What risks do people perceive in everyday life? A perspective gained from the experience sampling method (ESM)

Robin M. Hogarth,[a*] Mariona Portell,[b] & Anna Cuxart[c]

[a]ICREA & Universitat Pompeu Fabra, Barcelona

[b]Universitat Autònoma de Barcelona

[c]Universitat Pompeu Fabra, Barcelona

January 12, 2007

For correspondence, please contact Robin M. Hogarth at Universitat Pompeu Fabra, Department of Economics and Business, Ramon Trias Fargas 25-27, 08005, Barcelona, Spain. (Tel: +34 93 542 2561, Fax + 34 93 542 1746). Email: robin.hogarth@upf.edu, mariona.portell@uab.es, anna.cuxart@upf.edu

Abstract

The *experiential sampling method* (ESM) was used to collect data from 74 part-time students who described and assessed the risks involved in their current activities when interrupted at random moments by text messages. The major categories of perceived risk were short-term in nature and involved "loss of time or materials" related to work and "physical damage" (e.g., from transportation). Using techniques of multi-level analysis, we demonstrate effects of gender, emotional state, and types of risk on assessments of risk. Specifically, females do not differ from males in assessing the potential severity of risks but they see these as more likely to occur. Also, participants assessed risks to be lower when in more positive self-reported emotional states. We further demonstrate the potential of ESM by showing that risk assessments associated with current actions exceed those made retrospectively. We conclude by noting advantages and disadvantages of ESM for collecting data about risk perceptions.

Keywords: Experiential sampling method; risk perception; risk assessment; gender differences; multi-level analysis; simultaneous vs. retrospective judgment.

Risk perception, and the role of risk in decision making, has received considerable attention in both theoretical and applied work over the last five decades (see, e.g., Fischhoff, Bostrom, & Quadrell, 2002; Sjöberg, 2000; Slovic, 2000). Empirically, four major methodologies can be distinguished. In one, researchers use surveys (questionnaires) to elicit people's attitudes toward specified risks (see, e.g., Slovic, Fischhoff, & Lichtenstein, 1979). The second is the use of choices between experimental gambles (see, e.g., Kahneman & Tversky, 1979). In the third, people are asked to state how much they are willing to pay to avoid certain risks (see, e.g., Hammitt & Graham, 1999). And in the fourth, researchers infer people's attitudes to risk by observing their choices in *specific* situations, e.g., the decision to smoke (Viscusi, 1992) or to drive without using seat-belts (Richens, Imrie, & Copas, 2000).[1]

One conclusion from these studies is that risk – and the perception of risk – is not easily characterized. It is unclear, for example, what people understand by the term "risk" and assessments are subject to many individual and situational factors. In this paper, we ask a question that is not addressed by the methodologies enumerated above, namely: What risks do people perceive and assess in their daily activities?[2] However, answering this question has two requisites. One is that people should define the risks they encounter. The second is to ensure that these risks are representative of their daily activities.

The methodology we employ is inspired by Brunswik's concept of *representative* design (Brunswik, 1956) and is aimed at sampling people's behaviors (in our case perceptions of risk) as they go about their daily lives. In an early study, Brunswik (1944) illustrated the method by investigating a person's ability to estimate distances and the sizes of objects in her natural environment (the University of

---

[1] This kind of analysis has also been done at the societal level (see, e.g., Tengs & Graham, 1996).
[2] We are not, of course, the first investigators to ask this question. See, e.g., Fischer et al. (1991).

California at Berkeley). Over a period of four weeks, the person (a student) was followed by a researcher and instructed to behave in her normal fashion. However, at irregular – or random moments – the participant was asked to estimate the sizes of objects that happened to be in her visual field as well as the distances from the objects. The researcher then checked the accuracy of the estimates. It is important to note that the researcher did not choose the specific "experimental tasks" (i.e., the objects in the participant's visual field), but instead defined a *process* by which these were a random sample of the participant's experience. Thus, although the study was conducted with a single participant, the random sampling process used was sufficient to make generalizable statements about this person's ability to judge sizes and distances.

In 1944, Brunswik's methodology clearly required much labor to make inferences about a single person. However, recent years have seen the growth of what is now referred to as the *Experience Sampling Method* (ESM) using modern technology (e.g., beepers, palmtop computers) to prompt participants to respond to questionnaires at random moments (see, e.g., Csikszentmihalyi & Larson, 1987; Hurlburt, 1997; Bolger, Davis, & Rafaeli, 2003). For example, in a recent study using cell telephones, Hogarth (2006) sampled the decision making behaviors of undergraduate students and business executives. Specifically, the goal was to investigate how much feedback people receive and/or expect to receive on the many decisions they make each day. By sampling individual participants up to four times a day for up to two weeks, random samples of the participants' decisions were obtained such that estimates could be made of the feedback they received or expected to receive. This, in turn, allowed inferences to be made about the characteristics of decision making tasks that people face in their everyday lives. Incidentally, although not a major study, some 1,200 data points (i.e.,

decisions) were obtained from 34 participants such that good estimates could be obtained at both individual and group levels.

It is important to emphasize the kinds of inferences that can and cannot be drawn from ESM studies. At the individual level, the method can be used to make inferences about a single person's behavior (e.g., judgments and/or situations faced). That is, the population of behavior to which one can generalize is limited to the particular individual. However, the extent to which the individual is representative of a specific population depends on how the individual was sampled from that population. In many cases, individuals belong to so-called "convenience samples" (i.e., participants who could be recruited for studies) and thus are not necessarily representative of specific populations. Nonetheless, one can still aggregate individual responses and make inferences at the group level albeit without claims for generalization.

