Effect of risk attitudes on recall of assignment statements (part 1 of 2)

Experiment performed at the 2011 ACCU Conference
First published in CVu vol. 23 no. 6

Derek M. Jones
derek@knosof.co.uk

1 Introduction

One of the first major discoveries in experimental psychology was a feature of human memory that has become generally known as *short term memory*. People are able to temporarily retain a small amount of information in memory whose accuracy quickly degrades unless an effort is made to 'refresh' it, and the information is easily overwritten by new information.

The capacity of short term memory (STM) has been found to correspond to approximately two seconds worth of sound, with some people have less capacity and some more.

There have been a huge number of experiments investigating the characteristics of STM and its impact on human cognitive performance. Since 2004 I have been trying to experimentally[7, 8] measure the impact of STM on developer performance when recalling information about previously seen source code (usually a sequence of assignment statements). In these experiments subjects have always been given the option to answer "I would refer back", i.e., if they have to recall this information in a work environment they would refer back to the previously read code rather than use whatever information they currently recall. In all experiments there have been subjects giving a much higher percentage of "would refer back" answers than average.

The immediate explanation that comes to mind for a subject giving a high percentage of "would refer back" answers is that they have a lower capacity STM than other subjects; an alternative explanation is that these high "would refer back" subjects are risk averse (they may or may not also have a lower capacity STM).

This is the first of a two part article that reports on an experiment carried out during the 2011 ACCU conference investigating and analysing their performance on a memory task and measuring their risk attitude. This first article provides general background on the experiment and discusses the 'risk' related results, while part two discusses subject performance on recall of recently seem assignment statements.

1.1 The hypothesis

When people recall information from memory they get a feeling for the confidence level associated with the recalled information. A person who is comfortable taking risks is more likely to make use of information for which they have a low confidence level than a person who is risk averse.

Risk attitude is hypothesized to effect subject performance in a memory recall task in the following three ways:

1. a risk averse subject works more slowly through the experiment questions than a subject who is less risk averse,
2. a risk averse subject works through questions at a similar rate to other subjects but gives a higher percentage of "would refer back" answers than less risk averse subjects,
3. a risk taking subject will work through the questions at a rate that is faster than their cognitive abilities can reliably support; such behavior would be expected to generate a higher percentage of incorrect answers than somebody working within the bounds of their cognitive abilities.

2 Risk

Living in an uncertain world we are all used to taking risks and human risk behavior has been found to be influenced by many different factors. People vary in their willingness to take risks and an individual's approach to risk may vary across different domains (e.g., play vs. work).[11] People's willingness to take risks within a given domain may depend on their interaction with that domain (e.g., athletes taking greater risks during recreational activities, gamblers more gambling risks, smokers more health risks, etc).[6]

Risks might be taken for the thrill of it, because the risk taker believes the outcome will produce a benefit rather than a cost or because a person is unaware that the outcome of their actions is uncertain.

When making cost/benefit decisions people have been found to give answers that do not agree with the mathematically optimum answer. For instance, people are risk adverse for gains (e.g., given an 85% chance of winning £1,000 or unconditionally winning £800, the majority of subjects have opted for the unconditional

option) and risk seeking for losses (e.g., given an 85% chance of loosing £1,000 or unconditionally loosing £800, the majority of subjects have opted for the 85% option).[9]

When working on source code what perception of risk[10] do developers have and is any risk analysis they perform correct or misconceived[2]? Your author is not aware of any research investigating risk taking by developers while they are working on source code and so this experiment is something of a jump in the dark.

2.1 Measuring risk attitude

In 2002 Weber, Blais and Betz[11] created a questionnaire intended to measure people's risk attitude in six domains: Investing (e.g., money), Health/safety, Recreational, Gambling, Ethical and Social decisions; this set of statements has become widely used and was updated in 2006.[3] The questionnaire consists of various statements each specifying some action; subjects are asked to rate the likelihood they would perform the action, if they found themselves in that situation, on a scale of 1 to 7 (extremely unlikely to extremely likely). The answers are combined to create a measure of risk attitudes.

