How Much is Location Privacy Worth?

George Danezis, Stephen Lewis and Ross Anderson

University of Cambridge, Computer Laboratory,
15 JJ Thomson Avenue, Cambridge CB3 0FD, United Kingdom
{George Danezis,Stephen.Lewis,Ross.Anderson}@cl.cam.ac.uk
http://www.cl.cam.ac.uk/users/{gd216,srl32,rja14}/

Abstract. We use techniques from experimental economics and psychology to determine how much compensation must be offered to persuade someone to allow precise information about their location to be collected. We pretend that we are running a study that needs volunteers to have their location monitored (via their mobile phone) over a period of one month. Volunteers apply by specifying the amount of compensation which they would require to participate in the experiment. The experimental subjects are led to believe that we will run a sealed-bid second-price auction on these values, and thus we obtain an estimate of the value that users attach to their location data being used by third parties.

1 Introduction

Computer security researchers have over the past two decades developed a number of sophisticated 'Privacy Enhancing Technologies' (PETs) to minimise the personal data leaked by everyday online interactions. The science is now relatively mature, yet commercial uptake has been very disappointing. Most products designed to support private communications and transactions have failed to be deployed beyond an experimental stage, or have failed in the market.

Many explanations have been offered for these failures, from the global to the product-specific. Stalder [1] believes that privacy is a concept of the industrial age in which printing was the main means of knowledge distribution, and that it cannot be preserved in the information age. Others [2] attribute the deployment failure to the poor integration and user interfaces of PET products.

More compelling explanations use economic arguments to explain why PETs have failed. Odlyzko [3] uses the tension between privacy and the ability to price discriminate to explain why such technologies are not deployed by service providers and merchants. Economic analysis has also been used to understand consumer needs for privacy better. Acquisti *et al.* [4] presented a framework to reason about the economics of anonymity, where participants might have incentives to participate in the provision of the service if they benefit from more anonymity as a result. Acquisti and Grossklags [5] performed extensive user surveys to understand perceptions of privacy: how much it is valued and how it is discounted.

Our work follows the economics tradition and attempts to measure the extent to which location information is valued by those whose location is being monitored. We use tools from experimental economics and psychology to infer the price at which a group of volunteers would be ready to disclose their precise location for a period of time. This price may be used as a guide to understanding the amount users would be ready to spend for products that protect their location information from third parties.

2 What is location information?

Location information is a set of data describing an individual's location over a period of time. The time and location resolution vary with the technology used to collect the data.

Location information is sensitive private data for many reasons. It is leaked and can be collected as a side effect of most wireless communication. Mobile phones have to register to a cell, of varying size, to connect to a network; they move from cell to cell as users roam around. Mobile phone operators record handset location information and supply it to third parties – to police, to marketeers and even to subscribers themselves (for example, where a company wishes to locate phones issued to sales or service staff). Triangulation and other techniques can be used to increase the accuracy of location; and some third-generation mobile phones have an integrated GPS receiver which provides location information with an accuracy of a few metres. There are also proposals to install third generation receivers into cars [6].

Wireless 802.11 Ethernet cards connect by registering with wireless access points, which can be used to locate a network device. Pascual [7] has studied traces of devices moving around, and has been able to infer information about the high-level relationships between different users; friends and colleagues tend to be at the same place more often than randomly expected. Intel Laboratories Cambridge performed a similar experiment by issuing their employees with small transponders that record their encounters: an otherwise undisclosed intimate relationship was uncovered when two transponders recorded rather frequent encounters late at night. A more thorough study has recently been performed at the MIT Media Lab in their Reality Mining project [8]. This involved gathering mobile phone location data and Bluetooth visibility information from 100 subjects over an academic year.

Finally, cheap GPS receivers are increasingly embedded into other devices and can record their location. This data can be stored for access at a later time or transmitted over a wireless communication channel. The deployment of all this technology means that most of the population will be walking around with a beacon that transmits their location. This has led some computer-security researchers to design location-privacy mechanisms. For example, mobile users may be assigned pseudonyms that can only be decoded by a machine at their home location; this machine in turn can determine who gets to learn whose location [9]. Such services have not, however, been commercially successful.

3 Experimental Design

The overall aim of the work described is to discover what value people attach to their location privacy. In this section we describe how an experiment to answer this question was devised, and identify its limitations. We also examine the ethical considerations related to our methods.

3.1 Auction structure

In order to induce subjects in our experiment to reveal the true value they attach to their location privacy, we carry out something that might loosely be termed a 'compensation auction'. We invite volunteers to participate in a fictitious study that will require the collection and processing of location information from their mobile phones. As part of the application to participate in this notional study volunteers are asked to disclose the amount they will require as compensation.