In the study reported here, 74 part-time students answered questions about the risks they perceived themselves facing when they received text messages from the investigators. The participants each received three such messages at random moments during 10 consecutive working days. The main goal of the study was to understand better the kinds of risk that people experience in their daily lives. However, we were also intrigued by other questions: Do emotional state and gender moderate assessments, and how? Can overall judgments of risk be decomposed into judgments of the severity of the negative consequences, on the one hand, and the possibility of such events occurring, on the other hand? Do simultaneous risk assessments (i.e., made at the time the risk is experienced) differ from retrospective judgments (i.e., when recalled after the event)?

In the next section of this paper, we describe our methodology. This is followed by presenting the results. In short, participants' perceptions of risks are mainly short-

term and concrete. At the same time, they assess surprisingly high estimates of the possibility that the worst consequences of their current actions will be realized. We find that females provide, on average, higher assessments of risks than males. However, there is no difference in assessments of potential severity of risks. Rather, females judge that negative consequences are more likely to occur. When self-reported emotional states are better, assessments of risks and their likelihood of occurrence are lower. We explore possible differences between simultaneous and retrospective estimates of risk and find that the former exceed the latter. We conclude by discussing advantages and disadvantages of using ESM to illuminate issues of risk perception.

Method

Participants. Seventy-four students (46 women and 28 men) were recruited from the Universitat Autònoma de Barcelona. They ranged in age between 18 and 56 (median 22). A condition of their participation was that they had part-time jobs (defined by at least half of full working days). They were paid 30 € for their participation that, in addition to responding to the questions detailed below, required attendance at sessions before and after the experiment for instructions and debriefing.

Procedures. We sent text messages to participants between 8 am and 10 pm over a two-week period that excluded week-ends, i.e., for 10 days. Depending on their working hours, some participants received their messages between 8 am and 3 pm and the others between 3 pm and 10 pm (31 and 43 participants, respectively).[3] To determine when messages should be sent, we divided time into segments of 15 minutes and chose six segments at random each day (three for each group of participants).

---

[3] The objective was to send participants messages during the part of the day in which they were mainly at work.

When they received a message, participants were required to note the date and time and to answer eight questions.[4] Two of these questions were open-ended; the others required responses on scales. The questions, their scales, and the abbreviations we use, were:

1. *How would you evaluate your emotional state right now?* Scale from 1(very negative) to 10 (very positive): emotional state.

2. *What are you doing right now?* Open-ended and subsequently referred to as ACT: activity.

3. *Is ACT professional or personal in nature?* Binary response: type.

4. *How often do you do ACT?* Scale from 1 (first time) to 10 (many times): frequency.

5. *What is the WORST consequence that could result from ACT?* Open-ended: worst consequence.

6. *How do you rate the severity of the WORST consequence that could result from ACT?* Scale from 1 (very low) to 10 (very high): severity.

7. *At this moment, what is the chance that the WORST consequence of ACT occurs?* Scale from 0 (impossible) to 100 (certain): possibility.

8. *At this moment, what risk for your well-being do you associate with ACT?* Scale from 0 (very low) to 100 (very high): risk.

After completing the task, participants were thanked, debriefed, and paid in a session in which they also completed some questionnaires including one in which they re-assessed the risks associated with two activities they had reported in the preceding weeks.

---

[4] All questions were asked in Spanish.

Results

Response rates. Overall, we sent out 2,220 messages (= 3 x 10 x 74) and received 2,213 responses (99.7%).[5] For various reasons, people might not receive text messages when they are sent (e.g., cell telephones may have been turned off). We therefore checked the extent to which messages were received when they were sent. Participants reported receiving messages between 0 and 22 minutes after they were sent with an overall mean (median) of 3 (2) minutes. We deem both response rates and reported times of receiving messages quite satisfactory.

Overall trends. The goal of Table 1 is to provide an overview of the data. It is a cross-tabulation of activities (rows) by types of losses (columns) and was constructed from a content analysis of the two open-ended responses to the questionnaires.[6] To perform this analysis, definitions of the categories for activities and types of losses were first constructed. Then, two researchers independently allocated responses to categories. Disagreements between the two coders were resolved by having them discuss until they reached consensus. In Appendix A, we provide definitions of the categories used to classify activities and worst consequences as well as illustrative examples.

Insert Table 1 about here

Overall, the 74 participants provided information about 2,213 activities and types of loss. Two-thirds of all activities involved participants' basic occupation (34%), personal transportation (11%), eating and drinking (10%), and entertainment and recreation (10%). Most types of loss can be categorized as loss of time or materials (42%) or physical damage (37%). The largest single joint entry in the table (445) is for loss of time or materials associated with participants' basic occupation. This accounts

---

[5] We used commercially available software to dispatch the messages to the 74 participants.
[6] Here, as in other cases where qualitative responses were provided, we report the initial inter-coder agreement prior to the reaching of consensus.

7

for 20% of all the data. Within the basic occupation category, loss of time or materials accounts for 59% of all potential losses and this is more than twice the next largest category, physical damage (23%). The second largest joint entry is for potential physical damage arising from personal transportation (189 instances or 9% of the total). In fact, of all the activities, transportation is the one where type of loss is most associated with physical damage. Moreover, it is of interest to note the effects of the goal of transportation (personal or professional). Physical damage accounts for 79% of potential losses when transportation is personal, but 60% when this is professional.