The 30 statements are:

```
Approaching your boss for a raise (S)
Swimming far out from shore on an unguarded lake or ocean (R)
Betting a day's income at the horse races (G)
Investing 10% of your annual income in a moderate growth mutual fund (I)
Drinking heavily at a social function (H)
Taking some questionable deductions on your income tax return (E)
Disagreeing with an authority figure on a major issue (S)
Betting a day's income at a high-stake poker game (G)
Having an affair with a married man/woman (E)
Passing off somebody else's work as your own (E)
Going down a ski run that is beyond your ability (R)
Investing 5% of your annual income in a very speculative stock (I)
Going whitewater rafting at high water in the spring (R)
Betting a day's income on the outcome of a sporting event (G)
Engaging in unprotected sex (H)
Revealing a friend's secret to someone else (E)
Driving a car without wearing a seat belt (H)
Investing 10% of your annual income in a new business venture (I)
Taking a skydiving class (R)
Choosing a career that you truly enjoy over a more secure one (S)
Riding a motorcycle without a helmet (H)
Speaking your mind about an unpopular issue in a meeting at work (S)
Driving while taking medication that may make you drowsy (H)
Bungee jumping off a tall bridge (R)
Piloting a small plane (R)
Walking home alone at night in an unsafe area of town (H)
Moving to a city far away from your extended family (S)
Starting a new career in your mid-thirties (S)
Leaving your young children alone at home while running an errand (E)
Keeping a wallet you found that contains £150 (E)
```

Risk domains are not limited to the six domains addressed by the DOSPERT questionnaire. People have been shown to exhibit other recognizable risk attitudes when operating in different domains, e.g., driving a car.[1] The DOSPERT questionnaire has been used in a variety of domains (see www.dospert.org), continues to be used and researched and provides a starting point for the empirical investigation of developer risk attitude during software development.

3 Experimental setup

The experiment was run by your author during a 40 minute lunch time session at the 2011 ACCU conference (www.accu.org) held in Oxford, UK. Approximately 370 people attended the conference, 30 (8.1%; 3 joined 7 minutes after the experiment started) of whom took part in the experiment. Subjects were given a brief introduction to the experiment, during which they filled in background information about themselves, and then spent 20 minutes answering problems. All subjects volunteered their time and were anonymous.

3.1 The problem to be solved

The problem to be solved followed the same format as an experiment performed at a previous ACCU conference and the details can be found elsewhere.[7]

The following is an excerpt of the text instructions given to subjects:

The task consists of remembering the value of four different variables and recalling these values later. The variables and their corresponding values appear on one side of the sheet of paper and your response needs to be given on the other side of the same sheet of paper.

1. Read the variables and the values assigned to them as you might when carefully reading lines of code in a function definition.
2. Turn the sheet of paper over. Please do **NOT** look at the assignment statements you have just read again, i.e., once a page has been turned it stays turned.
3. For each of the following two statements, please indicate the likelihood that you would engage in the described activity or behavior if you were to find yourself in that situation.
4. You are now asked to recall the value of the variables read on the previous page. There is an additional variable listed that did not appear in the original list.

    - if you remember the value of a variable write the value down next to the corresponding variable,
    - if you feel that, in a real life code comprehension situation, you would reread the original assignment, tick the "would refer back" column of the corresponding variable,
    - if you don't recall having seen the variable in the list appearing on the previous page, tick the "not seen" column of the corresponding variable.

The following is an example of one of the problems seen by subjects. One side of a sheet of paper contained three assignment statements while the second side of the same sheet contained the five expressions and a table to hold the recalled information. A series of X's were written on the second side to ensure that subjects could not see through to identifiers and values appearing on the other side of the sheet. Each subject received a stapled set of sheets containing the instructions and 40 problems (one per sheet of paper).

```
---------------------- first side of sheet starts here ----------------------
 p =  5 ;
 b =  4 ;
 k =  9 ;
 t =  8 ;
---------------------- second side of sheet starts here ---------------------
For each of the following two statements, please indicate the
likelihood that you would engage in the described activity or
behavior if you were to find yourself in that situation.
Provide a rating from
Extremely Unlikely to Extremely Likely, using the following scale:
1.  Extremely Unlikely
2.  Moderately Unlikely
3.  Somewhat Unlikely
```

```
4.  Not Sure
5.  Somewhat Likely
6.  Moderately Likely
7.  Extremely Likely
Swimming far out from shore on an unguarded lake or ocean :
Approaching your boss for a raise :
               remember    would refer back     not seen
 b       =      ____            __                __
 t       =      ____            __                __
 p       =      ____            __                __
 s       =      ____            __                __
 k       =      ____            __                __
```

End of excerpt of text instructions given to subjects.

