Within-Person and Between-Sample Sensitivity to Dimensions of Risk

Authors:

- Graham Loomes, University of Warwick, Coventry CV4 7AL, United Kingdom. Corresponding author. E-mail: G.Loomes@warwick.ac.uk.
- Miguel Flores, University of East Anglia, Norwich NR4 7TJ, United Kingdom.

Acknowledgements:

- This study was undertaken as part of the ‘Social Contexts and Responses to Risk’ (SCARR) network, funded by Economic and Social Research Council Grant L326 25 3054.
- Miguel Flores acknowledges funding from the Economic and Social Research Council.

Within-Person and Between-Sample Sensitivity to Dimensions of Risk

Miguel Flores
University of East Anglia[y]

Graham Loomes
University of Warwick[z]

January 28th 2014

Abstract

This paper reports the results of a survey designed to elicit probability judgements' for different kind of events. We find limited sensitivity to time both within-person and between groups of people. Second, we find some sensitivity to scope within-person - although arguably less than we might expect - and limited sensitivity to scope between groups of people. Finally, the subjective probabilities about crime and injuries appear to be largely unaffected either by feedback about the known frequencies of such events or by people's experience/practice estimating the likelihood of pure chance events.

JEL Classification: C42, C81, D84.

Keywords: Subjective probability; Elicitation methods; Survey methods; Scope sensitivity.

---

[y] ESRC Centre for Competition Policy, University of East Anglia, Norwich NR4 7TJ, United Kingdom. M.Flores-Sandoval@uea.ac.uk.

[z] Corresponding author. University of Warwick, Coventry CV4 7AL, United Kingdom. G.Loomes@warwick.ac.uk.

1 INTRODUCTION

Many individual decisions involve risk and uncertainty, ranging from daily decisions (e.g. ‘Should I overtake this cyclist/car?’) to more infrequent choices (e.g. ‘Should I insure my pet?’). Such decisions are usually supposed to involve the individual making some judgment about the likelihoods of the relevant events – the chances of the overtaking manoeuvre resulting in injury, vehicle damage and/or loss of time, the chances that the pet will be lost or stolen or require veterinary treatment. It is widely assumed that most decision makers form reasonably coherent judgments about the relevant subjective probabilities so that – in aggregate, at least – their expectations are broadly realistic and their decisions are correspondingly well-founded.

However, rather than rely on such assumptions, it has been suggested that researchers should seek to elicit people’s subjective probabilities directly. Manski (2004) reviews some of the history of such endeavours and discusses a number of the issues arising. He concludes that information about subjective probabilities elicited in a number of surveys can often be a useful input into the analysis of people’s behaviour, while also noting that for some people and for some types of questions there may be uncertainty, imprecision and coarseness (rounding) – see, for example, Manski and Molinari (2010). He also acknowledges that the correspondence between subjective expectations and objective realities is variable: for example, while people’s judgments about the chances of losing their jobs within the next 12 months or not having health insurance in 12 months’ time corresponded quite closely with the proportions observed a year later, the judged probabilities of being burgled were on average about four times higher than the rates actually experienced.

Manski accepts the infeasibility of directly checking the efforts respondents make to give accurate and truthful revelation of their expectations, but considers that “it is possible to informally judge the *face validity* of responses by examining the degree to which persons give internally consistent, sensible responses to the questions posed" (2004, p.1343). He inclines to the view that "mainly, although not always, . . . responses do possess face validity when the questions concern well-defined events that are relevant to respondents' lives".

The aim of this paper is to undertake various checks of the internal consistency of people's stated beliefs and examine their broad correspondence with observable frequencies. In particular, we study whether people's subjective probabilities concerning 'real life' personal events, such as being a victim of crime or sustaining accidental injuries, are sensitive to variations of time and scope. We also investigate whether people's probability judgements about such events are affected by receiving frequency information; and whether some experience of estimating the likelihood of 'pure chance' events (e.g. dice throws, card draws) influences their judgments about the probabilities of the real life events in question. We find limited sensitivity to time both within-person and between groups of people. We find some sensitivity to scope within-person - although arguably less than we might expect - but limited sensitivity to scope between groups of people. It also seems that the stated subjective probabilities about crime and injuries are largely unaffected either by feedback about the known frequencies of such events or by people's experience/practice estimating the likelihood of pure chance events.

