduces the possibility of further bias.

In this context, payment schemes have often been employed in experimental human factors research. Often these payments are structured as flat fee for participation, typically ranging from \$5 USD to \$10 USD (MacEachren *et al.* 1998, Leitner and Buttenfield 2000, Brewer and Pickle 2002). Hourly rates may also be used, for example, by Finger and Bisantz (2002) where participants were paid \$6 USD per hour. Further incentives have also been included in some experiments, such as Bisantz *et al.* (2005, 2011) where participants were paid an hourly rate plus performance-based monetary bonus for the highest scoring participant.

The objective of all these payment schemes is to make explicit the incentives for participants. In the majority of experiments in the literature surveyed, no payments were offered to participants (Schweizer and Goodchild 1992, Monmonier and Gluck 1994, Blenkinsop *et al.* 2000, Cliburn *et al.* 2002, Aerts *et al.* 2003, Lucieer and Kraak 2004, Deitrick and Edsall 2006, Hope and Hunter 2007a, Kunz *et al.* 2011, Viard *et al.* 2011, Severtson and Myers 2013, Ash *et al.* 2014). It is unclear what in each of these instances the incentives were for those participants.

To combat such ambiguities, performance-based incentive payment schemes have been widely used in experimental economics research (Smith and Walker 1993, Camerer and Hogarth 1999, Hertwig and Ortmann 2001, 2003). For example, Camerer and Hogarth (1999) evaluated 74 experiments where zero, low, or high performance-based incentives were paid. They found that the presence and amount of financial incentive does seem to affect average performance in many tasks, particularly judgment tasks where effort responds to incentives and where increased effort improves performance (Camerer and Hogarth 1999).

Numerous studies have shown that performance-based incentive payments can have large effects on subject behavior when compared with zero incentive payments, particularly in the domain of risk (Smith and Walker 1993, Camerer and Hogarth 1999, Cubitt *et al.* 2001, Holt and Laury 2002, Bardsley *et al.* 2010). Thus, in our study, we expect positive incentive payments generate a set of motivations that interacts with visualization and time pressure. However, the same is true of an experiment with the

performance-based incentives set to zero: this experimental design is still expected to affect performance.

According to Bardsley *et al.* (2010), we can expect performance-related incentives influence behavior in our experiments in three main ways:

- Performance-based incentives provide motivation for human subjects to increase their effort in the decision-making process (Camerer and Hogarth 1999). As a result, the experiment provides data on how an individual acts when their deliberative process are engaged. For our experiments, we are interested in studying decision-making in a wildfire setting where individuals have significant incentives to make decisions accurately and quickly, and where the adverse outcomes are highly salient. It is likely that in this environment, deliberative processes will be engaged, and visualization may help guide individuals to make better decisions.
- Performance-based incentives assist in allowing for 'home-grown' motivations that are typically hard to measure and control (e.g., some subjects may wish to support the research of the experimenter and try to adjust their decisions to generate data that would 'support' the researchers hypothesis; others may wish to have fun or create noise while others may wish to support 'aesthetically pleasing' visualizations by intentionally making mistakes in other formats). A major reason for performance-related rewards is to make the specific decision problem we are interested in salient (i.e., in our experiments, we wish to see the effect of the treatments) and to crowd out intrinsic preferences that we cannot control (Smith 1976).
- Performance-based incentive payments help make an environment more salient to the subject so that the decisions made encode some of the emotional experiences that would be felt in the true environment. In the risk domain, individuals who are placed into hypothetical situations often fail to predict their emotional affect and act differently than they do when faced with payoff-relevant decisions. For example, subjects faced with real financial incentives are more risk averse (Holt and Laury 2002), exhibit endowment effects where they do not wish to give up things they already possess (Lowenstein and Adler 1995), and are more eager to mitigate losses (Cubitt *et al.* 2001). In this experiment, our interest is in situations where emotional affect is likely to be high. Thus, performance-based incentive payments are expected to generate data in experimental conditions that more faithfully capture the emotional responses of subjects in real scenarios.

In summary, offering performance-based incentives promotes engagement of subjects' deliberative systems, makes the environment more salient for subjects, and helps make the experimental environment more representative of the types of decisions that users of the visualization platform are likely to face. As the aim of this research is to evaluate spatial decision-making under uncertainty, this study adopted an incentive-based methodology from experimental economics in its design.

Thus, the key contributions of this study are to combine three underresearched aspects in the literature on visualization of uncertainty, specifically:

- the investigation of the task of decision-making under uncertainty, rather than map reading of uncertain information;

- the use of quantitative experiments in exploring decision-making performance under uncertainty; and
- the application of an incentivized experimental design to combat potential presentation effects in subjects.

3. Experiment 1 (primary)

Our experiments are designed to investigate the decision-making performance of individuals given multiple different representations of uncertainty. Of particular interest to us is the 'right decision, wrong outcome' problem. Under uncertainty, people may make the right decision, but due to the underlying uncertainty this decision may result in the wrong outcome.

In our experiments, participants were presented with a decision-making task related to a wildfire. Participants had to choose to *stay* or *leave* a home based upon their interpretation of the potential and uncertain impact of a wildfire. Our objective is to see to what extent the representation used affects not just the participants' decisions but also the outcomes. This primary experiment addresses the most basic scenario, with extensions in response to the results covered in [Section 4](#).

3.1. Participants

Sixty students from the University of Melbourne took part in the primary experiment. The students were selected from an existing experimental economics database of participants. The students were recruited from a wide variety of faculties within the university. Students currently engaged in studying geography or closely related disciplines were actively excluded, as the aim was to focus on participants without specialized knowledge of map-reading concepts. The mean age of participants was 23.8 years, ranging in age from 18 to 53. The majority of participants had no prior experience with wildfires. All participants in the experiment were tertiary educated with degrees ranging from bachelors to PhD, and there was an even distribution between male and female. Approximately 50% of participants listed English as their first language and the other 50% were of Asian, South East Asian, and European descent. All participants had a good grasp of mathematics and 54 of 60 had never formally studied Geographical Information Science (GIS) or cartography.

3.2. Stimuli

Five of the six experimental stimuli were maps consisting of a burn probability surface, a hillshade backdrop, a house location shown using a black cross, and a corresponding legend. The hillshade backdrop was chosen to provide some geographical context. However, roads and other topographic features were omitted from the map to minimize the chance of recognition by any subjects who might have prior knowledge of the area.

Each mapped probability surface showed the different likelihoods of being impacted by a hypothetical fire across the study region. The probability surfaces were generated by the Phoenix RapidFire fire spread modeling software, currently in use in the Victorian Government for fire prediction and management purposes (Tolhurst *et al.* 2008). The