The question answering task acts as both a time filler for the assignment remember/recall problem and as a method of gathering as much information as possible in the limited time available.

4 Results

All subjects answered questions for the same amount of time (20 minutes) and were requested to perform at the rate they would use during normal work. In the past some subjects ignored this request and attempted to answer all questions in the booklet they were given. To try to prevent this behavior occurring the booklet contained many more questions than it was thought subjects could complete and they were told during the introduction about this rationale.

The 30 subjects had a mean of 14.3 years (sd 8.2) experience writing software professionally and answered 582 complete questions (average of 38.8 risk ratings and 96.7 individual assignment answers).

Subject performance will depend on a spectrum of cognitive abilities, one of which is risk attitude. The Spearman rank correlation coefficient was the statistical test used to measure the correlation between the six risk domains in the DOSPERT questionnaire and the various performance measurements described below. The source code of R program written to analyse the data is available on the web page www.knosof.co.uk/dev-experiment/accu11.html along with the (anonymous) data extracted from subject answers.

There are 30 statements in the risk questionnaire and two were randomly chosen, without replacement, to appear in each complete experimental question. Subjects who answered fewer than 15 complete experimental questions will not have given ratings to all risk statements, while subjects answering more than 15 will have rated some risk statements twice (the last answer given was used). Each subject's risk rating answers were averaged within each of the six risk domains.

4.1 Working more slowly

Figure .1 is a scatter plot of the total number of answers given for the assignment/recall component of the experiment (i.e., all correct, incorrect, "would refer back" and "not seen" answers) against the six risk attitudes (each dot represents one subject).

There is no obvious pattern to the subject responses and the Spearman correlation coefficients don't stray too far from zero and have a p-value that is not statistically significant (values not given here to save space and can be obtained by running the R source available on the experiments web page).

4.2 Higher percentage of "would refer back"

Figure .2 is a scatter plot of the percentage of "would refer back" answers against the six risk attitudes.

There is no obvious pattern to the subject responses and the Spearman correlation coefficients don't stray too far from zero and have a p-value that is not statistically significant. Treating the three subjects having a significantly higher percentage of "would refer back" answers than other subjects as outliers and not including them in the correlation analysis does not change the analysis (values not given here to save space and can be obtained by running the R source available on the experiments web page).