This paper contributes to the literature that suggests people are not as sensitive to the scope of various characteristics of goods as might ideally be expected (e.g. Smith and Osborne, 1996; Frederick and Fischhoff, 1998; Loomes, 2006; Yechiam and Budescu, 2006). The paper closest to ours is Loomes and Mehta (2007), henceforth LM, which was designed to investigate the sensitivity of people's expectations about different types of event; in particular, whether people articulate their subjective probabilities for the more socially-situated events in a way comparable with their judgements about pure chance events. Our paper reports a more recent survey, following up on issues unresolved in LM. This paper is also related to the literature on belief updates which studies how information affect people's perceptions of risk (see, for example, Hoffman et al., 2011; Poinas et al., 2012).

In the next section we describe the design of the survey, explaining the rationale behind its structure. Section 3 reports the key results. The final section draws the different strands together and discusses some implications of our results for future research.

2 STUDY DESIGN AND IMPLEMENTATION

The primary aim of the study was to focus upon two classes of event: being a victim of crime; and suffering an injury. We were particularly interested in these types of events because the earlier study reported in LM had suggested that people's subjective probability judgments for 'personal' events such as these were rather less sensitive to key parameters than their judgments for 'public' events such as changes in the price of petrol, changes in political leadership, weather conditions, etc.

In the present study, within each type of event there were two levels of scope. For the crime context, the smaller scale event was expressed as "Your home is burgled", while the larger scale event was expressed as "You are a victim of any crime at all (burglary, theft, fraud, assault, murder, etc.". Thus it was explicit that the smaller event was a proper subset of the larger, and some indication of other elements in the larger set was clearly stated. We shall label the first event Burglary and the second event All_Crime.

Likewise for the injury context: the smaller event was described as "You are injured in a road accident seriously enough to require professional medical treatment" while the larger event was given as "You are injured in a road accident or any other kind of accident (e.g. at work or at home) seriously enough to require professional medical treatment". Here we shall label the first event RTA and the second event All_Injury.

Besides the two levels of scope, there were two periods of time: 12 months, or 5 years. Thus for each of the four events above, we could investigate how stated subjective probabilities varied with the time interval. If the likelihood of an event occurring in any one year, $\boldsymbol{p}$, is independent of its likelihood of occurrence in another year, then the probability of experiencing it at least once during a 5-year period is $1 - (1 - \boldsymbol{p})^5$. We aimed to see how far people's responses conformed with this relationship.

The design allocated respondents at random to different 'treatment' groups, some of which entailed keeping the time period fixed and varying the scope of the events, while others presented the same event over the two different lengths of time. This allowed both within-person and between-group comparisons to be made.

Although it is sometimes thought to be easier to publish studies which produce striking 'anomalies' as compared with studies that quietly confirm conventional assumptions, we did not set out to make it difficult for standard assumptions to succeed. On the contrary, we were interested in exploring elicitation methods that we thought might aid the accuracy and sensitivity of responses. The results reported by LM had suggested that people who had undertaken a task involving assessing the probability of 'pure chance' events before giving subjective probability responses were somewhat more likely to moderate their estimates[1]. We therefore included a pure chance task (PCT) which involved presenting respondents with the six events in Table I. Each event was printed on a separate card and the set of six cards (previously shuffled) was given to the respondent, who was then asked to lay them out on a template in order from the one they thought was most likely to occur down to the one that was least likely. They were then asked to provide numerical estimates of the probabilities of occurrence for each event[2]. Depending on the treatment group to which they had been assigned, the respondent might stop at that point and receive no feedback (denoted by PCT-NF) or might be given feedback in the form of

[1] The effects of this manipulation in LM were statistically significant only in a minority of the 12 cases examined; but in all 12 cases the changes in means and medians were in the same direction, suggesting that there might well be a systematic, although modest, effect.

[2] Respondents could give their estimates in whatever form they felt comfortable with – most commonly, fractions or percentages; and if, when they came to give these estimates, they wished to alter their initial ranking in the light of further thought, they were told that they should feel free to do so.

"the chances that a professional statistician would put on these events" (denoted by PCT+F).

Table I. The Six Pure Chance Events

- When a standard pack of 52 playing cards is shuffled, the first card to be picked at random is a red one.
- When a standard pack of 52 playing cards is shuffled, the first card to be picked at random is a diamond.
- When a single dice is thrown once, the number that comes up is six.
- A person chosen at random has a birthday in June.
- When a standard pack of 52 playing cards is shuffled, the first card to be picked at random is the 7 of diamonds.
- When a day of the year is randomly chosen, that day is your birthday.