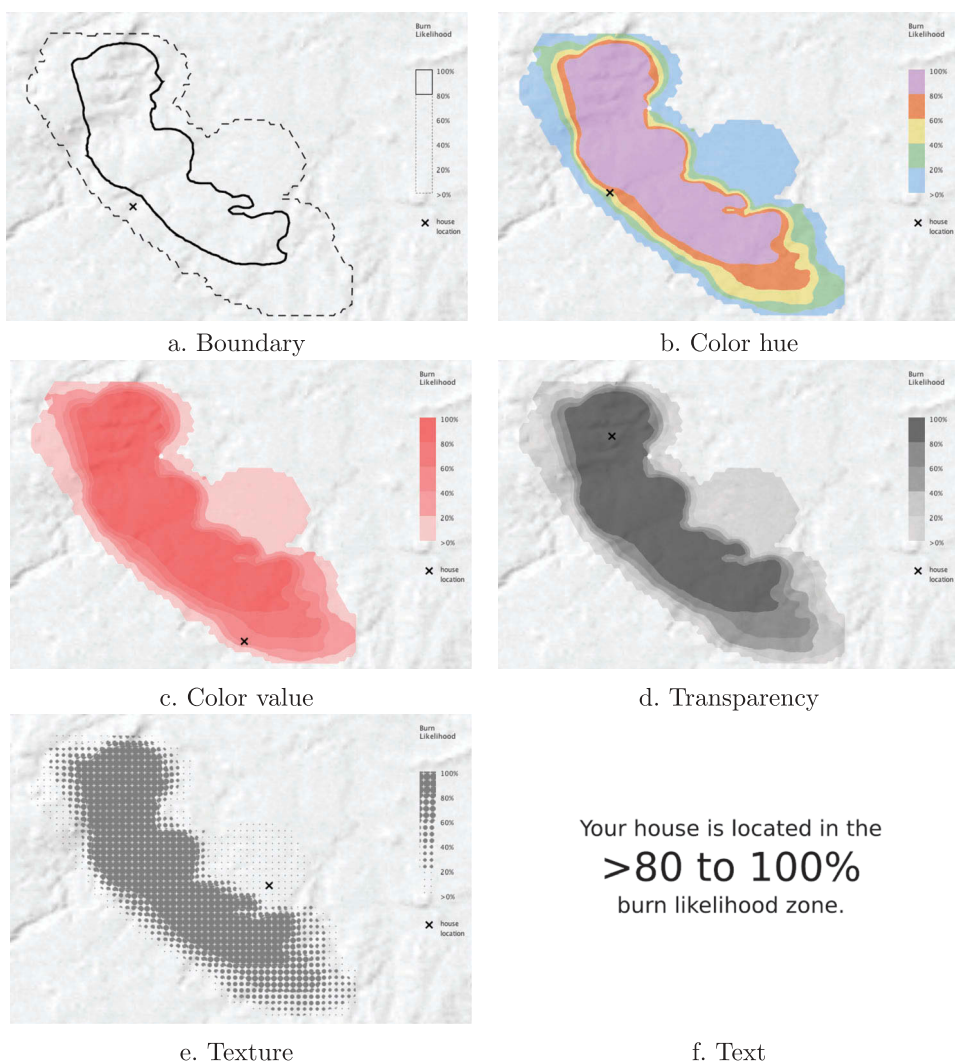


Figure 1. Examples of six experimental stimuli types for representing uncertainty in impact of fire on house location.

stimuli were not tested for being color-blind safe, which could be a potential bias. However, due to the existing complexity of the task and the fact that such a small proportion of the population is color blind, we decided not to introduce this extra factor into our experiment. The probability surface was represented using one of five different cartography representations of uncertainty (Figure 1):

- (a) The simple map *boundary* representation (Figure 1a) was chosen as a similar style of map is currently in use in the fire services sector in Victoria. This simplified representation encoded less information than the other representations, showing only two classes of region: 80–100% likelihood and less than 80% likelihood of burn impact.

- (b) The color hue representation used a spectral color ramp (Figure 1b) to distinguish between five classes of burn likelihood (0–20%, 20–40%, 40–60%, 60–80%, and 80–100%). This representation was again chosen as it is used in practice as the default representation in the FSPRO fire modeling software widely used in the USA (USDA Forest Service 2015).
- (c) A color value representation was chosen as variation to the color hue representation (Figure 1c) but with the addition of containing a clear value message (with lighter value representing decreasing uncertainty). This representation has been found to be successful for depicting uncertainty in a regional planning scenario by Leitner and Buttenfield (2000). In that study, the best results were observed where a darker value was used to represent more certain information.
- (d) Transparency was chosen as the fourth representation (Figure 1d), as it proved successful for representing uncertainty in an experiment conducted by Bisantz *et al.* (2011) using a military decision-making task. A previous human subject study on transparency conducted by Bisantz *et al.* (2009) found transparency to be suitable for ranking levels of uncertainty in thunderstorm mapping.
- (e) A point texture-based representation was also tested (Figure 1e). Texture was described by Leitner and Buttenfield (2000) in their regional planning study, to be the next most successful following color value.

Finally, a sixth, text-based representation of the predicted wildfire likelihood was also included (Figure 1f). This inclusion is in line with research by Ware (2008), who concluded that text may often be superior to images for presenting abstract ideas, logic, and conditional information. Using text enabled an evaluation of whether representing uncertainty through simple text-based descriptions might be more effective for decision-making than through map-based representations, in contrast to some results in the literature (see Section 2.2).

In selecting the different representations, the issue of the precision of information encoded in each of the different representations was also considered (i.e., how detailed the information is in each of the different views). The four ‘normal’ representations (i.e., color hue, color value, transparency, and texture) are most precise and present the most detail. Although it gives a percentage for burn likelihood for the house location, the text representation is arguably less precise as it does not provide the wider geographic context, for example, representing how close the house location is to the boundary of a probability region. This representation is also not entirely the same as the graphic representations as it does not require value retrieval in its interpretation. The boundary representation is arguably least precise as it has no internal gradation from 0% to 80% probability.

3.3. Procedure

The experiment was conducted using between-subjects testing where each participant viewed every scenario once. The experiment was administered in blocks of treatments (six blocks with 10 participants per block). There was randomized ordering within the blocks. Each participant viewed 90 scenarios. For each scenario, participants had simply

to decide whether to 'stay' or 'leave' a home based on the presented information about the potential and uncertain impact of a wildfire.

As was foreshadowed in the previous sections, a key feature of this experiment is the use of a performance-based incentive scheme, adopted directly from standard experimental economics methods. Each participant was paid a flat \$7 AUD participation fee. Up to an additional \$9 AUD could be earned throughout the experiment in performance-based payments. These amounts were chosen to ensure the performance-based payments (i.e., those a participant could earn) could potentially be more than the participation fee, a clear motivation for performing the experiment well (see [Section 2.2](#)). Thus, the total amount that could be earned by a participant was \$16 AUD, slightly more than previous experiments as surveyed by the literature (see [Section 2.2](#)). However, because of the performance-related element, in practice, the most that was earned in any of our experiments was \$14.40 AUD (approximately \$11.50 USD). On average, it took participants approximately 15 minutes to complete the entire experiment.

For each correct *outcome*, a participant was paid \$0.10 AUD. For each incorrect outcome, participants did not receive a payment. An outcome was deemed to be correct if either (a) the participant decided to leave the house and it was in fact impacted by the fire or (b) the participant decided not to leave the house and it was in fact not impacted. Whether the house was in fact impacted (burned) or not was randomly chosen with a probability equal to that represented on the map. Note that this experimental design ensured the experiment was not simply a map-reading task. The information provided to the participants gives the 'likelihood' of an outcome. In a small number of cases (10%), participants who elected to leave a house located in the 80–100% burn likelihood zone (logically, the right decision) would in fact suffer the wrong outcome (the house would not be impacted by the fire). Similarly, in 30% of cases houses located in the 20–40% burn likelihood zone would in fact be impacted by the fire. Our experiment also imposes a cost for choosing to leave when the house is *not* impacted, in which case a participant is paid \$0.00 AUD. This incentive structure reflects the premise that in reality successfully staying and defending a house significantly decreases the chance of property loss, at both that house and neighboring houses. Thus, our experiments embody the 'right decision, wrong outcome' problem, where participants who make rational decision based on reading the map may still result in the wrong outcome (and conversely, some participants who make the wrong decision may by luck result in the right outcome).

A matrix showing decision/outcome payments was displayed on the screen at all times. However, to minimize the influence of a person's fine-grained performance on their progress, participants were not shown the total amount they would be paid until the end of the experiment. A time limit of 30 seconds was chosen for decisions to avoid excessive delays in progress through the experiment. The limit of 30 seconds was chosen, following an initial pilot experiment which indicated it would be ample time to answer questions without stress or discomfort. A countdown timer was shown on the experiment screen showing a countdown from 30 seconds, so participants had knowledge of how much time they had to make a decision.

The experiments were run in a purpose-built experimental economics computer lab. The experiment was administered through an internet browser connected to the

experimental database. Data from participants were anonymized through a unique multiple-digit identifier assigned to each participant. Participants were asked to begin the experiment after reading the instruction screen on the experimental interface (Figure A1), which included sample images for each representation they would be shown during the experiment. See Appendix B for screenshots of the experiment interface.

At the end of the experiment, participants were asked to complete a short questionnaire which included demographic questions, scenario preferences, general comments, questions about the participants' math ability, map-reading ability, and wildfire experience.