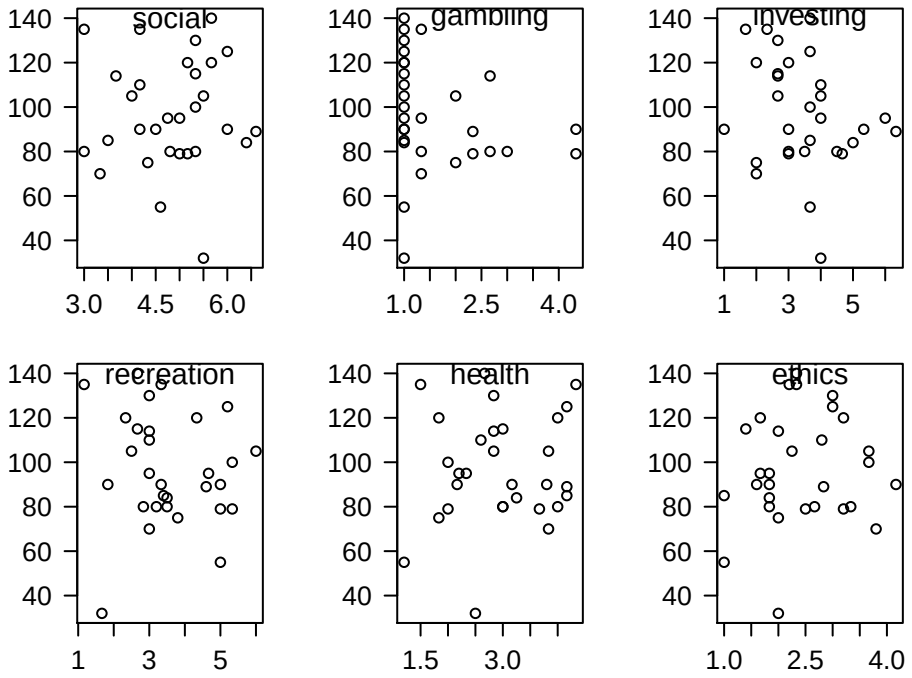

Figure .1: For each subject a scatter plot of risk attitude (x-axis) in six domains against total number of assignment/recall questions answered (y-axis).

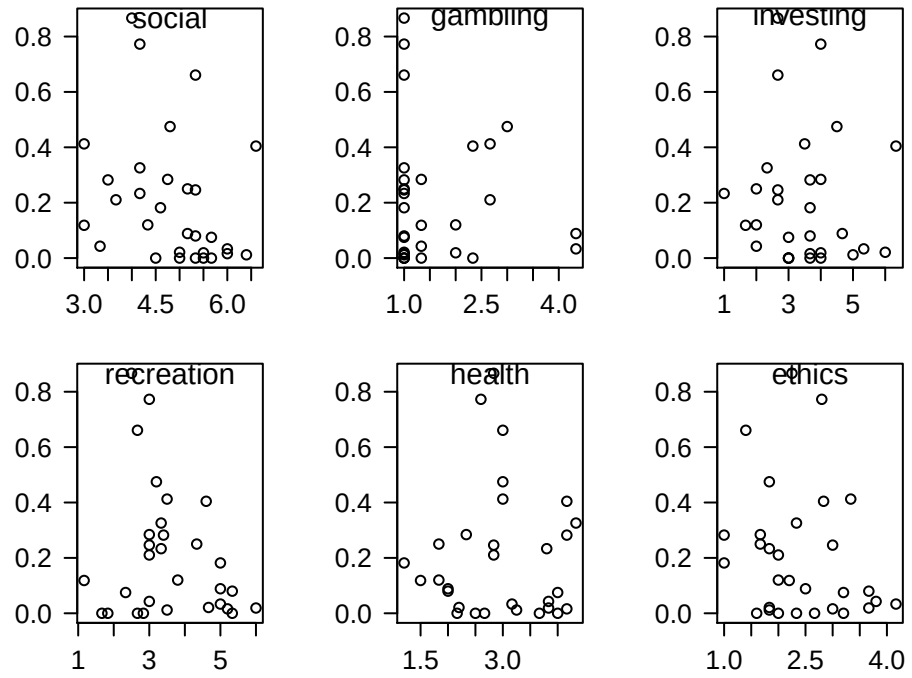

Figure .2: For each subject a scatter plot of risk attitude (x-axis) in six domains against percentage of "would refer back" answers (y-axis).