The procedure whereby respondents first produced a ranking and then went on to give numerical estimates was also employed for what we shall refer to as the 'mixed events' (ME) task. In this task, a respondent was given 8 cards. Two of these were the PCT cards relating to the chance of picking a diamond and to the chance of picking the 7 of diamonds from a standard pack. In treatment groups where the PCT+F task preceded the ME task, those two cards showed the 'professional statistician' estimates; in the other treatment groups where the PCT-NF task came later, those cards omitted those estimates.

Two other cards involved different increases in the price of postage stamps over different periods. These were included partly to add variety and partly to provide cases where the judged probabilities were likely to be higher, to provide broader coverage of the 0-1 probability interval[3]. They do not form part of the analysis in this

[3] Since Parducci (1965) proposed his range-frequency model, it has often been shown that if people are presented with a set of stimuli and a range or scale on which to map them, there is likely to be a tendency to spread the stimuli out to fill the range. We were concerned that if we presented respondents only or mainly with low-probability events, they would be inclined to 'spread' their estimates up towards the higher part of the 0-1

paper.

The other four cards contained the personal event scenarios which are the focus of this paper, involving various combinations of scope and time. For example, in what we label Treatment Group A, the four scenarios were being burgled during the next year (B1), being burgled during the next 5 years (B5), being injured in a road accident in the next year (R1) and being injured in a road accident in the next 5 years (R5). As another example, the four personal scenarios in Treatment Group E involved burglary, all crime, road accident and all kinds of injury each during the next 5 years (denoted respectively by B5, A_C5, R5 and A_I5).

The ME procedure first asked each respondent to place the 8 cards on a template in order from the event they considered most likely down to the event they judged least likely; and then asked them to provide a probability estimate for each in turn. By proceeding to the estimation task via a less demanding ranking exercise and by including two pure chance cards with 'known' probabilities (for six of the eight groups), the intention was to facilitate consistency and accuracy. So, for example, someone who thinks that their chance of suffering an injury of any kind during the next 5 years is greater than the chance of picking the 7 of diamonds from a shuffled pack but less than the chance of picking any diamond from that pack might be encouraged to give an estimate somewhere between 0.25 and just under 0.02; and may further refine their estimate by considering whether the chance of A_I5 is (much) closer to one of those two pure chance events than the other.

The other 'manipulation' we investigated involved the provision of frequency information before (or after) dealing with ME task: we shall label this task FREQ. For six of the eight treatment groups, the ME task was preceded by questions which asked respondents to give their best estimate of the frequencies of burglaries and of road accident casualties in their local area during the previous 12 months. Having provided a response, they were given information about the actual numbers officially recorded and were then asked whether they thought that their own chances were

range. By including higher-probability events (the median estimates for three of the four postage stamp scenarios were 0.5) we hoped to mitigate any such tendency.

higher, lower or about the same as the local average. If it is the case that individuals who don't already have a good idea of their own absolute levels of risk may nevertheless have some approximate idea of where they stand in relation to a relevant average, such information might be influential. In this respect, the two treatment groups who were not presented with these questions and items of information until after they had undertaken the ME task serve as the control group.

Altogether, then, the key features of the design can be summarised as in Table II.

Table II. Summary of the Survey Design

Treatment Group	n	Order of Main Tasks			Composition of ME Tasks
		First	Second	Third	Two postage stamp and two pure chance scenarios **PLUS**
A	32	PCT+F	FREQ	ME	B1, B5, R1, R5
B	33	PCT+F	FREQ	ME	B1, B5, R1, R5
C	32	PCT+F	FREQ	ME	B1, A_C1, R1, A_I1
D	34	FREQ	ME	PCT-NF	B1, A_C1, R1, A_I1
E	32	PCT+F	FREQ	ME	B5, A_C5, R5, A_I5
F	33	FREQ	ME	PCT-NF	B5, A_C5, R5, A_I5
G	31	PCT+F	ME	FREQ	B1, B5, R1, R5
H	30	PCT+F	ME	FREQ	B1, B5, R1, R5

The survey was administered to a total of 257 individuals drawn from a list of members of the general public in Norwich who had previously responded to a mail shot inviting them to participate in research conducted at the University of East Anglia. Face-to-face interviews lasting approximately 40 minutes were conducted with respondents in their own homes with individuals randomly assigned to one of the eight treatment groups.