3.4. Results

Participants earned an average of \$13.34 AUD in this experiment, comprising an average performance-related incentive payment of \$6.34 AUD plus a \$7.00 AUD participation fee. Overall, performance-related incentive payments ranged from \$4.50 AUD to \$7.40 AUD.

Figure 2 compares the results of a logistic regression (logit) analysis for stay/leave decisions and correct/incorrect outcomes for the primary experiment. Logit is best suited to comparing the response across different treatments because of the binary nature of the data (stay/leave decisions, correct/incorrect outcomes) (Cox and Snell 1989, Collett 1991). In comparing different treatments, logit requires one treatment to be identified as the benchmark against which the other methods were tested. The boundary representation was selected as the benchmark for our experiments, since it was the representation that encoded the least precise information of all the representations. The choice of the boundary representation as the benchmark does not in any way affect our results. While logit requires the selection of a baseline reference representation, the other representations can still be compared with one another. Thus, if we were

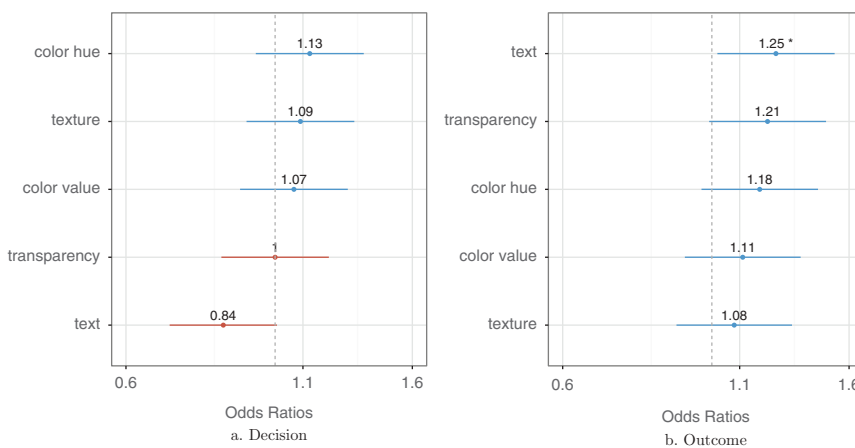


Figure 2. Odds ratios for (a) decisions and (b) outcomes for different representations relative to boundary representation for Experiment 1. Upper and lower 95% confidence limits are represented by the extent of the horizontal bars. A single asterisk indicates significance at the 5% level. As the boundary representation is the referent representation, it is not represented above.

to select another representation as the referent representation, it would not affect the relative odds ratios between other representations.

Figure 2 shows the odds ratios for the different results. An odds ratio is a measure of association between an exposure and an outcome. The odds ratio represents the odds that an outcome will occur given a particular treatment compared to the odds of the outcome occurring in the absence of that treatment. For example, in Figure 2a the odds ratios indicate that participants were 0.84 times less likely to decide to stay in the case of text than in the case of the boundary representation. Conversely, the results in Figure 2b indicate that participants were 1.25 times more likely to make a decision that led to the correct outcome in the case of text compared with the boundary representation.

The logit analysis produces significance values for each representation when compared to the boundary representation. A single asterisk indicates a difference is significant at the 95% confidence level. Figure 2a shows that none of the differences in stay versus leave decisions across each of the different representation was statistically significant at the 5% level. By contrast, in terms of the outcomes in Figure 2b, the text-based representation led to significantly better outcomes than the boundary representation, although none of the other representations resulted in a significant difference in outcome success.

The mean time taken by participants to decide to each scenario was found similar across all the representations (between 3 and 4 seconds). Users performed fastest on the boundary representation (3.41 seconds), which also resulted in the least successful outcomes (see Figure 2a). The slowest responses were found for color value, color hue, and textural representations (3.45 seconds), noting that these were the next three most accurate behind text. Although analysis of variance revealed the differences in times to be statistically significant, the effect size is very small (at most four hundredths of a second).

Finally, analysis of participant preferences (Figure 3) indicated that the color hue was the most preferred representation, with the text representation – which led to the most successful outcomes – the next most preferred.

3.5. Discussion

These results highlight the distinction between decision and outcome. Participants apparently took somewhat different decisions based on the different representations (e.g., with rather fewer participants electing to stay given the text-based representation), even though the observed differences were not statistically significant.

However, the differences in decision do give rise to a statistically significant difference in outcomes. Subjects were least successful when using the boundary representation, and most successful when using the text-based representation, which led to 25% more successful outcomes than the boundary representation. This result apparently contradicts results of other studies that demonstrate the utility of graphical representations (see Section 2.2), although it is in line with the conclusions of Ware (2008) (that text may sometimes be superior to images for abstract information).

Conversely, the boundary representation was least successful for users, although there is evidence that the boundary representation can be handled significantly more rapidly. One interpretation of this result is that the boundary representation encodes less information

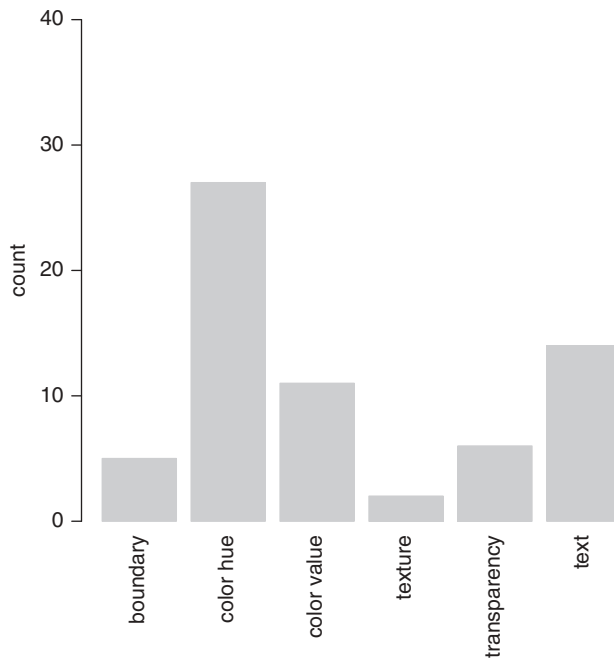


Figure 3. Histogram showing user preference count for primary experiment. Order of the bars is the same as in Figure 6.

than the other representations (i.e., with only one category for 0–80% likelihood of fire impact, compared with four categories in the other representations). Hence, it seems plausible that users can make faster decisions when faced with less information to process. Set against that, participants were able to complete the task rapidly on average across the board, with almost all responses requiring less than 5 seconds.

The strong user preference for the color hue was counter to expectations (and graphic design principles). The spectral color hue representation lacks any ‘value message’ and so should normally only be considered for the representation for categorical rather than quantitative data. Despite color hue having no clear risk value message, strong preference for this representation could be attributed to participants being able to read it more easily as there is higher contrast between the levels of burn likelihood. The most successful outcomes were achieved with the substantially less-preferred textual representation. Thus, preference alone may not provide a clear guide to the effectiveness of a mechanism for representing uncertainty.

In summary, Experiment 1 provided a useful baseline, with subtle indications of differences in performance, but differences between outcomes that were in most cases not statistically significant. Except in the case of text versus boundary representations, it made little difference to the outcome which representation was chosen. In some senses, this result might be interpreted to indicate that the cartography ‘doesn’t matter’ in this instance of decision-making under uncertainty. Subjects performed better using textual information than with a map, and aside from the less precise boundary representation, none of the different graphical depictions of uncertainty had much impact upon the outcomes.

However, another possible interpretation of this result was that the task was too easy, with participants able to consider at their leisure all the information presented, in turn leading to little differences in outcomes. Increasing the difficulty of the experiment, however, might reveal greater differences in outcomes, the focus of the next two experiments.