Apart from transportation, the only other categories with a majority of losses associated with physical damage are the activities of eating and drinking and smoking. In Appendix B, we reproduce some of the responses given by participants. It is immediate rather than longer-term damage that predominates. In the case of eating and drinking, 57% of losses involve "choking/suffocating," 26% "indigestion," and 18 % "burns, cuts, or other injuries associated with the act of eating or drinking." In the case of smoking, there are many more mentions involving immediate and minor consequences (64% for "to getting burned") than to the more serious consequences that are stressed in anti-smoking advertisements (36% refer to cancer or respiratory problems). Given the limited number of observations, however, these results should be viewed with caution. Nonetheless, we believe they are noteworthy because the data were collected at a time when sensitivity to the dangers of smoking has reached an all-time high in Spain. They are also generally representative of many of the potential losses cited by our respondents in that these are dominated by short-term consequences of risky activities. We believe that these findings could be important in thinking about how the risks associated with different activities should be communicated.

Some 60% of all activities are perceived as being personal in nature. At the same time, 73% of participants perceive their activities as being more personal than professional. We draw attention to these facts because the data were collected during times when most participants were working. (For similar results, see Mullet, Ciutad, & Rivière-Shafighi, 2004). This suggests, therefore, that their occupations allowed much time for personally-initiated activities, a result that was also observed in Hogarth's (2006) ESM study of business executives and undergraduates.

Modeling the data using HLM. The design of our study involved data that can be thought of as being collected at two levels. One – termed level 1 – is represented by participants' responses to the 30 occasions on which they received SMS messages (i.e., at the level of events). The other – level 2 – is at that of the participants themselves (i.e., characteristics of the participants that do not change across the 30 events). Thus, for example, it is of interest to know whether emotional state at the moment judgments are elicited affects assessments of risk (i.e., at level 1) and also whether such judgments reflect differences between the participants in, say, gender (i.e., at level 2). As such, our data can be efficiently modeled using the techniques of hierarchical linear models (Bryk & Raudenbush, 2002; Goldstein, 1995; Longford, 1994).

In what follows, we first ask what variables (at both levels) explain responses for severity, possibility, and risk (treated as three separate dependent variables). We subsequently explore how best to model risk as a function of, inter alia, severity and possibility. But, before presenting results, we motivate our use of the hierarchical linear model technique with an example.

Assume we wish to model risk as being affected by emotional state (at level 1) but that this is moderated by gender (at level 2).

Define the model at level 1 as

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \overline{X}_{.j}) + r_{ij} \qquad\qquad (1)$$

where

Y_{ij} is the judgment of risk on the ith occasion (i = 1,....,30) for the jth individual (j = 1,..., 74);

$(X_{ij} - \overline{X}_{.j})$ is the deviation of the reported emotional state X_{ij} on the ith occasion for the jth individual from his or her average emotional state $\overline{X}_{.j}$;

β_{0j} is the individual-specific intercept;

β_{1j} is the individual-specific slope (regression coefficient) of Y_{ij} on $(X_{ij} - \overline{X}_{.j})$; and

r_{ij} is the error term which we assume normally distributed with constant variance, $r_{ij} \sim N(0, \sigma^2)$.

Define the model at level 2 as

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j} \qquad\qquad (2)$$

and

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j} \qquad\qquad (3)$$

where

Z_j = gender of participant (0, female, or 1, male);

γ_{00} is the constant part of the intercept β_{0j};

γ_{01} is the regression coefficient of β_{0j} on Z_j;

u_{0j} is an error term (the individual effect on judgments of risk);

γ_{10} is the constant part of the slope β_{1j};

γ_{11} is the effect of gender on the slope β_{1j} (i.e., the interaction of gender and emotional state on judgments of risk);

u_{1j} is an individual error term, that is, the random interaction effect of individual j and emotional state on judgments of risk.

Thus, to interpret the above, γ_{00} is the average risk score across all women, while $\gamma_{00}+\gamma_{01}$ is the average risk score across all men; γ_{10} is the average effect across women of emotional state on judgments of risk, while $\gamma_{10}+\gamma_{11}$ represents the average effect across men of emotional state on judgments of risk. We assume that u_{0j} and u_{1j} are random variables with zero means, variances τ_{00} and τ_{11}, respectively, and covariance τ_{01}; they represent the variability in β_{0j} and β_{1j} that remains after controlling for Z_{j}.

Substituting equations (2) and (3) into (1), we obtain

$$Y_{ij} = \gamma_{00} + \gamma_{01}Z_{j} + u_{0j} + (\gamma_{10} + \gamma_{11}Z_{j} + u_{1j})(X_{ij} - \overline{X}_{.j}) + r_{ij} \tag{4}$$

which can be re-arranged as

$$Y_{ij} = \gamma_{00} + \gamma_{01}Z_{j} + \gamma_{10}(X_{ij} - \overline{X}_{.j}) + \gamma_{11}Z_{j}(X_{ij} - \overline{X}_{.j}) + u_{0j} + u_{1j}(X_{ij} - \overline{X}_{.j}) + r_{ij} \tag{5}$$

Note that, in the latter expression, one can distinguish the fixed part, $\gamma_{00} + \gamma_{01}Z_{j} + \gamma_{10}(X_{ij} - \overline{X}_{.j}) + \gamma_{11}Z_{j}(X_{ij} - \overline{X}_{.j})$ (with main effect for Z, main effect for X and their interaction), from the random part, $u_{0j} + u_{1j}(X_{ij} - \overline{X}_{.j}) + r_{ij}$ (with individual random effect, random interaction between the jth individual and X, and occasion-specific error term).

We have only illustrated the model by considering one independent variable at level 1 (emotional state) and one independent variable at level 2 (gender). However, it is straightforward to construct more complete models by considering vectors of independent variables at both levels. Moreover, we can consider variations of the models where coefficients associated with different independent variables are random or fixed.

Severity, possibility, and risk. The main results of modeling judgments of severity, possibility, and risk are provided in Table 2.[7] In estimating the models, we used the maximum likelihood method and the HLM6 software.[8] We first comment on variables included in the analyses.