At the outset of the experiment, the subjects are told that they have the opportunity to participate in a future study on location privacy. The study will involve their location being recorded (with stated temporal and spatial resolution) for a stated period of time. It is specified that the collection will use the capabilities of the mobile phone network. An explanation of these capabilities is given to go some way towards satisfying our technically savvy participants.

Details of the auction are then provided: volunteers are asked to state the amount of (monetary) compensation that they would require to participate in the study. Our cover story is that since only a limited number of participants can be part of the study, we will invite the n people with the lowest values, and we will pay them the amount of compensation requested by the *lowest bidder not chosen*.

This structure is very similar to the 'multiple sales by sealed bids' auction described in [10], and it was chosen in order to ensure as far as possible that bidders are motivated to report the true value that they attach to their location privacy.

3.2 Choice of subjects

We chose the population of undergraduate computer science students at the University of Cambridge as the subjects for the initial study. This will limit the generalizability of our findings, but has three significant advantages. First, we can virtually eliminate the possibility of multiple submissions, through the use of the university's single-sign-on system, Raven [11]. We also know the number of potential participants, and can thus assess accurately the level of non-participation.

Second, we have greater control over the information that potential participants in the experiment receive before they take part. Since our experiment relies on participants believing that their sensitive data will be collected and processed, it is important that they believe that the notional study using it will take place. Third, computer science students get a course in their second year that introduces them to auction theory. Third-year students, at least, should be

aware that the experimental design is incentive-compatible, and that disclosing their true preference should be the optimal strategy.

3.3 Practical details

Our experiment is administered as a simple web application, which gives a number of advantages. The benefits of Internet-based experimenting are described in [12]; in our case, the most significant of these are:

- Avoidance of time constraints
- Highly voluntary participation
- Detectability of motivational confounding
- Reduction of experimenter effects
- Cost savings
- Ease of access for participants

For this study, it is important that we are able to assess the level of non-participation, and the level of dropout. Students who have information about the experiment, but choose not to take part, are non-participants. Students who begin participation in the study (by proceeding to the 'questionnaire' part of the website), but do not finish answering the questions are deemed to have dropped out.

In order to measure the level of non-participation, we disseminated a basic description of our experiment widely within the undergraduate computer scientist population. The experiment was described as 'a research project on the use of location data gathered from mobile phones', and the corresponding URL was given. By announcing the study in lectures and on the undergraduate mailing lists, we ensured that with high probability students at least knew of the existence of the study.

In entering the web application for the study, students first authenticated themselves. The reasons for this are twofold: we want to get an accurate idea of how many students are interested in getting further details of the study, and we want to use up-front gathering of personal information as a 'high-hurdle technique' (as described in [12]). We made an attempt to estimate the number of unique users who choose not to authenticate after visiting the site. It is important to note that this requirement for authentication might have biased some particularly privacy-sensitive students against participating, thus skewing our results.

After authentication, a full and detailed description of our notional future study was presented. This follows the standard practice in experimental psychology of presenting the most vexatious aspects of a study initially, in order to prevent dropout at a later stage. The text of this description is as follows:

> The Security Group in the Computer Laboratory is running a study that involves gathering location data for a number of volunteers over a period of 28 days. The location data will be retained, and may be used again for future research.

> The mobile phone of each participant in the study will be queried for its location every few minutes, 24 hours a day, for 28 days. The technological constraints imposed by the mobile networks mean that the locations returned to us will be at 'current cell' resolution (approximately 500 m radius circle in built-up areas, and 10 km radius elsewhere). These location requests will not interfere with the normal functioning of the handset. Each participant in the study will receive monetary compensation, and we are running an auction to select those who will take part. We invite you to submit a bid for the amount of money you require to take part in such a study. Successful bidders will be those who bid the lowest amounts, and each will be paid the amount of compensation demanded by the lowest unsuccessful bidder. (We have yet to decide how many participants we will require.)

The technical details are accurate enough, given the technology we claim to use, not to give rise to any suspicion or challenge the technological knowledge of our subjects.

At this stage, subjects were asked to select one of three options:

- I have a mobile phone, and am interested in participating in this study
- I have a mobile phone, but am not interested in participating in this study
- I do not have a mobile phone

The choice of those who selected one of the last two options was recorded, and their participation ended. Those who selected the first option were presented with a short list of questions. The questions were:

- What network do you use for your main mobile phone?
  - O_2
  - Orange
  - Vodafone
  - Other
- Do you carry a mobile phone with you most of the time?
  - Yes
  - No
- How often do you travel outside Cambridge?
  - More than once a week
  - About once a week
  - About once a month
  - Less frequently than once a month
- With whom do you communicate using your phone?
  - Friends
  - Family
  - Partner
- How much compensation would you require to participate in our study for 28 days (in whole pounds)?