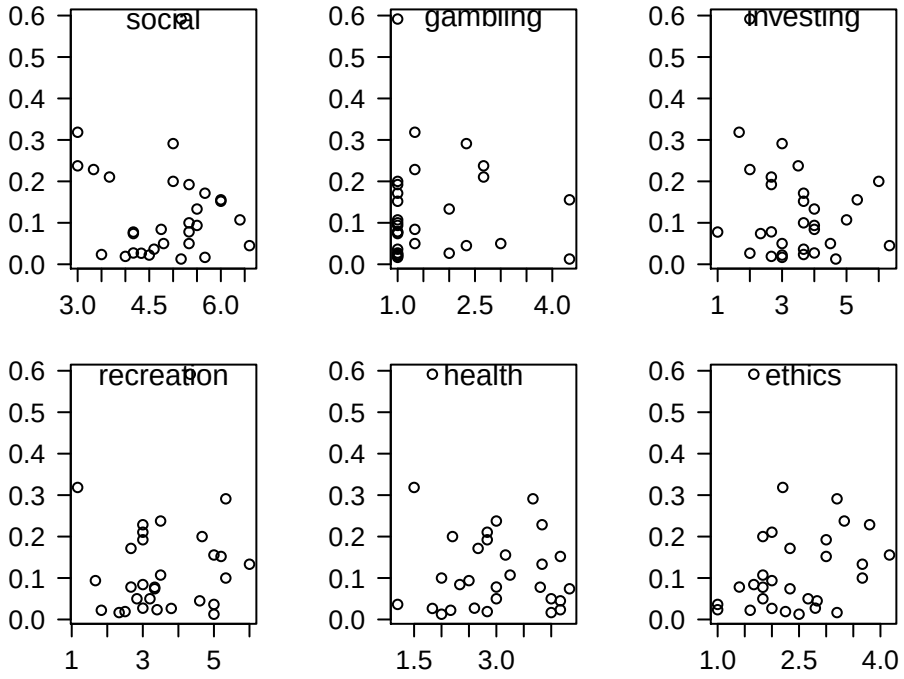

Figure .3: For each subject a scatter plot of risk attitude (x-axis) in six domains against percentage of incorrectly recalled answers (y-axis).

4.3 Higher percentage of incorrect answers

Figure .3 is a scatter plot of the percentage of incorrect answers given for the assignment/recall component of the experiment against the six risk attitudes.

There is no obvious pattern to the subject responses and the Spearman correlation coefficients don't stray too far from zero and have a p-value that is not statistically significant (values not given here to save space and can be obtained by running the R source available on the experiments web page).

The gambling risk correlation p-value (0.039) was less than the often used significance level of 0.05. However, if enough correlation tests are performed one will eventually be found that has a p-value below 0.05. The Bonferroni correction adjusts for multiple tests by dividing the significance level by the number of tests, in this case 0.05/6 gives 0.003 as the level below which a p-value will be considered significant. The gambling risk correlation is not significant at the level adjusted for the number of tests.

4.4 Summary of risk attitude

Table .1 gives the mean and standard deviation of risk attitudes, over all subjects, for each of the six domains, along with Cronbach's reliability coefficient alpha. Also included are values obtained by Blais[4] from 382 subjects from a variety of backgrounds.

The mean value in all risk domains for the ACCU subjects is less than the corresponding Internet subject means, but still within one standard deviation of the Internet mean values. For six cases there is a 1 in 64 chance of this pattern occurring through random selection.

The standard deviation of the mean in all risk domains for the ACCU subjects is less than the corresponding Internet standard deviations.

For the gambling domain the ACCU subject mean and standard deviation is a lot less than for the Internet subjects.

Cronbach's reliability coefficient alpha is a measure of the internal reliability (or intercorrelation among

items) of a set of test scores that are combined to create a single score. The third colum of values in Table .1 is the Cronbach alpha for the ACCU subjects. In those cases where a subject did not rate all of the DOSPERT statements missing ratings were treated as having a value that was the mean of the rating given for that domain by the subject.

A Cronbach alpha less than 0.5 is considered unacceptable (by statisticians), between 0.5 and 0.6 poor, between 0.6 and 0.7 questionable and between 0.7 and 0.8 acceptable. Only one risk domain had an acceptable value, one questionable and two poor, with two being unacceptable. The domain values derived from the Internet subjects all had a Cronbach alpha above 0.77.

It is possible that having to remember assignment information affected subject risk ratings. Like the ACCU subjects the Internet subjects spanned a wide range of ages.

Table .1: Mean and standard deviation for subject responses in each risk domain and Cronbach's alpha for the mean score. Last two columns are reported by Blais[4] based on responses from 382 subjects (after filtering to meet various criteria) recruited via the internet and paid for their time.

Domain	Mean	Standard deviation	Cronbach alpha	Internet mean	Internet SD
Social	4.84	0.95	0.58	5.27	1.09
Recreation	3.57	1.22	0.60	3.78	1.57
Gambling	1.6	0.96	0.73	2.85	1.92
Investing	3.44	1.24	0.58	4.12	1.53
Health	2.96	0.89	0.29	3.65	1.41
Ethical	2.42	0.82	0.34	3.14	1.31

4.5 Threats to validity

The structure of the experiment is such that giving a "would refer back" answer has a much lower cost than would have to be paid in real life, i.e., ticking the appropriate answer row vs. spending time searching back through code. A study by Fu and Gray[5] showed that this difference can be significant. They asked subjects to copy a pattern of colored blocks (on a computer-generated display). To carry out the task subjects had to remember the color of the block to be copied and its position in the target pattern, a memory effort. An effort cost was introduced by graying out the various areas of the display where the colored blocks were visible. These grayed out areas could be made temporarily visible using various combinations of keystrokes and mouse movements. When performing the task, subjects had the choice of expending memory effort (learning the locations of different colored blocks) or perceptual-motor effort (using keystrokes and mouse movements to uncover different areas of the display).