Insert Table 2 about here

Two sets of variables were included to check the assumption that we were sampling events randomly from our participants. At level 2, we included a dummy variable, "Moment of day," to capture whether participants belonged to the morning or afternoon groups (see above). At level 1, we also tested whether the position of the message in a day (i.e., first, second, or third) had an influence on responses. This possible "Sequence" effect was captured by two dummy variables ("Second in day," "Third in day").

Participants reported their emotional state on each occasion they received an SMS message. (This was the first of the eight questions they answered, see above.) We used these data in two ways. One was to calculate the mean of each person's emotional states to estimate overall individual tendencies (i.e., at level 2); the other was to calculate deviations from those means to capture responses to different situations (i.e., at level 1).

We used dummy variables to assess the effects of three types of "worst loss" relative to the category "none." These were "Physical damage," "Psychological/social conflict,"[9] and "Loss of time or materials." In coding responses, we had noted that situations in which risk was perceived varied as to whether people were alone or

---

[7] The dependent variables possibility and risk were rescaled from 0-100 to 0-10 to maximize comparability across all three dependent variables.
[8] The student version of this can be downloaded free from the internet, see http://www.ssicentral.com
[9] This represented the combination of the categories "Psychological discomfort" and "Social conflicts" in Table 1.

interacting with others. We therefore also classified situations as to whether they involved "Interactions" and with whom – "Friends or family" or "Work."

Several conclusions can be drawn from our analysis.

First, the three models all have independent variables that are significant at level 1; however, at level 2, the model for severity has no significant variables.

Second, neither "Moment of day" (level 2) nor the "Sequence" variables (level 1) are significant in any of our models. These results are consistent with the hypothesis that we are dealing with random samples of the participants' behavior.

Third, in all three models there is no significant effect for emotional state at level 2 but this is significant at level 1 for both possibility and risk. For these variables, being in a better mood implies making lower assessments. We emphasize that, in modeling emotional state at level 1, we tested and found that a random coefficient should be preferred to a fixed coefficient.[10] In other words, there was considerable individual heterogeneity in the relation between mood state, on the one hand, and judgments of possibility and risk, on the other hand.

Fourth, gender effects (level 2) mediated judgments of possibility and risk but not severity (females assessed possibility and risk higher than males).

Fifth, there were no effects due to whether participants classified risks as being private or professional in nature and frequency (i.e., how often particular risks are faced) had only a small, significant effect on severity.

Sixth, there were significant effects for types of risks. The mean judgments for severity and risk are highest for "Physical damage" and lowest for "Loss of time or materials" (excluding, of course, the category "none"). Interestingly, the negative

---

[10] The likelihood ratio test statistic for the comparison between the constrained model (fixed coefficient, meaning variance of the slope coefficient equal to 0) and the unconstrained model (random coefficient) is $\chi^2_{2df} = 12.5$ (p-value = 0.002) for the dependent variable possibility and $\chi^2_{2df} = 15.3$ (p-value < 0.001) for the dependent variable risk.

consequences associated with "Psychological/social conflict" are seen as being more likely (greater possibility) than the others.

Seventh, we note that assessments of possibility were, on average, quite high (e.g., the effect for physical damage was 2.57). Indeed, if one treats these assessments as probabilities, they suggest considerable overestimates (e.g., approximately one fourth of risks of physical damage clearly did not occur subsequent to when the risks were perceived). We discuss this result further below.

Before discussing the implication of these results, we consider modeling judgments of risk as being affected by the prior judgments of severity and possibility.

Modeling judgments of risk. How do judgments of severity and possibility contribute to assessed risk and what other variables explain risk when the contributions of severity and possibility have been taken into account?

Insert Table 3 about here

In Table 3, we present the results of four hierarchical linear models that progressively answer these questions. Model 1 is a "null model" that decomposes the variance without any explanatory variable, capturing the overall mean and the variance both between and within individuals. [11] Model 2 adds the independent effects of severity and possibility at level 1 (occasions). Here, the coefficients for both severity and possibility are significant and of the expected sign (i.e., larger assessments of risk are associated with greater severity and possibility). Model 3 also considers the interaction between severity and possibility. In this formulation, severity and the interaction are significant but possibility is not. In Model 4, we investigated variables

---

[11] The "null model" is the simplest hierarchical linear model and is equivalent to a one-way ANOVA with random effects. The estimation of the null model is a first step in a hierarchical analysis. It can be used to estimate the intraclass correlation coefficient (0.36 in Model 1) that measures the proportion of total variance corresponding to differences between level 2 units (individuals).

that had further significant effects. These are gender at level 2 (men assess risk lower than women), emotional state at level 1 (the better the emotional state, the lower the assessment of risk), and positive effects for both risks involving physical damage and interactions with friends and family (also level 1).

In short, our data are consistent with the idea that both severity and possibility contribute to assessments of risk but that these assessments are also influenced by gender, emotional state, and types of risk. Indeed, given the importance of both gender and emotional state in our findings, we display the fitted effects of both these variables on assessments of risk in Figure 1.

Insert Figure 1 about here

Retrospective versus simultaneous judgments of risk. In the post-experimental questionnaire (completed between 2 and 7 days after participants had provided their last experimental responses), we showed participants two activities (i.e., ACTs) they had reported over the two-week period, namely for the 4th and 27th messages they had received.[12] We then asked them to re-assess the levels of risk that were associated with each of these activities.