Every respondent began the survey with the same introduction and the same 'practice' set of three questions intended to familiarise respondents with the type of questions they were going to encounter. Having answered any queries from respondents about the procedures, the interviewer moved to the first main task as indicated in Table II.

3. ANALYSIS

The first aim of the paper is to study the sensitivity of subjective probability to variations in time and scope (Subsections 3.1 and 3.2). The second aim of the paper is to explore whether people's probability judgements are affected by information about the events under consideration and/or by the experience of estimating the likelihood of pure chance events (Subsection 3.3).

3.1 Time sensitivity

We begin by analysing how the distributions of the estimates for the 5-year time period compare with the corresponding distributions for the 1-year time period, holding the nature and scope of the event constant. Groups A, B, G, and H allow us to examine within-subject time sensitivity. Groups C and D saw only 1 year scenarios while Groups E and F saw only 5 year scenarios for crime/injury events, allowing us to examine between-subject time sensitivity.

For the within-subject time sensitivity we proceed as follows. For each individual in a given group, we take the $\boldsymbol{p}$ response given for the 1-year scenario and compute $1 - (1 - \boldsymbol{p})^5$ to obtain that individual's inferred response for the 5-year scenario. We then test whether there is a significant difference between the distribution of inferred responses compared with the distribution of responses stated by those same individuals when presented directly with the 5-year scenario. Table III reports the mean of the inferred and stated probabilities for both burglary and RTAs, together with the results of testing for the difference between the distributions, using the paired t-test and the Wilcoxon Sign-Rank test[4]. Both tests reject the null hypothesis of no difference between the inferred probability and the stated probability, with the stated probabilities being significantly lower than the inferred probabilities,

[4] Because the distributions were often skewed we complement the parametric paired t-test with non-parametric test.

suggesting a lack of time sensitivity even when both time periods are presented to the same individuals.

Table III. Time Sensitivity on a Within-Subject Basis

Groups	Burglary - 5 year		RTA - 5 year	
	Inferred	Stated	Inferred	Stated
A and B	0.211	0.146	0.205	0.116
Paired t-test statistic	0.025		0.003	
Wilcoxon test statistic	0.013		0.000	
G and H	0.234	0.146	0.204	0.132
Paired t-test statistic	0.002		0.014	
Wilcoxon test statistic	0.002		0.001	

For time sensitivity *between* groups, we test whether there is a significant difference between the stated probability of the events for those individuals in groups E and F (presented with 5-year scenarios for all four types of event) and the corresponding responses from Groups C and D (presented with 1-year scenarios throughout).

The upper part of Table IV reports the means of the stated probabilities and the significance levels of the relevant t-tests and Mann-Whitney tests. For all four comparisons, t-tests allow us to accept the null hypothesis that there is no significant difference between the 1-year and 5-year sample means. The Mann-Whitney test allows us to accept the null at the 5% level in three cases, but rejects it for RTA.

The lower part of Table IV shows the means of the distributions of estimates inferred from groups C and D by putting their members' 1-year responses into the formula $1 – (1 – \boldsymbol{p})^5$. When we compare these with the probabilities directly stated by Groups E and F in the upper part of the table, we can strongly reject the null hypothesis of no difference between the inferred and the stated probabilities for all four types of event.

Table IV. Time Sensitivity Between Groups

	Burglary	RTA	All_Crime	All_Injury
Groups C & D 1-year	0.080	0.088	0.180	0.127
Groups E & F 5-year	0.095	0.104	0.173	0.170
t-test statistic	0.573	0.561	0.856	0.232
M-W test statistic	0.134	0.046	0.638	0.074
5-year Inferred from Groups C & D	0.217	0.232	0.386	0.322
t-test statistic	0.002	0.003	0.000	0.002
M-W test statistic	0.003	0.017	0.000	0.007

These results are in line with the findings reported by Yechiam and Budescu (2006), which they argue are mainly due to what Kahneman (2003) called ‘extension neglect’. LM offer a similar explanation for their findings, suggesting that for personal events of this kind, the subjective judgment of likelihood may be driven mainly by the affective reaction to the aversive nature of the event, with the time information commanding only secondary attention. In cases such as those presented to groups C, D, E and F, the fact that the time period was the same for all four personal event cards could have encouraged respondents to ‘cancel out’ the time information to such an extent that the nature of the event predominated, giving the result that in each case there was no significant between-sample difference in the mean estimates generated. Presenting the two different time periods to the same individuals, as in Groups A, B, G and H, may have encouraged them to pay more attention to that dimension, resulting in somewhat greater sensitivity; but even this ‘treatment’ failed to produce anything close to the theoretically-required degree of sensitivity.