4. Experiments 2 and 3

Based on the results of the primary experiment, two further adaptations of Experiment 1 were designed that would increase the difficulty of the task. In particular, time pressure and distractions are common features of emergency decision-making scenarios. Subsequent Experiments 2 and 3 aimed to examine the performance of the different representations in such scenarios. The experiments again used similar performance-based incentives to Experiment 1. It is important to emphasize all three experiments follow the same controlled experiment design. Across each of the experiments, the same information is presented to the subjects under the same conditions but using different representations. Thus, we ascribe differences in performance evident across the set of subjects to the effect of the different representations.

4.1. Procedure: time pressure (Experiment 2)

Time pressure is a factor in real wildfire decisions. Reducing the amount of time participants were given to make a decision seemed a natural next step in making the task more difficult. Experiment 2 used identical experimental stimuli and between-subjects methodology as the primary Experiment 1. Participants were tested in the same experimental laboratory environment, using the same computers, and given the same exit questionnaire to control for subject differences. However, instead of 30 seconds, participants were given only 5 seconds decision-making time per stimulus with a 1 second pause in between each scenario (where participants were shown whether they were correct or incorrect and the amount earned). While this timing was a considerable reduction in the time available when compared with the 30 seconds in Experiment 1, it is still longer than the 3–4 seconds taken on average to make a decision in that experiment.

Other minor changes to the experimental interface were made to assist participants in performing the task under time pressure. Specifically, the information on the correct/incorrect screen was further simplified and enlarged, and the stay/leave buttons were also enlarged (Figures A2 and A3). Further, stimuli were presented in blocks of 15, with a pause in between requiring the participant to confirm when they were ready to continue to the next 15 stimuli. These periodic breaks provided some respite for participants, useful in case participants became confused or overwhelmed rather than simply time-pressured. Finally, in line with the objective of rewarding rapid decisions, participants were not paid for a decision if they ran out of time.

4.2. Procedure: dual task (Experiment 3)

Distraction is also a factor in emergency decision-making. Experiment 3 used a dual-task design, so participants would have to make decisions with competing information and attention. Experiment 3 followed the Experiment 1 design, but where the original stimuli

were paired with a secondary task. The secondary task involved a series of sentences to memorize. For each sentence, participants were later required to answer whether two statements about the prepositional content of each sentence was true or false. The sentences used had been shown in previous research on working memory to work well in dual-task experiments (Fedorenko *et al.* 2007). As an example, one sentence used was: 'The detective who recognized the spy crossed the street at the light.' One of the corresponding (true) statements was: 'The detective crossed the street at the light' (i.e., as opposed to the spy). [Appendix A](#) contains a full list of sentences and their corresponding true/false questions as used in this experiment.

Six different sentences and their corresponding statements were spliced between the blocks of 15 stimuli (i.e., participants memorized a sentence, completed 15 stimuli, indicated whether two statements about the sentence were true or false, and then repeated the sequence). The timing for this experiment was the same as that for primary Experiment 1 (30 seconds per scenario) including that for secondary tasks (i.e., up to 30 seconds to memorize each sentence and indicate whether each corresponding statement was true or false).

4.3. Participants

Sixty new participants were recruited for each of Experiments 2 and 3. In each experiment, any participants who had participated in previous experiments were excluded, to guard against learning effects. The population of participants had similar characteristics to those in Experiment 1: mean age approximately 24, ranging from 18 to 63 years; education ranging from year 12 to PhD; relatively even distribution between the sexes with 60% male; approximately 50% of participants listing English as their first language with the remainder of South East Asian, European, or Asian descent; and all participants having a good grasp of mathematics, with only 5% having previously studied GIS or cartography.

4.4. Results

The logistic regression analysis used for Experiment 1 was again used to examine any differences in decisions and outcomes for the time pressure and dual task experiments (2 and 3). The boundary representation was again chosen as the benchmark for all analyses. The odds ratios of a correct response of each of the representations relative to the boundary method were also computed.

[Figure 4](#) provides a graphic representation of odds ratios and confidence intervals for Experiment 2, time pressure. The results show that while only transparency led to significantly different decisions when compared to the boundary representation ([Figure 4a](#)), with the exception of text all of the representations led to significantly better outcomes than boundary. In particular, the color hue representation again performed particularly well, leading to the best overall outcomes (1.37 times more likely to lead to a correct decision) significant at the 99% confidence level.

[Figure 5](#) provides the same results for Experiment 3, dual task. The picture in Experiment 3 is different again, with none of the representations leading to a

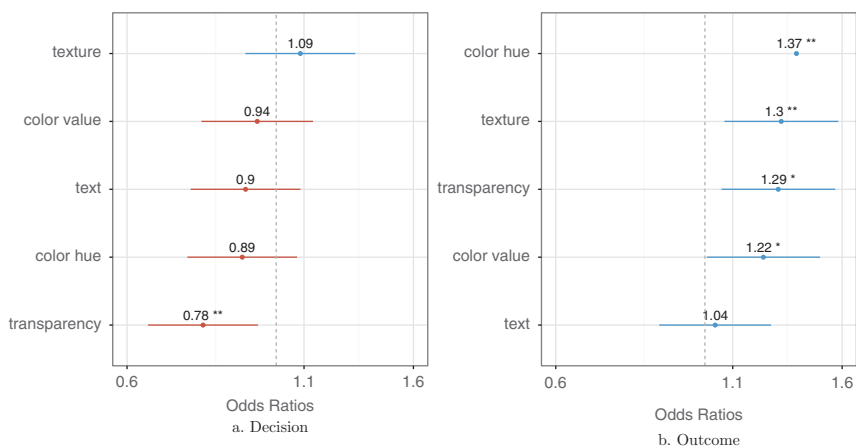


Figure 4. Odds ratios for (a) decisions and (b) outcomes for different representations relative to boundary representation for Experiment 2 (time pressure). Upper and lower 95% confidence limits are represented by the extent of the horizontal bars. A single asterisk indicates significance at the 5% level and the double asterisk indicates significance at the 1% level. As the boundary representation is the referent representation, it is not represented above.

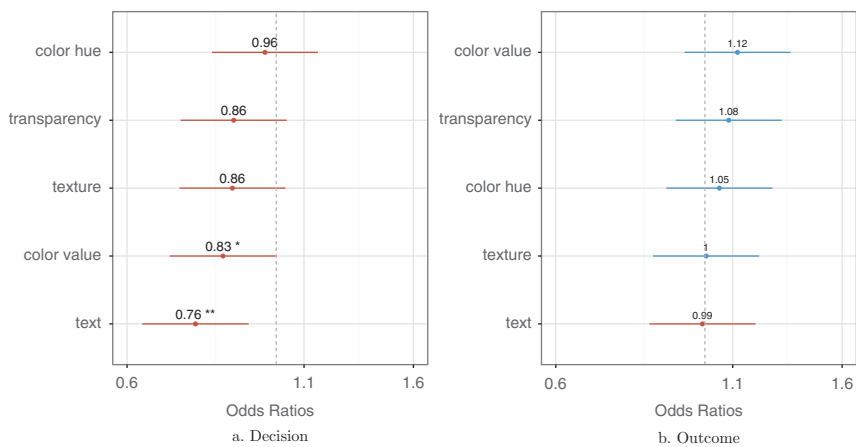


Figure 5. Odds ratios for (a) decisions and (b) outcomes for different representations for Experiment 3 (dual task). Upper and lower 95% confidence limits are represented by the extent of the horizontal bars. A single asterisk indicates significance at the 5% level and the double asterisk indicates significance at the 1% level. As the boundary representation is the referent representation, it is not represented above.

significantly different outcome and with text and color value leading to significantly fewer 'stay' decisions.

On average, subjects earned \$13.00 AUD for participation in the time-pressure experiment, of which \$6.00 AUD was performance related (ranging from \$4.70 AUD to \$7.20 AUD). By contrast, subjects earned average of \$5.36 AUD performance-related incentives for the dual experiment with amounts varying from \$4.40 AUD to \$6.40 AUD.

4.4.1. Preferences

Figure 6 shows the histograms of user preferences for Experiments 2 and 3. The pattern of preferences is broadly consistent with that in Experiment 1 (Figure 3), but with a marked reduction in preference for the text-based representation in Experiment 2, time pressure.