Insert Table 4 about here

Table 4 reports the mean risk judgments for the 4th and 27th messages made retrospectively (i.e., after the experiment) and simultaneously (i.e., during the experiment and thus when experiencing the risk). The differences (retrospective – simultaneous) of -13.8 and -10.7 are both large and statistically significant (t = - 4.86, p

---

<.001; and t = -3.65, p <.001 for the 4th and 27th messages, respectively).[13]   In other words, risks are judged to be higher at the time they are experienced than in hindsight.

Discussion

We discuss our results by considering the types of risks identified by our participants; differences between simultaneous and retrospective assessments of risk; and individual differences in risk assessment. In doing so, we enumerate advantages and disadvantages of using the ESM to study the perception and assessment of risk and also suggest avenues for further research.

The manner in which participants were asked to identify risks was by linking these explicitly to their current actions. Such assessments have the advantage of being based in the participants' reality. However, this can limit the types of risks reported. For example, imagine a worker who is using a potentially dangerous work tool but is pre-occupied by the risk of wasting valuable materials if a mistake is made. Because the latter risk is salient when interrogated, it is likely to be reported even though the consequences of the dangers inherent in using the tool are greater.   Similarly (and as noted above), our respondents who were smoking were far more likely to mention short-term risks such as being burned by a cigarette rather than longer-term and more serious risks to health.

On the other hand, from a policy perspective of risk prevention and communication, it is critical to understand the risks to which people are attending when involved in different activities, e.g., being burned by a cigarette as opposed to dying from lung cancer.

---

[13] These data can also be used to test for reliability in that the correlation between the 4th (27th) response assessed simultaneously and retrospectively was 0.45 (0.62), p < .01 for both.

These considerations suggest that the risks that people consciously perceive are a subset of the risks that they actually face in their everyday lives. Thus, many of the risks that are commonly studied in surveys were totally absent from our participants' responses. For example, whereas 21 of our 74 participants were traveling by train or waiting at a station when they received SMS messages, on only one occasion did anyone mention the risk of terrorism. And yet, one year prior to when the data were collected, many people were killed in an attack in Madrid while simply commuting to work. In future work, it would be of interest to investigate explicitly the risks that people do not mention.

A second, important finding was that assessments of risk made simultaneously (i.e., when responding to SMS messages) were greater than those made retrospectively (i.e., at the debriefing session after the study). There are clearly differences between the two judgments. One takes place in "the field" at the time when the risk is perceived; the other is assessed in laboratory conditions and involves memory. In addition, we suspect that the former is more emotionally charged than the latter (cf., Epstein, 1994) such that the direction of the difference in assessed risk should not be too surprising. At the same time, the finding suggests that surveys of perceived risk (conducted in laboratory-like conditions) are likely to underestimate the levels of risk that people experience in their daily lives. This issue clearly requires further research.

A related issue concerns our observation that participants' assessments of the "possibility" that they would suffer the worst consequences associated with their current actions seemed unrealistically high (roughly 30%). It is not clear to us, however, that the participants understood the possibility scale as a measure of probability and so we are cautious in how we interpret this result. At the same time, it does suggest another ESM study in which participants are provided with an explicit probability scale and/or

verbal probabilistic statements and where, after the study, they report on how many of the worst consequences actually occurred.[14] More generally, it suggests that ESM could be profitably used to study issues of calibration in the assessment of subjective probabilities as well as possible hindsight biases (cf., Fischhoff, 1975). In this way one would be sampling events that are representative of the participants' lives (cf., Brunswik, 1944) in a much more satisfactory manner than laboratory studies that attempt to achieve the same goal (Gigerenzer, Hoffrage, & Kleinbölting, 1991; Klayman et al., 1999).

The methodology of ESM clearly allows one to draw inferences at the individual level; however, it presents a greater challenge in finding ways to represent the aggregate-level data (cf., Bolger, Davis, & Rafaeli, 2003). In presenting Table 1, for example, we attempted to provide aggregate information but recognize that these fail to capture the richness of the data. On the other hand, the hierarchical linear models we used provided a tidy manner of capturing systematic variance in the data at the levels of both individuals and occasions.

Our findings on gender differences are consistent with many findings in the literature that suggest that women perceive higher risks than men in relation to specific activities such as smoking and air travel (Slovic, 1992; Flynn, Slovic, & Mertz, 1994; Jungermann, Pfister, & Fischer, 1996) as well as in domains such as environmental and nuclear power risks (Brun, 1994; Greenberg & Schneider, 1995). More generally, Flynn et al. (1994) found that white women perceive risk to be higher than white men (in the US).[15] A recent study by Harris, Jenkins, and Glaser (2006) presents results similar to our own in that the source of higher risk assessments for negative events by

---

[14] Of course, one would have to exclude from analysis events that could not yet have taken place. However, most risks on which participants report are short-term in nature. Thus, this should not unduly restrict the number of possible observations.

[15] See also the meta-analysis of gender differences by Byrnes, Miller, and Schafer (1999).

women is linked to larger assessments of probabilities as opposed to the severity of risks *per se*.

A striking feature of our data was that, despite using only a simple self-reported measure of mood (emotional state), we obtained significant effects on judgments of both risk and possibility. This finding therefore adds to the growing literature that attests to the importance of emotions in the assessment of risk (see, e.g., Loewenstein et al., 2001; Slovic et al., 2002). We emphasize four points. First, we found no overall mood effect at the level of individuals (i.e., using the average of each individual's emotional state judgments as a dispositional measure) but this was significant at the level of situations that individuals faced (i.e., level 1). In other words, the effect of emotional state in our data appeared to reflect the situations faced as opposed to personal dispositions.[16] Second, the effect was better modeled by a random as opposed to fixed coefficient thereby attesting to considerable individual heterogeneity. Third, similar to the result concerning gender, emotional state would appear to affect judgments of risk by impacting the imagined possibility of the occurrence of negative events as opposed to assessments of potential severity of losses.