Other forms of extension neglect have been identified in terms of insufficient sensitivity to differences in the scale/scope of events (see, for example, Smith and Osborne, 1996; Frederick and Fischhoff, 1998; and Loomes, 2006). So we now consider sensitivity to the scope of the scenarios in the present study.

3.2 Scope sensitivity

As noted above, Groups C, D, E, and F were asked to consider the likelihood of being burgled on one card in the ME set and the likelihood of being a victim of any crime (explicitly including burglary among a number of other crimes) on a different card, holding the time dimension constant; and in the same set they were asked to consider a card concerned only with a road accident as well as another card referring to all accidental injuries (including road accidents) during the same period.

This allows us to compute for each individual their judgment of the ratio between the chances of the wider-scope and the narrower-scope events and to examine how these ratios compare with the ratios derived from national statistics.

Table V provides information about the patterns generated by the respondents in groups C, D, E and F, separating C and D from E and F to avoid any possibility that the different lengths of time period might be confounding factors. Reading from left to right, the columns show the numbers of respondents for whom the ratio was less than 1 (i.e. who gave a smaller probability for the wider-scope event); for whom the ratio was exactly 1 (same response to wider and narrower scope); and for whom the ratio was greater than 1 (as would normally be expected). The fourth column shows the proportion whose ratio was less than 10 (which, as we shall see, is approximately the ratio suggested by official statistical sources); and the final column shows the median response (and in brackets, the median among just those whose ratio was strictly greater than 1).

Table V. Within-Subject Scope Ratios

	< 1	= 1	> 1	% < 10	Median
Groups C & D					
All_Crime/Burglary	9	12	45	86.40%	2.00 (2.50)
All_Injury/RTA	7	13	45	89.20%	1.50 (2.50)
Groups E & F					
All_Crime/Burglary	6	8	51	93.80%	2.00 (2.00)
All_Injury/RTA	8	10	46	89.10%	1.67 (2.00)

Clearly, there is noise in the data, and there may be some respondents who misunderstand or make mistakes to the extent of assigning lower probabilities to events that must logically be at least as likely to occur. We might also allow that when respondents assign the same probability to the wider as to the narrower event, it may be a manifestation of the kind of 'rounding' often observed in such data (see Manski and Molinari, 2010). However, between 86% and 94% of respondents judge that the probability of the wider event is less than ten times the probability of the narrower event, and the median judgment for all those who give ratio strictly greater than 1 is either 2.50 or 2.00.

How do these compare with the ratios we might expect on the basis of official statistics? There is some variability between sources, but in the field of crime the two most influential sources are the British Crime Survey and the incidents recorded by the police. For 2006-7, the BCS found that the chance of being a victim of burglary was about 0.024 while the chance of being a victim of any crime was 0.24 (Nicholas et al., 2007, Table 2a) while police statistics suggested it represented about 11% of all crime recorded by the police (see Nicholas et al., 2007, Figure 2.3). By comparison with these relativities, responses in our survey were seriously insensitive to scope in this context.

For accidents, it is more difficult to find official figures that give exactly the relationship we are studying. In 2006, official statistics based on road accidents attended by policy recorded 28,673 serious non-fatal injuries in Great Britain (all of which by definition should have required professional medical treatment of some kind) and 226,559 slight injuries, a number of which would not require medical attention beyond, perhaps, roadside first-aid, although there is no official indication of the proportions involved. On the other hand, some accidents which are not notified to the police may involve subsequent medical treatment, either via a hospital Accident and Emergency (A&E) department or via a visit to a General Practioner (cuts or fractures resulting from cycling accidents are an example of the former; whiplash injuries that have delayed effects are examples of the latter). However, if we suppose that the officially-reported slight injuries that do *not* require medical attention are roughly equal to the unreported injuries that *do* require some professional attention, an annual figure of about 250,000 RTA injuries might be a reasonable one to use.

Most other non-fatal injuries requiring medical attention occur in the home, in the course of leisure pursuits, or in the workplace. For 2006, the Health and Safety Statistics record about 140,000 workplace injuries resulting in more than 3 days absence from work. It is not known what proportion of these required medical attention, but the period of absence suggests that most may have been of sufficient severity to merit some treatment. At the same time, there may have been many injuries that required medical treatment (cuts requiring some stitches, possible fractures requiring diagnosis) but that involved less (or no) absence from work.