4.4.2. Time analysis

Examining the average time taken across the three experiments shows only relatively small differences in time taken (Table 1). In particular, the effect on time taken of the time-pressure experiment is a reduction in time taken of approximately two tenths of a second on average. Interestingly, in Experiment 2, users took slightly longer to make decisions with the text and boundary representations, and they were also less accurate with these two representations.

4.5. Discussion

The results of the time-pressure experiment did indeed lead to a different pattern of results. All of the other map-based representations outperformed the boundary representation. This result is further evidence that users were able to account for and include in their decision-making the varying levels of uncertainty encoded in these representations, but not encoded in the boundary representation. Only text led to outcomes as poor as the boundary representation. It seems likely that under time pressure, map-

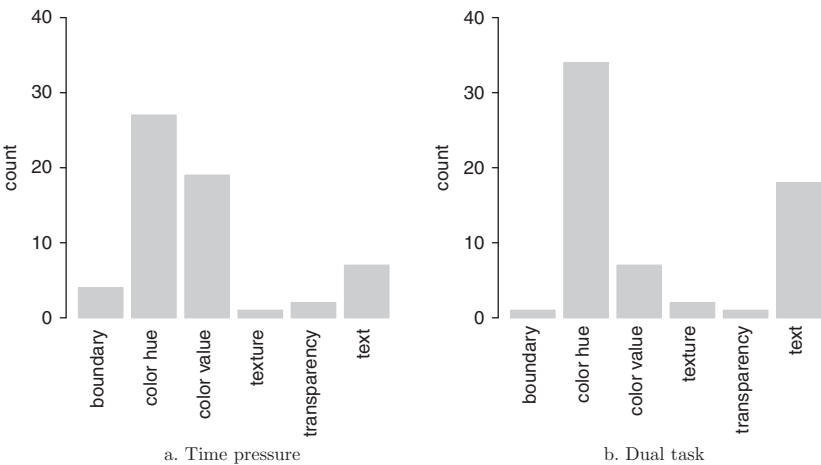


Figure 6. Histograms of user preferences for (a) Experiment 2 (time pressure) and (b) Experiment 3 (dual task). Order of the bars is the same as in Figure 3.

Table 1. Mean time taken (seconds) to decide across Experiments 1–3.

Experiment	Representation					
	Boundary	Color hue	Color value	Transparency	Texture	Text
1. Primary	3.41	3.45	3.45	3.44	3.45	3.42
2. Time pressure	3.23	3.21	3.21	3.20	3.20	3.25
3. Dual task	3.41	3.42	3.40	3.42	3.41	3.39

based information is easier to use rapidly, an impression somewhat supported by the longer time required to decide using the text-based representation (a small effect size – at most one twentieth of a second – but significant at the 1% level).

A clear increased preference for the color value representation was evident in Experiment 2, when compared with Experiment 1, with text showing a clear decreased preference. However, color hue remained the most preferred, and in this case was also the most successful in terms of outcomes. The participants' preference for color hue is one possible reason for the improved performance (e.g., participants liked the color hue representation more, so perhaps paid more attention to it). However, this interpretation is somewhat contradicted by the results in Experiment 1, where color hue was more strongly preferred, but was not the most successful in terms of outcomes. In general, visualization literature indicates that user preferences do not typically influence performance (Wilkening 2009, Quispel and Maes 2014). This success might instead be attributed to the color hue's ability to provide a 'quick-and-dirty' answer. This concept aligns with research by Wilkening (2009), who found that people using maps for decision-making indeed prefer different types of map displays according to whether under time pressure or not. Although the representation does lack a value message, there are in most cases clearer distinctions between the different zones than are possible with more elegant and graphically correct choices of visual variable (i.e., color value or transparency). This lack of value message is comparable to the 'rainbow color map' issue commonly encountered in visualization literature (Light and Bartlein 2004, Borland and Taylor II 2007). The use of the rainbow color map is cartographically inelegant, confusing, and obscuring, but it is still widely used because it successfully encodes continuous rather than discrete surfaces (Borland and Taylor 2007).

The time-pressure experiment did indeed lead to a small but significant decrease in the time taken by participants to decide (up one quarter of a second on average). The slowest responses were also observed when using the worst-performing representations (boundary and text). Further, subjects were somewhat less accurate overall in Experiment 2, making more incorrect decisions than Experiment 1 and earning marginally less performance-related incentive payments (on average \$6.00 AUD) for Experiment 2 versus \$6.34 AUD for Experiment 1). It seems reasonable to conclude from these results that Experiment 2 did indeed increase the difficulty of the experiment, and the observed differences between Experiments 1 and 2 might be attributable to this increased difficulty. For example, the differences in outcomes using the text representation may imply that while text can be the best representation in cases where one has time, it may be the worst in cases where time is limited.

In the dual-task experiment, however, none of the representations were significantly different in outcomes from the boundary representation, although both text and color value were associated with significantly fewer 'stay' decisions. The performance of participants was the worst of all three experiments, with participants earning performance-related incentive payments of on average just \$5.36 AUD. In fact, this performance translates into numbers of correct decisions in the dual-task experiment that were not significantly better than purely random decisions. In short, the dual-task experiment effectively reduced participants to decisions that gave rise to outcomes no better than 'flipping a coin.' Our conclusion is that Experiment 3 was too difficult, confounding the participants to the extent that again the representation used did not really matter, and participants performed equally poorly with all representations.

5. Conclusion

This work explored experimentally the effects of different graphical and textual representations of uncertainty on human subjects' abilities to make decisions under uncertainty. The question of how decision-makers perform under uncertainty requires the 'right decision, wrong outcome' problem to be addressed. Thus, our experimental design incorporates as a key feature the possibility that subjects will make the *right* (i.e., rational) decision – for example, decide to leave a home predicted to have an 80–100% chance of being impacted by a fire – that leads to the *wrong* outcome – for example, in a small proportion of instances, homes predicted to have an 80–100% chance of being impacted by a fire will in fact not be burned. Previous work has tended to ignore this issue, focusing instead on evaluating the map reader's success at correctly reading the level of uncertainty from a map.

The experiments compared five map-based and one text-based representation of uncertainty in an emergency decision-making scenario. As all three experiments follow an identical controlled experimental design (i.e., same stimuli, same method for recruiting participants, same experimental conditions, incentives scheme, laboratory, and computers), we can infer that through these experiments we are as accurately as possible measuring people's responses to the stimuli. The results show no evidence of a link between user preference and task performance. For example, color hue was found to be the most preferred representation across all experiments, and while it performed best in terms of outcomes for Experiment 1, this was not the case for Experiments 2 and 3. These findings support those in research conducted by Quispel and Maes (2014) and Wilkening (2009), for example, indicating that user preferences do not typically influence performance. In terms of outcomes, participants performed best when allowed time and focus to consider the decision. In this case, the purely text-based representation outperformed all the map-based representations. In this respect, our results accord with Ware (2008), indicating that text may be superior to graphics for presenting abstract information about uncertainty. This concept draws parallels with work on positional uncertainty by Hope and Hunter (2007a) who also found quicker decisions were made with probability represented textually/numerically than with color value or a representation similar to our boundary representation.

However, as the difficulty of the task was increased, with time pressure introduced, this pattern was reversed with time- and attention-consuming textual information leading to among the worst outcomes. Under time pressure, the best outcome was achieved with a spectral color hue representation. Such a representation is among the most inelegant tested, and acknowledged to be inappropriate for use with quantitative data. Research by Gehlenborg and Wong (2012) on using heat maps reported that 'color can be unreliable when used to represent discrete values' because there is a 'discordance between what we "should" see and what we "actually" see.' However, in our case of time-pressured decision-making under uncertainty, the stark differences in regions of uncertainty, perhaps combined with users' preferences, led to the most rapid and high-accuracy decisions of any representation tested.

Across both experiments, the simplified 'boundary' map, encoding only two rather than five regions of uncertainty, was among the worst performers. Hence, whether using map or text-based representations, subjects were still able to incorporate the additional

information encoded by more detailed representations of uncertainty. A third experiment increased further the difficulty for subjects, distracting them with a dual task. In this experiment, participants' ability to use the information in the uncertainty representation failed completely, with none of the representations improving outcomes above those that might occur by random chance.