The first question was merely a distractor to make participants believe that we need such details in order to plan our future study. The answers to the other questions were be used in our analysis of the results.

Ten days after the initial invitation to participate was issued, a follow-up email was sent out to those who filled in the questionnaire. We asked them if they would participate in a commecial study, and solicited revised bids. The delay was necessary in order to avoid biasing those who had not yet filled in the questionnaire. The email was worded as follows:

> Thank you for filling in our questionnaire about the Security Group's location study.
> There is now some possibility of commercial interest in the data collected during our study, and we would be grateful if you could let us know whether you would
> - not be willing to participate in the study if the data might be used by a commercial entity;
> - allow your data to be used by a commercial entity for the same amount of compensation you originally bid; or
> - allow your data to be used by a commercial entity only if you were allowed to revise your bid for compensation?
>
> If the last option applies to you, please let us know the value to which you would revise your original bid (of n pounds).

The last sentence of the email was customized to show the true first bid of each recipient. The replies to these emails were correlated automatically with the questionnaire answers, and coded as

- Not willing to participate in the light of commercial interest
- Willing to participate with same bid
- Willing to participate with revised bid
- Requested further details

We recorded the revised bid for each participant (where applicable), and whether it appeared that they had 'figured out' the real purpose of the experiment.

3.4 Ethical considerations

The experiment described in this paper involves an element of deception, and we aimed to adhere to the guidelines set down in the British Psychological Society's 'Code of Conduct, Ethical Principles & Guidelines' [13]. Deception is not unknown in psychology experiments, and the BPS consider it appropriate as long as it does not lead to 'discomfort, anger or objections' when it is revealed. A decision on this may not be taken by the Principal Investigator; at our institution the procedure is for the department's Senior Management Committee to take a view on whether the experiment should be referred to the Psychology Ethics

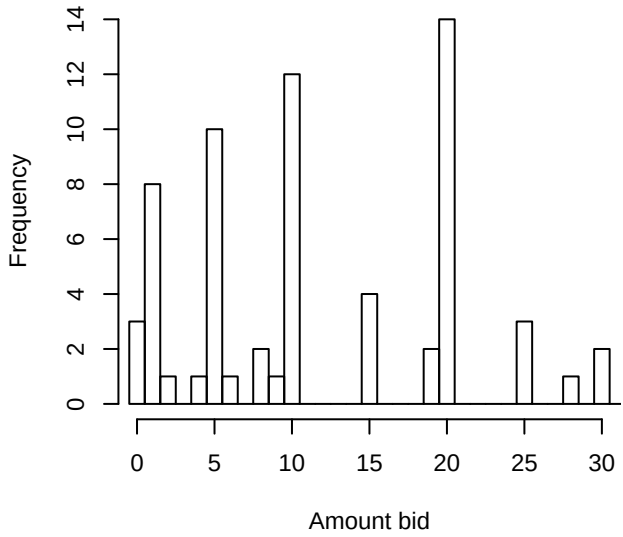

Fig. 1. Histogram showing amounts bid between £0 and £30

Committee for full consideration. In this case, our management decided that it did not.

Only one student expressed any displeasure at the element of deception in the experiment. Three other students also indicated that they believed deception was involved before the debrief email was sent. After the debrief was sent, there was significant interest in receiving the results of the study, and no further students indicated any objections to the methods used.

4 Data Analysis

Our analysis of the data collected during our experiment is split into four parts. We begin by looking at the distribution of the bids, and how it is affected by the suggestion of commercial interest. We then look at the distributions conditioned on the answers to the questions 'who do you communicate with?' and 'how often do you travel outside Cambridge?'. Finally, we comment on the data gathered from our web server logs on levels of participation and 'attention span'.