However, these numbers appear small compared with those injured at home or in the course of leisure activities. The collection of data by the UK government in the form of the Home Accident Surveillance System and the Leisure Accident Surveillance System was discontinued after 2002, but at that time it was estimated that there were 2.7 million visits to A&E for treatment for home and leisure injuries.

Even if one were to adjust this figure down for 2006 to allow for the same kinds of reductions in home and leisure accidents as in road and workplace accidents between 2002 and 2006, it still seems safe to say that the probability of All_Injury should on average be at least 10 times greater than the probability of an RTA injury. But in the injury context too, Table V shows that there was widespread understatement of this relativity.

What factors might be contributing to this? Since the average estimate of the probability of being a victim of burglary reported in Table IV was 0.08 while the average risk according to the British Crime Survey was 0.024, it might be thought that there was a widespread tendency to overestimate the chances of the narrower event[5]. But this was the mean from a skewed distribution where the median was 0.01 – that is, rather lower than the UK average – but where 17% of the sample gave estimates greater than 0.10. So the mean was inflated by a minority of excessively high responses and the fact is that the majority of estimates were well below the 'objective' average. This was even more the case for those same people's estimates of the chances of being a victim of *any* crime: while the BCS figure for this was 0.24, the median for Groups C and D was just 0.04, with more than 70% of respondents giving estimates below the BCS figure. So it was not the case at the individual level that most people were overestimating the narrower risk and underestimating the wider risk: for the crime scenarios, many understated the narrower risk but understated even more the wider risk, consistent with Kahneman's notion of extension neglect – failing to take adequate account of all of the other types of crime, even though a number of them had been mentioned explicitly on the card in question.

For accidental injuries the picture is a little different, in that in this context the majority *did* overestimate the probability of the narrower event. If the figure of around 250,000 RTA injuries in 2006 is broadly correct, the average risk per member of the

[5] Groups C and D were not atypical: the mean estimate based on all 192 respondents presented with the 1-year burglary scenario was 0.078.

population was less than 0.005, whereas the median response was 0.011 (with the mean once again being inflated by a minority of relatively high responses – 25% of respondents gave an estimate of 0.1 or more). For All_Injury, the median response was 0.05, which is arguably not too far from the right figure; but the overestimation of the RTA risk still leaves the relativity between All_injury and RTA far too low.

The calculation of geometric means gives a broadly similar picture in terms of lack of sufficient sensitivity. Table VI shows the relevant geometric means. As in Table V, the wider event produces a figure only between 2 and 2.5 the size of the figure for the narrower event[6].

Table VI. Geometric Means

	Burglary	RTA	All_Crime	All_Injury
Groups C & D 1-year	0.021	0.017	0.052	0.038
Groups E & F 5-year	0.030	0.031*	0.060	0.061

* One respondent gave a zero response: replacing his/her zero response by 0.0001 or by 0.00001 (respectively, one half and one twentieth of his/her All_Injury response) produces the same geometric mean to the third decimal place; omitting this respondent gives 0.034 and 0.067 for RTA and All_Injury respectively.

3.3 Information and Experience

In this subsection we consider whether people's subjective probabilities regarding burglary and RTA injuries are affected by feedback about the known frequencies of such events. We also check whether the exercise of estimating the likelihood of pure chance events affects the judged probabilities of being burgled and of being injured in an RTA.

The summary of the survey design in Table II shows that the only difference between the treatments for Groups A and B and for Groups G and H was that Groups A and B were asked about the local risks of burglary and road accidents and

[6] The geometric means tend to show *slightly* more sensitivity to time than the arithmetic means in Table IV, but there is still much less sensitivity than the $1 – (1 – p)^5$ formula entails.

were given facts about those rates before they undertook the ME task and made their subjective probabilities, while Groups G and H made their judgments before they received any frequency information. Table VII reports the differences in means for the two pairs of groups and the results of testing for differences between the two sets of responses. Both tests accept the null hypothesis that there is no difference between the two in any of the four scenarios.