Fourth, given that we obtained these results using a simple self-report measure, it is intriguing to imagine what might be obtained using more sophisticated measuring devices. Indeed, in a not too distant future, one might imagine that technology for ESM studies could also include physical monitoring devices capable of detecting galvanic skin responses and perhaps even neurological measures.

To the best of our knowledge, this is the first study that has used the ESM to study the perception and assessment of risk. Like all other methodologies, it is not without its own problems. For example, there may be a bias in the particular risks that

---

[16] On the other hand, we note that some studies that did not use the ESM have reported effects of some personality variables on assessments of risk (Bouyer et al., 2001).

participants report although this, in itself, raises the interesting issue of identifying the risks of which people are not consciously aware. A second bias may be that the methodology (i.e., explicit questioning) directs participants' attention to specific types of risks. On the other hand, the methodology has the enormous advantage of being able to collect random samples of people's perceptions of the risks they face in everyday life, and – as demonstrated in this study – the ubiquity of cell telephones provides researchers with the possibility of collecting data *in situ* at relatively modest cost. Moreover, assessments associated with these perceptions probably differ from results collected through the more usual survey methods. This therefore raises ambiguities in interpreting results of more conventional studies of risk perception and questions what risk perception means when not elicited in the context of actually taking decisions. The methodology used also means that the perception of risk can be described at both individual and aggregate levels.

Finally, we are impressed by the potential of ESM for studying the perception and assessment of risk. As any single methodology, it is clearly incomplete relative to the complexities of the means. However, it suggests novel ways of studying important issues and therefore can and should be used as a means to complement other methodologies.

References

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, 54, 579-616.

Bouyer, M., Bagdassarian, S., Chaabanne, S., & Mullet, E. (2001). Personality correlates of risk perception. *Risk Analysis*, 21(3), 457-465.

Brun, W. (1994). Risk perception: Main issues, approaches and findings. In G. Wright and P. Ayton (Eds.), *Subjective probability* (pp. 395-420). Chichester: John Wiley and Sons.

Brunswik, E. (1944). Distal focusing of perception: Size constancy in a representative sample of situations. *Psychological Monographs*, 56 (254), 1-49.

Brunswik, E. (1956). *Perception and the representative design of experiments (2d ed.).* Berkeley, CA: University of California Press.

Byrk, A. S., & Raudenbush, S. W. (2002). *Hierarchical Linear Models: Applications and data analysis methods*. 2nd ed. Newbury Park: Sage Publications.

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin*, 125 (3), 367-383.

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *Journal of Nervous and Mental Disease*, 175, 526-536.

Epstein, S. (1996). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, 49, 709-724.

Fischer, G. W., Morgan, M. G., Fischhoff, B., Nair, I., & Lave, L.B. (1991). What risks are people concerned about? *Risk Analysis*, 11 (2), 303-314.

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance,* 1, 288-299.

Fischhoff, B., Bostrom, A., & Quadrel, M. J. (2002). Risk perception and communication. In R. Detels, J. Mc Ewen, & G. Ommen (Eds.), *Oxford textbook of public health* (pp. 1105-1123). London: Oxford University Press.

Flynn, J., Slovic, P., & Mertz, C. K. (1994). Gender, race and perception of environmental risks. *Risk Analysis*, 14 (6), 1101-1108.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, 98 (4), 506-528.

Goldstein, H. (1995). *Multilevel Statistical Models*. 2nd ed. Kendall's Library of Statistics 3. London: Edward Arnold.

Greenberg, M., & Schneider, D. (1995). Gender differences in risk perception: Effects differ in stressed vs. non-stressed environments. *Risk Analysis*, 15 (4), 503-511.

Hammitt, J. K., & Graham, J. D. (1999). Willingness to pay for health protection: Inadequate sensitivity to probability? *Journal of Risk and Uncertainty*, 18 (1), 33-62.

Harris, C. R., Jenkins, M., & Glaser, D. (2006). Gender differences in risk assessment: Why do women take fewer risks than men? *Judgment and Decision Making*, 1 (1), 48-63.

Hogarth, R. M. (2006). Is confidence in decisions related to feedback? Evidence – and lack of evidence – from random samples of real-world behavior. In K. Fiedler & P. Juslin (eds.), *In the beginning there is a sample: Information sampling as a key to understand adaptive cognition*. Cambridge, UK: Cambridge University Press.

Hurlburt, R. T. (1997). Randomly sampling thinking in the natural environment. *Journal of Consulting and Clinical Psychology*, 67 (6), 941-949.

Jungermann, H., Pfister, H.R., & Fischer, K. (1996). Credibility, information preferences, and information interests. *Risk Analysis,* 16, 251-261

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. E*conometrica*, 47 (2), 263-292.

Klayman, J., Soll, J. B., González-Vallejo, & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Performanc*e, 79 (3), 216-247.

Loewenstein, G.F., Weber, E. U., Hsee, C. K., Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, 127(2), 267-286.

Longford, N. T. (1994). "Random Coefficient Models", en Arminger y Sobel, eds., *Handbook of Statistical Modelling for the Social and Behavioural Sciences*. New York: Plenum Press.

Mullet, E., Ciutad, N., & Rivière-Shafighi, S. (2004). Cognitive processes involved in the assessment of health hazards' severity. *Health, Risk & Society*, 6(3), 277-288.

Richens, J., Imrie, J. & Copas, A. (2000). Condoms and seat belts: The parallels and the lessons. *Lancet*, 355, 400-403.

Sjöberg L. (2000). Factors in risk perception. *Risk Analysis,* 20 (1), 1-11.