The subjects were split into three groups: one group had to expend a low effort to uncover the grayed out areas, the second acted as a control, and the third had to expend a high effort to uncover the grayed out areas. The results showed that the subjects who had to expend a high perceptual-motor effort uncovered grayed out areas fewer times than the other two groups. These subjects also spent longer looking at the areas uncovered and moved more colored blocks between uncoverings. The subjects faced with a high perceptual-motor effort reduced their total effort by investing in memory effort.

DOSPERT was originally developed and tested using university students as subjects. Based on an average of 14.3 years of professional software development the average age of the ACCU subjects is likely to be in the mid-thirties. Answers to some of the questions may change as people get older (e.g., older people are more likely to be in a stable relationship and having unprotected sex); the effect of age on the reliability of the measurement is not known. The average age of the Internet subjects was probably late twenties (82 were 21 years or younger, 223 between 22 and 35, and 71 were 36 or older).

5 Conclusion

This experiment failed to find any statistically significant correlation between subject risk attitude in six domains, as measured using the DOSPERT questionnaire, and subject performance in recalling information about a previously seen sequence of assignment statements. Either risk attitude is not a significant factor in

recalling information about assignment statements or the attitudes measured by the DOSPERT questionnaire are not applicable.

The ACCU subject risk attitudes were more risk averse than those from the Internet survey, the ACCU subjects were also more self consistent (i.e., the standard deviation in their scores was lower). However, Cronbach alpha values suggest that in some risk domains the single value obtained from combining all subjects ratings is not reliable.

6 Further reading

Statistics Explained by Perry R. Hinton provides a very good introduction to statistics.

The Art of R Programmibng by Norman Matloff teaches the R language as a language rather than as a tool to use for statistical analysis.

7 Acknowledgments

The author wishes to thank everybody who volunteered their time to take part in the experiment and the ACCU for making a conference slot available in which to run it.

References

1. J. Adams. *Risk and Freedom: The record of road safety regulation*. Transport Publishing Projects, 1985.
2. T. Aven. *Misconceptions of Risk*. Wiley, 2010.
3. A.-R. Blais and E. U. Weber. A domain-specific risk-taking (DOSPERT) scale for adult populations. *Judgment and Decision Making*, 1(1):33–47, Apr. 2006.
4. A.-R. Blais and E. U. Weber. The domain-specific risk taking scale for adult populations: Item selection and preliminary psychometric properties. Technical Report TR 2009-203, Defence R&D Canada, Dec. 2009.
5. W.-T. Fu and W. D. Gray. Memory versus perceptual-motor tradeoffs in a blocks world task. In *Proceedings of the Twenty-second Annual Conference of the Cognitive Science Society*, pages 154–159, Hillsdale, NJ, 2000. Erlbaum.
6. Y. Hanoch, J. G. Johnson, and A. Wilke. Domain specificity in experimental measures and participant recruitment. *Psychological Science*, 17(4):300–304, Apr. 2006.
7. D. M. Jones. Experimental data and scripts for short sequence of assignment statements study. `http://www.knosof.co.uk/cbook/accu04.html`, 2004.
8. D. M. Jones. Developer beliefs about binary operator precedence. *C Vu*, 18(4):14–21, Aug. 2006.
9. D. Kahneman and A. Tversky. Choices, values, and frames. In D. Kahneman and A. Tversky, editors, *Choices, Values, and Frames*, chapter 1, pages 1–16. Cambridge University Press, 1999.
10. P. Slovic. *The Perception of Risk*. Earcthscan Publications Ltd, 2000.
11. E. U. Weber, A.-R. Blais, and N. E. Betz. A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavior and Decision Making*, 15(4):263–290, Apr. 2002.