Table VII. Impact of Frequency Information

Groups	Burglary 1-year	Burglary 5-year	Injury 1-year	Injury 5-year
(A & B) minus (G & H)	0.001	0.0006	-0.0017	-0.0159
t-test statistic	0.969	0.987	0.941	0.623
M-W test statistic	0.618	0.451	0.664	0.658

Finally we check whether or not experience with the pure chance events influences people's probability judgements about the various personal events. The motivation behind this part of the design was twofold. First, the PCT exercise was a way of encouraging respondents to engage with thinking about probabilities that were mostly fairly straightforward: the LM study had established that majority of the kind of people in our sampling frame were able to answer these questions quite easily and accurately. Second, if people find it easier initially to *rank* events in order of likelihood, and if they can locate personal events relative to pure chance events, this may make the task of assigning probability magnitudes easier.

One direct comparison is between Group C and Group D, where everything was the same except that Group C undertook the PCT exercise and got feedback before the ME task while Group D did the ME task before the PCT exercise. The second direct comparison is between Group E (PCT+F first) and Group F (ME first). If the PCT+F experience were significantly influential, we should expect differences between the subjective probabilities for C compared with D and for E compared with F.

Table VIII reports the relevant differences in means and the results of the tests for the significance of these differences. Both the t-test and the Mann-Whitney test cannot reject the null hypothesis of no differences between the distributions of responses: experience with the pure chance events does not appear to have a significant effect on the probability judgements about the personal events in our study[7].

Table VIII. Differences According to PCT Experience

Groups	Burglary	RTA	All_Crime	All_Injury
C vs D	-0.053	0.014	-0.086	-0.044
t-test statistic	0.184	0.726	0.175	0.391
M-W test statistic	0.580	0.684	0.528	0.371
E vs F	-0.007	-0.011	-0.049	-0.062
t-test statistic	0.829	0.745	0.364	0.244
M-W test statistic	0.683	0.792	0.958	0.838

4 CONCLUDING REMARKS

This paper has considered whether people’s subjective probabilities for ‘real life' personal events involving crime victimisation and accidental injury are more or less in line with the frequencies with which such events actually occur, and whether the judged probabilities are appropriately sensitive to time and scope variation. We find clear differences between expectations and frequencies and insufficient sensitivity to both time and scope.

[7] We offer one possible *caveat*: it was noted in footnote 1 that in the earlier LM study, all 12 analogous comparisons suggested that PCT experience shifted estimates in the same (downward) direction, albeit seldom to a significant extent. Table VIII shows the same direction of change in seven of the eight comparisons. So we hesitate to reject too strongly the possibility that the PCT exercise has *some* systematic effect which might be statistically significant in larger samples.

Attempts to help respondents by asking them to undertake ranking exercises before stating actual magnitudes, and by embedding some ‘pure chance’ events in these exercises to help anchor on appropriate magnitudes, appear to have had limited effects at best. With hindsight, we might have included one or two other pure chance events with probabilities in the critical region – say, events with probabilities of 0.05 and 0.1 – to provide extra comparators. On the other hand, one of the things most often reported back informally by interviewers was the difficulty some respondents expressed when trying to compare pure chance events with personal events: although risk analysts may work on the assumption that the probabilities of all kinds of events can be mapped to the 0-1 interval, the commensurability of different kinds of events is not intuitive for at least some respondents.

Another somewhat discouraging result was the lack of impact of information about frequencies of the events in the local area. Respondents appeared to have little difficulty in saying whether, by comparison with the local population, they felt their own risks were above, below or about equal to the local average for the events in question; but this, in conjunction with quantitative information about those local averages, had no discernible effect on the distributions of responses.

We acknowledge the limitations of our study. It was based on a sample of modest size using respondents from a convenience sample, so we must be guarded about the power and generality of our conclusions. On the other hand, our sample was drawn from volunteers with above-average levels of education who were disposed to put effort and care into their responses. (The mean characteristics – gender, age and education – of each treatment group are reported in the Appendix.)

Also, our focus was upon a particular class of events, namely those with quite low probabilities but with consequences liable to prompt quite strong affective reactions. This type of event may be particularly vulnerable to neglect of information about more abstract dimensions such as the period of time involved. Indeed, while Manski and various co-authors have found much evidence of reasonably well-tuned expectations for a number of important life events, they too have noted less accuracy for burglary (Dominitz and Manski, 1997).

However, the kinds of events we have considered may be influential in people's social and economic behaviour – not least, in their decisions about what insurance to buy and what preventive/precautionary measures to spend time and money upon. Other kinds of events within a longer temporal dimension, such as catastrophic risks of various sorts, were beyond the scope of our study; but expectations in these areas may also be vulnerable to biases and insensitivities which inhibit optimal responses in people's behaviour towards them.