Slovic, P. (1992). Perception of Risk: Reflections on the Psychometric Paradigm. In S. Krimsky & D. Golding (Eds.), *Social theories of risk* (pp. 117-152). New York: Praeger.

Slovic, P. (2000). *The perception of risk.* London: Earthscan Publications Ltd.

Slovic, P., Finucane, M., Peters, E., & MacGregor, D. (2002). The affect heuristic. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.). *Intuitive judgment: Heuristics and biase*s (pp. 397-420). New York, NY: Cambridge University Press.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1979). Rating the risks. *Environment*, <u>21</u> (3), 14-20.

Tengs, T. O., & Graham, J. D. (1996). The opportunity costs of haphazard social investments in life-saving. In R. Hahn (Ed.), *Risks, costs, and lives saved: Getting better results from regulation* (pp. 167-182). New York, NY: Oxford University Press.

Viscusi, W. K. (1992). *Making the risky decision*. New York, NY: Oxford University Press.

| Type of loss[1]: | | | Loss of time or materials | Physical damage | Psychological discomfort | Social conflicts | None | Other |
| | | % | 42 | 37 | 12 | 5 | 4 | 0 |
Activity[2]:	%	Total						
Basic occupation	34	751	445	171	57	58	20	0
Transportation -- personal	11	240	38	189	11	0	2	0
Eating and drinking	10	224	45	147	18	2	10	2
Entertainment, recreation, sports	10	215	73	78	51	2	10	1
Acquiring information	7	151	68	21	49	3	10	0
Housework, personal time, and managing funds	5	117	55	54	2	0	6	0
Communication -- personal	5	112	41	14	15	36	6	0
Personal care (e.g., hygiene, dressing)	5	102	38	40	21	1	2	0
Rest, sleep	3	76	25	19	29	1	2	0
Transportation -- professional	3	72	25	43	4	0	0	0
Pauses at start/end of working day	3	68	30	20	6	5	6	1
Shopping	1	30	20	3	3	0	4	0
Waiting	1	29	24	4	1	0	0	0
Smoking	1	15	2	11	0	2	0	0
Other	0	11	4	5	2	0	0	0
Total		2,213	933	819	269	110	78	4

[1]Initial inter-coder agreement. Kappa= 0.65
[2]Initial inter-coder agreement. Kappa= 0.83

	Severity		Possibility		Risk	
	Coefficient	t-ratio	Coefficient	t-ratio	Coefficient	t-ratio
Fixed effects:						
Level 2 variables						
Intercept	0.01	0.01	-0.51	-0.40	-0.20	-0.13
Mean emotional state	0.27	1.76	0.21	1.15	0.18	0.79
Gender (1= male; 0 = female)	0.08	0.30	-0.81	-2.47	-0.94	-2.27
Moment of day	-0.31	-1.11	0.03	0.08	-0.04	-0.10
Level 1 variables						
Emotional state	-0.03	-0.88	**-0.20**	-4.91	**-0.16**	-3.68
Private/professional	-0.10	-0.82	0.07	-0.68	-0.07	-0.65
Frequency	0.07	3.24	-0.01	-0.74	0.00	0.07
Type of worst loss:						
Physical damage	**5.00**	13.05	**2.57**	7.30	**2.96**	7.99
Psychological/social conflict	**3.48**	8.71	**3.35**	9.14	**2.07**	5.36
Loss of time or materials	**3.24**	8.28	**2.52**	7.00	**1.62**	4.28
Interactions:						
Friends and family	0.22	0.92	-0.07	-0.32	0.58	2.63
Work	0.20	0.91	0.44	2.29	0.46	2.26
Sequence						
Second in day	-0.01	-0.07	0.11	0.96	0.18	1.57
Third in day	-0.12	-0.99	-0.04	-0.34	0.09	0.74
Random effects:						
Variance components						
Level 2 (individuals)						
Intercept	**1.141**		**1.683**		**2.775**	
Slope (emotional state)			**0.043**		**0.054**	
Level 1 (occasions)	5.487		4.298		4.621	

Notes: (1) Coefficients/variance components significant at p<.001 are in **bold**, significant at p <.05 are underlined using
t-tests or X^2 as appropriate.
(2) Degrees of freedom for level 2 variables and random coefficients are at least 70. For all other variables they are
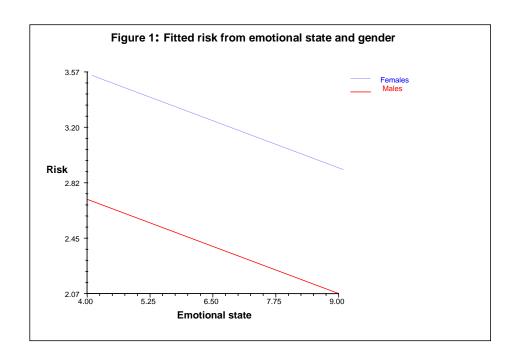at least 2162.

Table 3: Results of HLM models for risk

	Model 1		Model 2		Model 3		Model 4	
	Coefficient	t-ratio	Coefficient	t-ratio	Coefficient	t-ratio	Coefficient	t-ratio
Fixed effects:								
Level 2 variables								
Intercept	**2.88**	14.04	**-0.66**	-3.41	0.29	1.42	<u>0.52</u>	2.20
Gender (1= male; 0 = female)							<u>-0.66</u>	-2.10
Level 1 variables								
Severity			**0.43**	26.71	**0.26**	11.88	**0.22**	10.05
Possibility			**0.33**	17.75	-0.05	-1.37	-0.06	-1.50
Severity*Possibility					**0.06**	11.07	**0.06**	11.12
Emotional state							<u>-0.08</u>	-3.01
Type of worst loss:								
Physical damage							**0.56**	5.84
Interactions:								
Friends and family							<u>0.55</u>	3.04
Random effects:								
Variance components								
Level 2 (individuals)	**2.96**		**1.72**		**1.65**		**1.61**	
Level 1 (occasions)	5.20		3.51		3.32		3.27	

Notes: (1) Coefficients/variance components significant at p<.001 are in **bold**, significant at p <.05 are <u>underlined</u> using t-tests or χ^2 as appropriate.
(2) Degrees of freedom for level 2 variables and random coefficients are at least 70. For all other variables they are at least 2166.