Our results have application to the design of maps and graphics for conveying uncertain geographic information to decision-makers, in particular in risk-laden domains such as emergency response. Uncertainty is an essential component of many applications of map products that needs to be accounted for by map designers. In particular, outcome uncertainty – uncertainty in the mind of a decision-maker about which locations may be impacted by a near-future event – is inherent not only in emergency response, but common in other applications such as weather forecasts, climate change, transportation, and urban planning. Experimental designs such as those used in this research can provide quantitative data about user responses to different representations that incorporate the effects of uncertainty on decision-making, not only the map-reading task.

Our experiments and results are targeted specifically at static, paper maps. While other displays are certainly possible, static paper maps are a vital media for communication in emergencies. Such paper maps provide a relatively intuitive and accessible tool for communication of emergency information to the public and between different levels of emergency response personnel. Paper maps are, for example, overwhelmingly the media of choice in use in practice by Victorian emergency responders (Garvey 2013), with whom we work extensively. While higher-level professional emergency personnel may be well-versed in map reading, many emergency response personnel are volunteer members of the local community without any specialist GIS skills or training (Garvey 2013). Static maps, such as those tested, are also frequently provided to members of the public. These maps are used by non-expert individuals to make decisions about their own safety, particularly in Australia where there are no forced evacuations. Despite this ubiquity, for those without high levels of map expertise, the assumptions and limitations of maps are widely acknowledged to hold the potential for misinterpretation (Monmonier 1996). By using student participants with little or no GIS or map-reading experience in our experiments, we aim to capture the responses of this wider audience.

An auxiliary contribution of this paper has been to introduce performance-based incentives into the design of these experiments. Such experimental designs are widely used in experimental economics to control for 'risky' choices, where participants need to judge or predict outcomes, but only infrequently found in visualization and cartography. Experimental examinations of decision-making under uncertainty should account for exactly these risky choices. Without such incentives, it is arguably less likely that results will reflect decision-makers' true incentives and behavior in actual emergency decision-making settings.

A limitation of the study is the focus on a single decision-making task. It is possible that the task itself influences the role and effects of uncertainty upon the decision-maker. In previous work, Riviero *et al.* (2014) found no increases in the time required for air traffic control decision tasks when using transparency to represent uncertainty. St John *et al.* (2000) similarly found that time to make a decision (to attack or wait) did not increase when including symbology to represent the uncertainty in the size of an enemy unit. In contrast, Shattuck *et al.* (2009) found that the time taken to make a decision *increased* when including uncertainty in a troop deployment decision-making scenario.

Future work could extend this research in a number of ways. This initial performance-based incentive experiment deliberately focused on a highly simplified example of a spatial decision: whether a house would be impacted by an uncertain wildfire extent. While this situation was of a spatial nature, knowledge of spatial concepts was not crucial to the decision-making process. Although the task difficulty was increased through both time pressure and distractions, further experiments are planned with fundamentally more complex spatial tasks, such as routing and navigation. Rather than having participants select whether to stay or leave based upon the location of the house, participants would be presented with multiple exit routes and would need to select the best in the presence of uncertain outcomes. Our experiment has also deliberately focused on static cartographic interfaces, but examining decision-making under uncertainty with dynamic and interactive interfaces (e.g., animation or brushing) is another natural extension.

Of course, further complexity would also arise in practical decision-making in the application in the field. Our results shed light on the effects of using different representations of uncertainty in the rarefied atmosphere of the lab, albeit with unambiguous incentives, time pressures, and distractions. However, in practice, decision-making tasks are more complex still, especially in domains such as emergency response. Research into the effects of decision-making under uncertainty in a wider range of contexts might reveal more complex different patterns.

Finally, this experiment has examined just one aspect of uncertainty: the likelihood of impact by wildfire. The experimental method set out in this work could be used for examining not only the effect of different cartographic representations upon spatial decision-making but also the impact of different aspects of uncertainty, including decision-making with imprecise information and decision-making with vague spatial knowledge.

Acknowledgements

This research project has been approved by the University of Melbourne Faculty Human Ethics Advisory Committee, number 1238459.2. We gratefully acknowledge Amy Corman for her invaluable assistance with the human subject experiments, Alexander Klippel for fruitful discussions and generous advice on the original experimental design, Christoph Kinkeldey for his valuable feedback on earlier versions of the paper, and Ingrid Burfurd for the original idea of using performance-based incentives in the evaluation of the different visualizations. Finally, we are grateful for the thoughtful and constructive comments of the anonymous reviewers.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

We gratefully acknowledge the financial support of the Australian Research Council (ARC) including the Discovery Projects Scheme [project DP120100072] 'From environmental monitoring to management: Extracting knowledge about environmental events from sensor data' and ARC Discovery Early Career Research Award [DE140101014] (Wilkenning). This research is supported by funding from the Bushfire Cooperative Research Centre.

ORCID

Lisa Cheong  <http://orcid.org/0000-0001-6046-2205>

References

- Aerts, J.C.J.H., Clarke, K.C., and Keuper, A.D., 2003. Testing popular visualization techniques for representing model uncertainty. *Cartography and Geographic Information Science*, 30 (3), 249–261. doi:10.1559/152304003100011180
- Andre, A. and Cutler, H., 1998. Displaying uncertainty in advanced navigation systems. *Proceedings of the Human Factors and Ergonomics Society*, 42 (1), 31–35. doi:10.1177/154193129804200108
- Ash, K., Schumann III, R., and Bowser, G., 2014. Tornado warning trade-offs: evaluating choices for visually communicating risk. *Weather, Climate and Society*, 6, 104–118. doi:10.1175/WCAS-D-13-00021.1
- Bardsley, N., et al., 2010. *Experimental economics: rethinking the rules*. Princeton: Princeton University Press.
- Bearman, N., Jones, P., and Lovett, A., 2013. Using sound to represent uncertainty in UKCP09 data with Google Maps API. *Atmospheric Science Letters*, 14 (4), 220–226. doi:10.1002/asl2.443
- Bisantz, A., et al., 2011. Comparing uncertainty visualizations for a dynamic decision-making task. *Journal of Cognitive Engineering and Decision Making*, 5 (3), 277–293. doi:10.1177/1555343411415793
- Bisantz, A., Schinzing Marsiglio, S., and Munch, J., 2005. Displaying uncertainty: investigating the effects of display format and specificity. *Human Factors*, 47 (4), 777–796. doi:10.1518/001872005775570916
- Bisantz, A.M., et al., 2009. Visual representations of meta-information. *Journal of Cognitive Engineering and Decision Making*, 3 (1), 67–91. doi:10.1518/155534309X433726
- Blenkinsop, S., et al., 2000. Evaluating the perception of uncertainty in alternative visualization strategies. *Cartographica*, 37 (1), 1–11. doi:10.3138/3645-4V22-0M23-3T52
- Borland, D. and Taylor II, R., 2007. Rainbow color map (still) considered harmful. In: *IEEE Computer Graphics and Applications*, March/April, 14–17. 10.1109/MCG.2007.323435
- Brewer, C.A. and Pickle, L., 2002. Evaluation of methods for classifying epidemiological data on choropleth maps in series. *Annals of the Association of American Geographers*, 92 (4), 662–681. doi:10.1111/1467-8306.00310
- Brooks, T. and Swaminathan, S., 2010. Integrating the paper and digital environments for crisis/emergency response: lessons learned. In: *Proceedings of the 2010 Global Spatial Data Infrastructures Conference*. Singapore: Global Spatial Data Infrastructure Association.
- Buttenfield, B., 1993. Representing data quality. *Cartographica*, 30 (2–3), 1–7. doi:10.3138/232H-6766-3723-5114
- Cai, G., et al., 2006. Human-GIS interaction issues in crisis response. *IJRAM*, 6, 388–407. doi:10.1504/IJRAM.2006.009538
- Camerer, C. and Hogarth, R., 1999. The effects of financial incentives in experiments: a review and capital-labor-production framework. *Journal of Risk and Uncertainty*, 19 (1), 7–42. doi:10.1023/A:1007850605129
- Cliburn, D., et al., 2002. Design and evaluation of a decision support system in a water balance application. *Computers & Graphics*, 26, 931–949. doi:10.1016/S0097-8493(02)00181-4
- Collett, D., 1991. *Modelling binary data*. Boca Raton, FL: Chapman & Hall/CRC.
- Cox, D.R. and Snell, E.J., 1989. *Analysis of binary data*. Boca Raton, FL: Chapman & Hall/CRC.
- Cox, J., House, D., and Lindell, M., 2013. Visualizing uncertainty in predicted hurricane tracks. *International Journal for Uncertainty Quantification*, 3 (2), 143–156. doi:10.1615/Int.J.UncertaintyQuantification.2012003966
- Cubitt, R., Starmer, C., and Sugden, 2001. Dynamic choice and the common ration effect: an experimental investigation. *The Economic Journal*, 108 (450), 1362–1380. doi:10.1111/1468-0297.00346