We have every sympathy with Manski's contention that data about people's expectations and subjective probabilities are potentially important contributors to understanding the choices they make and explaining their behaviour, and that more efforts should be made to collect such data. However, there are at least some types of event where people's judgments elicited through surveys cannot be assumed to display the necessary accuracy and sensitivity to important variables, and we suggest that additional questions should be included specifically to check the internal consistency and sensitivity of those data. Part of the programme of research in this field must involve trying to gain a better understanding of which types of events are more or less vulnerable in this respect, the reasons for any such vulnerability, and the possibilities for overcoming the difficulties during the data collection process and/or making appropriate allowance for the limitations of the data when they are incorporated into analysis.

APPENDIX

Table A1. Mean characteristics

Treatment Group	Gender	Age	Education
A	0.38	52	2.2
B	0.36	44	2.1
C	0.34	47	3.1
D	0.32	45	2.0
E	0.34	46	3.0
F	0.33	48	2.3
G	0.35	42	2.1
H	0.33	47	1.9
All Groups	0.35	46	2.3

The categorization of the variables is as follows.

Gender: = 0 if female; = 1 if male.

Age: in years.

Education qualification:

= 1 if degree-level qualification or equivalent (including higher degree);

= 2 if teaching or nursing;

= 3 if ‘A’ levels, ‘AS’ levels, or Highers, or equivalent;

=4 if GCSE grades A to C, ‘O’ level passes, or equivalent;

= 5 if GCSE grades D to G, CSE grades 2 to 5, or equivalent;

= 6 if recognised Trade Apprenticeship completed;

= 7 if clerical or commercial qualification (e.g. typing, book-keeping, commerce);

= 8 if no formal qualifications;

= 9 if qualifications other than those listed.

REFERENCES

Dominitz J, Manski CF. Perceptions of Economic Insecurity: Evidence from the Surveys of Economic Expectations. Public Opinion Quarterly, 1997; 61: 261-87.

Frederick S, Fischhoff B. Scope (In)sensitivity in Elicited Valuations. Risk, Decision and Policy, 1998; 3: 109–123.

Hoffman RM, Kagel JH, Levin D. Simultaneous versus sequential information processing. Economics Letters, 2011; 112: 16–18.

Jansson K. British Crime Survey – Measuring crime for 25 years. 2007. Available at: http://webarchive.nationalarchives.gov.uk/20110218135832/rds.homeoffice.gov.uk/rds/pdfs07/bcs25.pdf

Kahneman D. Maps of Bounded Rationality: Psychology for Behavioural Economics. American Economic Review, 2003; 93: 1449-75.

Loomes G. (How) Can We Value Health, Safety and the Environment? Journal of Economic Psychology, 2006; 27: 713-736.

Loomes G, Mehta J. The sensitivity of subjective probability to time and elicitation method. Journal of Risk and Uncertainty, 2007; 34: 201-216.

Manski CF. Measuring Expectations. Econometrica, 2004; 72(5): 1329-76.

Manski CF, Molinari F. Rounding probabilistic expectations in surveys. Journal of Business and Economic Statistics, 2010*;* 28(2): 219–31.

Nicholas S, Kershaw C, Walker A. Crime in England and Wales 2006/7. Home Office

Statistical Bulletin, 2007. Available at: http://webarchive.nationalarchives.gov.uk/20110218135832/http:/rds.homeoffice.gov.uk/rds/pdfs07/hosb1107.pdf

Parducci A. Category Judgment: A Range-Frequency Model. Psychological Review, 1965; 72(6): 407-418.

Poinas F, Rosaz J, Roussillon B. Updating beliefs with imperfect signals: Experimental evidence. Journal of Risk and Uncertainty, 2012; 44: 219-241.

Road accidents and safety statistics. Department of Transport, UK. Available at: https://www.gov.uk/government/statistical-data-sets/ras30-reported-casualties-in-road-accidents

Smith VK, Osborne L. Do Contingent Valuation Estimates Pass a `Scope' Test? A Meta-analysis. Journal of Environmental Economics and Management, 1996; 31(3):287-301.

Yechiam E, Budescu D. The Sensitivity of Probability Assessments to Time Units and Performer Characteristics. Decision Analysis, 2006; 3(3): 177-93.