<u>Table 4:</u>   <u>Mean risk judgments for 4th and 27th messages</u>

	<u>Message number</u>	
	<u>4</u>	<u>27</u>
Judgments		
Retrospective	13.9	20.8
Simultaneous	27.7	31.5
Difference	-13.8	-10.7

Figure 1: Fitted risk from emotional state and gender

Appendix A1: Coding schedule for activities

Category	Definition and examples
Professional activity	Activities that the participant describes as professional
Basic occupation	Activities based on the professional activities of each participant. Examples: "look for information on the internet", "hand in documents to register property", "speaking on the telephone", "combing the hair of the child I look after", "preparing this afternoon's class", "taking boxes up to the warehouse", "explaining a policy to a client".
Transportation – professional	Traveling, be it walking, by public or private vehicle, and whether as driver, accompanying, or passenger. Examples: "driving,"I am in the train", "traveling by metro", "crossing the street," "walking to work".
Pauses at working day, start/end of working day	Pauses during the workday and transitional activities between the beginning and end of the workday that the participant classifies as professional. Examples: "I am sitting waiting for a client to come", "taking a coffee", "arriving at work", "collecting my things before leaving".
Personal activity	Activities that the participant categorizes as personal.
Housework, personal time organization and managing funds	Housework (cleaning, tidying, preparing food, repairs, gardening...), managing money, home finances, and organizing time. Examples: "lighting butane stove", "I am preparing the meal", "doing the dishes", "going over the plan for tomorrow".
Eating and drinking	Examples: "taking a coffee in the bar", "having lunch", "eating an apple".
Smoking	Examples: "lighting a cigarette".
Transportation – personal	Traveling, be it walking, by public or private vehicle, and whether as driver, accompanying, or passenger. Examples: "walking quickly to my apartment", "going home in the train", "driving", "driving my scooter from work to home"
Communication -- personal	Interaction, personal or otherwise, with the goal of transmitting information, ideas, feelings, etc. Examples: "speaking by telephone", "speaking with my parents", "discussing with my partner", "speaking with a friend on the computer".
Shopping	Any activity related to shopping or acquiring goods. Examples: "loading supermarket bags", "looking at things for the house in IKEA", "trying on a pair of trousers",
Acquiring information	Activities for acquiring information or education. Examples: "printing a document", "reading notes", "reading the paper at the university", "connected to the internet", "looking at a car catalogue".
Entertainment, recreation, sports	Activities aimed at recreation but excluding those that imply acquiring information or education. Examples: "playing at football", "looking at a football game", "looking at TV "listening to music", "playing the guitar".
Personal care	Activities related to active personal care whether the goal aesthetic, diagnostic, or therapeutic. Examples: "showering", "shaving", "applying humidifying cream to my hands", "doing makeup", "getting undressed to take a bath", "getting dressed".

Rest, sleep	Activities associated with resting. These include ending periods of rest. Examples: "sleeping", "taking the sun", "relaxing on the sofa", "getting up from bed"
Waiting	Waiting for a person, an object or a moment in time. Includes preparing to end work when this is classified as a personal activity. Examples: "waiting for my friends in the football team", "waiting for the bus", "getting my things together to go home."
Other	Uninterpretable responses

Appendix A2: Coding schedule for worst consequence

Category	Definition and examples
None	Indicated specifically that there were none.
Physical damage	Physical damage that implies a loss of health, Examples: "that I would cut myself with the paper", "bite my tongue", "burn myself", "a heart attack", "fall down", "get killed in an accident".
Psychological discomfort	1. Psychological disturbances. Includes psychiatric disturbances, such as behavioral, cognitive and emotional disorders. Examples: "get depressed", "stress", "that I'll get an anxiety attack". 2. Physical, psychological, or social discomfort. Examples: "that they are not broadcasting my favorite series", "that my cell telephone wakes me up", "fail in my work", "that I lose my team", "become bored", "that I don't understand".
Social conflicts	Deterioration or loss of relationships with others. Examples: "that the client gets angry", "that the interviewee starts to shout and insult me", "fight with somebody", "that there was a misunderstanding", "that we get angry".
Loss of time or materials	Deterioration or loss of something material or undesired changes in organizing one's time. Examples: "that the printer breaks down", "lose information", "miss the train and arrive late", "arrive late to class", "that the house has been robbed", "that the connection to the internet has been cut off", "not find any parking and arrive late".

Appendix B: Examples of responses

Smoking
-lung cancer
-burn myself with a cigarette
-burn myself
-burn my eyelashes
-cancer
-that I burn myself
-burn myself, etc...
-that I get asthma
-burn myself
-get cancer
-burn myself
-that the police come
-that my parents learn that I smoke
-burn my trousers
-that my short gets burned

Eating and drinking
-to choke
-to choke or indigestion
-to suffocate
-that I choke on my food
-that I choke on a potato
-to choke
-that I don't feel well
-have a stomach cramp
-that the food does not agree with me
-to choke
-that I eat somthing bad and get poisoned
-that something I eat makes me feel ill and I get indigestion
-get acid in my stomach
-burn myself