- Deitrick, S. and Edsall, R., 2006. the influence of uncertainty visualization on decision making: an empirical evaluation. In: A. Riedl, W. Kainz, and G.A. Elmes, eds. *Progress in spatial data handling*. New York: Springer, 719–738.
- Evans, B., 1997. Dynamic display of spatial data-reliability: does it benefit the map user?. *Computers & Geosciences*, 23 (4), 409–422. doi:[10.1016/S0098-3004\(97\)00011-3](https://doi.org/10.1016/S0098-3004(97)00011-3)
- Fedorenko, E., Gibson, E., and Rohde, D., 2007. The nature of working memory in linguistic, arithmetic and spatial integration processes. *Journal of Memory and Language*, 56, 246–269. doi:[10.1016/j.jml.2006.06.007](https://doi.org/10.1016/j.jml.2006.06.007)
- Finger, R. and Bisantz, A., 2002. Utilizing graphical formats to convey uncertainty in a decision-making task. *Theoretical Issues in Ergonomics Science*, 3 (1), 1–25. doi:[10.1080/14639220110110324](https://doi.org/10.1080/14639220110110324)
- Fisher, P., 1993. Visualising uncertainty in soil maps by animation. *Cartographica*, 30, 20–27. doi:[10.3138/B204-32P4-263L-76W0](https://doi.org/10.3138/B204-32P4-263L-76W0)
- Fisher, P., 1994. Hearing the reliability in classified remotely sensed images. *Cartography and Geographic Information Systems*, 21 (1), 31–36. doi:[10.1559/152304094782563975](https://doi.org/10.1559/152304094782563975)
- Garvey, M., 2013. GIS for emergency management Australian trends. In: *Proceedings of the Pacific GIS and Remote Sensing Conference 2013*, 16–19 November 2013, Suva.
- Gehlenborg, N. and Wong, B., 2012. Heat maps. *Nature Methods*, 9 (3), 213.
- Hertwig, R. and Ortmann, A., 2001. Experimental practices in economics: a methodological challenge for psychologists?. *Behavioral and Brain Sciences*, 24, 383–451.
- Hertwig, R. and Ortmann, A., 2003. Economists' and psychologists' experimental practices: how they differ, why they differ and how they could converge. In: I. Brocas and J. Carrillo, eds. *Economics and psychology*. New York: Oxford University Press, 253–272.
- Holt, C. and Laury, S. Risk aversion and incentive effects. Available from: <http://ssrn.com/abstract=893797> 2002.
- Hope, S. and Hunter, G., 2007a. Testing the effects of positional uncertainty on spatial decision-making. *International Journal of Geographic*, 21 (6), 645–665. doi:[10.1080/13658810601073273](https://doi.org/10.1080/13658810601073273)
- Hope, S. and Hunter, G., 2007b. Testing the effects of thematic uncertainty on spatial decision-making. *Cartography and Geographic Information Science*, 34 (3), 199–214. doi:[10.1559/152304007781697884](https://doi.org/10.1559/152304007781697884)
- Howard, D. and MacEachren, A., 1996. Interface design for geographic visualization: tools for representing reliability. *Cartography and Geographic Information Systems*, 23 (2), 59–77. doi:[10.1559/152304096782562109](https://doi.org/10.1559/152304096782562109)
- Kinkeldey, C., MacEachren, A., and Schiewe, J., 2014. How to assess visual communication of uncertainty? A systematic review of geospatial uncertainty visualisation user studies. *The Cartographic Journal*, 51 (4), 372–386. doi:[10.1179/1743277414Y.0000000099](https://doi.org/10.1179/1743277414Y.0000000099)
- Kirschenbaum, S. and Arruda, J., 1994. Effects of graphic and verbal probability information on command decision making. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 36 (3), 406–418.
- Kirschenbaum, S., et al., 2013. Visualizing uncertainty: the impact on performance. *Human Factors*, 56 (3), 509–520. doi:[10.1177/0018720813498093](https://doi.org/10.1177/0018720813498093)
- Klimenko, S., et al., 2007. Using virtual environment systems during the emergency prevention, preparedness, response and recovery phases. In: H.J. Pasman and I.A. Kirillov, eds. *Resilience of cities to terrorists and other threats: learning from 9/11*. Dordrecht: Springer.
- Krygier, J., 1994. Sound and geographic visualization. In: A. MacEachren and D. Taylor, eds. *Visualization in modern cartography*. Oxford: Pergamon, 1–12.
- Kunz, M., Grêt-Regamey, A., and Hurni, L., 2011. Visualization of uncertainty in natural hazards assessments using an interactive cartographic information system. *Natural Hazards*, 59, 1735–1751. doi:[10.1007/s11069-011-9864-y](https://doi.org/10.1007/s11069-011-9864-y)
- Leitner, M. and Battenfield, B.P., 2000. Guidelines for the display of attribute certainty. *Cartography and Geographic Information Science*, 27 (1), 3–14. doi:[10.1559/152304000783548037](https://doi.org/10.1559/152304000783548037)
- Light, A. and Bartlein, P., 2004. The end of the rainbow? Color schemes for improved data graphics. *EOS, Transactions, American Geophysical Union (EOS)*, 85 (40), 385–391. doi:[10.1029/2004EO400002](https://doi.org/10.1029/2004EO400002)

- Lodha, S., Wilson, C., and Sheehan, R., 1996. LISTEN: Sounding uncertainty visualization. In: R. Yagel and G. Nielson, eds. *Proceedings of the 7th conference on visualization* 96. Los Alamitos, CA: IEEE Computer Society Press, 189–195.
- Lowenstein, G. and Adler, D., 1995. A bias in the prediction of tastes. *The Economic Journal*, 105 (431), 929–937. doi:[10.2307/2235159](https://doi.org/10.2307/2235159)
- Lucieer, A. and Kraak, M.-J., 2004. Interactive and visual fuzzy classification of remotely sensed imagery for exploration of uncertainty. *International Journal of Geographical Information Science*, 18 (5), 491–512. doi:[10.1080/13658810410001658094](https://doi.org/10.1080/13658810410001658094)
- MacEachren, A., et al., 2012. Visual semiotics & uncertainty visualization: an empirical study. *IEEE Transactions on Visualization and Computer Graphics*, 18 (fall), 2496–2505. doi:[10.1109/TVCG.2012.279](https://doi.org/10.1109/TVCG.2012.279)
- MacEachren, A.M., 1992. Visualizing uncertain information. *Cartographic Perspectives*, 13, 10–19. doi:[10.14714/CP13.1000](https://doi.org/10.14714/CP13.1000)
- MacEachren, A.M., Brewer, C.A., and Pickle, L.W., 1998. Visualizing georeferenced data: representing reliability of health statistics. *Environment and Planning A*, 30 (9), 1547–1561. doi:[10.1068/a301547](https://doi.org/10.1068/a301547)
- Mason, J., Retchless, D., and Klippel, A., 2014. Dimensions of uncertainty: A visual classification of geospatial uncertainty research. In: *Workshop Visually-Supported Reasoning with Uncertainty, in conjunction with GIScience 2014* [online]. Vienna. Available from: <http://cognitivegis.science.psu.edu/uncertainty2014/Schedule.html> [Accessed 1 November 2014].
- McGranghan, M., 1993. A cartographic view of spatial data quality. *Cartographica*, 30 (2&3), 8–19. doi:[10.3138/310V-0067-7570-6566](https://doi.org/10.3138/310V-0067-7570-6566)
- Monmonier, M., 1996. *How to lie with maps*. Chicago: University of Chicago Press.
- Monmonier, M. and Gluck, M., 1994. Focus groups for design improvement in dynamic cartography. *Cartography and Geographic Information Systems*, 21 (1), 37–47. doi:[10.1559/152304094782563948](https://doi.org/10.1559/152304094782563948)
- Pang, A., 2001. Visualizing uncertainty in geo-spatial data. In: *Proceedings of the Workshop on the Intersections between Geospatial Information and Information Technology*. Washington, DC: National Academies Committee of the Computer Science and Telecommunications Board.
- Pang, A., Wittenbrink, C., and Lodha, S., 1997. Approaches to uncertainty visualization. *The Visual Computer*, 13, 370–390. doi:[10.1007/s003710050111](https://doi.org/10.1007/s003710050111)
- Quispel, A. and Maes, A., 2014. Would you prefer pie or cupcakes? Preferences for data visualization designs of professionals and laypeople in graphic design. *Journal of Visual Languages & Computing*, 25, 107–116. doi:[10.1016/j.jvlc.2013.11.007](https://doi.org/10.1016/j.jvlc.2013.11.007)
- Riviero, M., et al., 2014. Effects of visualizing uncertainty on decision-making in a target identification scenario. *Computers & Graphics*, 41, 84–98. doi:[10.1016/j.cag.2014.02.006](https://doi.org/10.1016/j.cag.2014.02.006)
- Schweizer, D. and Goodchild, M., 1992. Data quality and choropleth maps: an experiment with the use of color. In: *Proceedings of GIS/LIS '92*. Washington, DC: ACSM AND ASPRS, pp. 686–689.
- Severtson, D. and Myers, J., 2013. The influence of uncertain map features on risk beliefs and perceived ambiguity for maps of modeled cancer risk from air pollution. *Risk Analysis*, 33 (5), 818–837. doi:[10.1111/j.1539-6924.2012.01893.x](https://doi.org/10.1111/j.1539-6924.2012.01893.x)
- Shattuck, L., Miller, N., and Kemmerer, K., 2009. Tactical decision making under conditions of uncertainty: an empirical study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 53 (4), 242–246. doi:[10.1177/154193120905300417](https://doi.org/10.1177/154193120905300417)
- Smith, V., 1976. Experimental economics: induced value theory. *American Economic Review*, 66 (2), 274–279.
- Smith, V. and Walker, J., 1993. Monetary rewards and decision cost in experimental economics. *Economic Inquiry*, 31, 245–261. doi:[10.1111/ecin.1993.31.issue-2](https://doi.org/10.1111/ecin.1993.31.issue-2)
- St John, M., Callan, J., and Proctor, S., 2000. *Tactical decision-making under uncertainty: Experiments I and II. Technical report*. San Diego, CA: SPAWAR Space and Naval Warfare Systems Center.
- Tolhurst, K., Shields, B., and Chong, D., 2008. Phoenix: development and application of a bushfire risk management tool. *The Australian Journal of Emergency Management*, 23, 47–58.
- USDA Forest Service. 2015. WFDSS help. Available from: http://wfdss.usgs.gov/wfdss_help/toc.htm.

- Van Der Wel, F., Hootsmans, R., and Ormeling, F., 1994. Visualization of data quality. In: A. MacEachren and D. Fraser-Taylor, Eds. *Visualization in modern cartography*. New York, NY: Elsevier Science Ltd., 313–331.
- Viard, T., Caumon, G., and Levy, B., 2011. Adjacent versus coincident representations of geospatial uncertainty: which promote better decisions? *Computers and Geosciences*, 37 (4), 511–520. doi:10.1016/j.cageo.2010.08.004
- Ware, C., 2008. *Visual thinking for design*. Burlington, MA: Morgan Kaufmann Publishers.
- Wilkening, J., 2009, User preferences for map-based decision making under time pressure. In: C. Davies (ed.), *Proceedings of the Doctoral Colloquium at the 9th International Conference on Spatial Information Theory (COSIT)*, 25 September L'Aber Wrac'h. 91–97.
- Zlatanova, S. and Li, J., 2008. *Geospatial information technology for emergency response*. Boca Raton, FL: CRC Press.

Appendices

Appendix A

List of sentences and true/false questions used for the dual-task experiment adapted from Fedorenko *et al.* (2007).

- Sentence 1: The detective who recognized the spy crossed the street at the light.
- Question 1a: The detective crossed the street at the light.
- Question 1b: The spy recognized the detective.
- Sentence 2: The waitress who hugged the bartender dropped the tray on the floor.
- Question 2a: The waitress hugged the bartender.
- Question 2b: The waitress dropped the tray.
- Sentence 3: The acrobat who mocked the clown performed the trick at the show.
- Question 3a: The clown mocked the acrobat.
- Question 3b: The clown performed the trick.
- Sentence 4: The lecturer who provoked the dean left the university in the summer.
- Question 4a: The dean left the university in the summer.
- Question 4b: The dean provoked the lecturer.
- Sentence 5: The singer who blamed the organizer cancelled the concert in Los Angeles.
- Question 5a: The singer cancelled the concert.
- Question 5b: The orchestra blamed the singer.
- Sentence 6: The biker who the driver ignored made the turn at the crossing.
- Question 6a: The driver made the turn at the crossing.
- Question 6b: The biker ignored the driver.

Appendix B

Screenshots of the experiment interface.

Please enter your experiment number:

Please click the following button to confirm that your experiment number worked and the instruction text below has loaded.

Please read the following information carefully.

Overview
During the course of this experiment you will be shown a series of images. The images show maps, or in some cases text, that depict the likelihood that a bushfire will destroy your house. For each image, you will be asked to decide whether to leave or stay in the house. If you decide to stay when the house is not burned, or leave in cases when the house is burned, you will be rewarded with additional payments.

Activity
For each image you are asked to choose whether to **stay** or **leave** by pressing the corresponding button. You can choose only one of the options. Once you have opted to stay or to leave, you cannot change your decision.

Time
You will have up to **30 seconds** per image to make your decision. In most cases this should be ample time to decide. The time remaining for an image is shown at the top of the screen. If for any reason you do not reach a decision before the time reaches zero, then you will be deemed to have decided to stay. In total you will be shown approximately **90 images**, and should expect the experiment to take between **20 and 30 minutes** to complete.

Payment Schedule
You will be paid a **base rate of \$7.00** for participating in the experiment. You can earn up to an **additional \$9.00** depending on your decisions. Each decision to leave when the house is burned or decision to stay when the house is not burned will earn you an additional experimental payment of **\$0.10**. You will not earn any additional payment for other decisions. A summary of the payments is shown below, and above every image in the experiment.

House	Leave	Stay
Burnt	\$ 0.10	\$ 0
Not Burnt	\$ 0	\$ 0.10

If at any point you do not wish to continue the experiment, please inform the experiment staff and you will be paid the base rate of \$7.00 for your participation.

Questionnaire
After completing the series of scenarios, you will be presented with a short questionnaire to complete.

The total accumulated amount you will be paid, including any additional payments, will appear on the screen after you have completed the questionnaire at the end of the experiment.

Start the experiment
Please familiarize your self with the example images below. On each of the maps, the location of your **house** is shown as a **black "X"** and the **map legend** is shown on the right hand side of the image.
When you are ready, please click the 'Start' button at the bottom of this page. Thank you for your participation in this experiment.

Figure A1. Introduction screen for the primary experiment.



Figure A2. Experiment interface for the time pressure experiment showing payment matrix, count-down timer, and color value stimuli.



Figure A3. Experiment interface for the time pressure experiment showing payment matrix, count-down timer, and color hue stimuli.