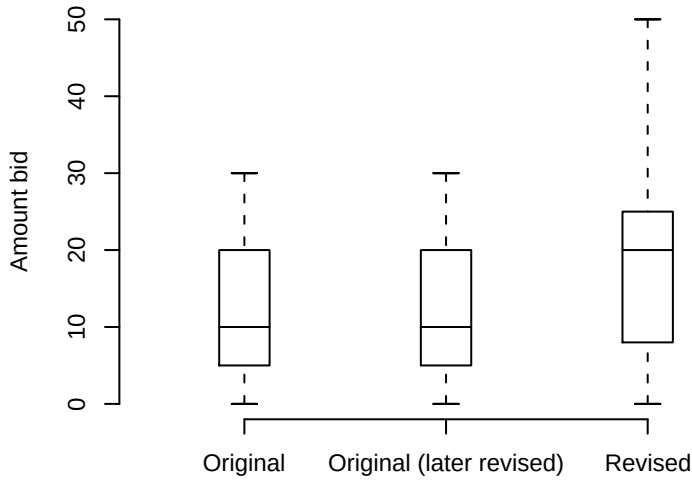

Fig. 2. Distributions of original bids (for the entire sample), original bids of those who indicated that they would still participate after commercial interest was mentioned, and their revised bids

4.1 Distribution of bids

The histogram in Figure 1 shows the frequency of amounts bid for the interval £0–30. It is important to note that this is not the entire picture; there were nine students who bid values greater than £30. The histogram does, however, show most of the detail in which we are interested.

The first immediately apparent feature of the histogram is that there was a preference for the values 1, 5, 10 and 20. This suggests that students were happier bidding 'round numbers', and imposed an artificial discretization on their true privacy valuation. There is also some evidence of strategic bidding: no bids from 16–18, but then 2 at 19, presumably aiming to undercut the bidders at 20.

Summary statistics for this initial distribution, of the 74 people who filled the full form, are shown below:

Min.	1st Q	Med.	Mean	3rd Q	Max.
0.0	5.0	10.0	27.4	20.0	400.0

We can also examine how the distribution of bids changed when the students were told that there was commercial interest in the data. Figure 2 shows three

distributions: the original distribution of bids, the distribution of original bids of those who indicated that they were still willing to take part in our experiment if there was commercial interest, and the revised bids. Outliers are not shown.

The graph shows that the distribution of bids of those who revised their bids (or re-bid the same amount) in the light of commercial interest is very similar to the distribution of the entire sample, and also what the distribution of revised bids looks like.

The summary statistics for the revised bids are:

Min.	1st Q	Med.	Mean	3rd Q	Max.
0.0	8.0	20.0	32.8	25.0	300.0

Of those who responded to our email about commercial interest (53 students, 72% of our sample), 2 were no longer interested in participating, 31 would participate for the same compensation and 16 would participate for a higher amount. The remaining 4 requested further information before making their decision.

4.2 Travel and communication patterns

The data gathered from the questionnaire about travel patterns has the following frequency distribution:

	n
More than once a week	1
About once a week	10
About once a month	38
Less frequently than once a month	25

There is a fairly good indication of a significant difference between the median bid of the students who travel outside Cambridge once a week or more and the rest of the sample; this is shown in Figure 3. A one-tailed Mann-Whitney-U test of whether the population median for the second group is higher gives $p = 0.05$, although this result must be interpreted with caution because of the small sample of students who travel outside Cambridge once a week or more.

We also examined how the distribution of bids varied between those who communicated with a partner using their phone ($n = 28$), and those who did not ($n = 46$). The difference in distribution is shown in Figure 4. A one-tailed Mann-Whitney-U test indicates that the median of the first population is not significantly higher than the median of the second ($p = 0.13$), but the graph suggests that the upward spread of bids made by those who communicate with a partner is greater.

4.3 Levels of participation and 'attention span'

All 277 undergraduates studying computer science were invited to fill out our online questionnaire, both with announcements during lectures, and by email.

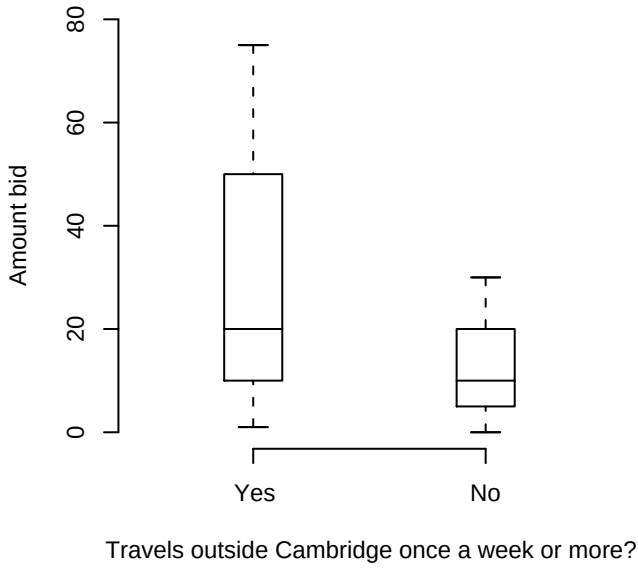

Fig. 3. Distributions of bids of those of travel outside Cambridge once a week or more versus the rest of the sample

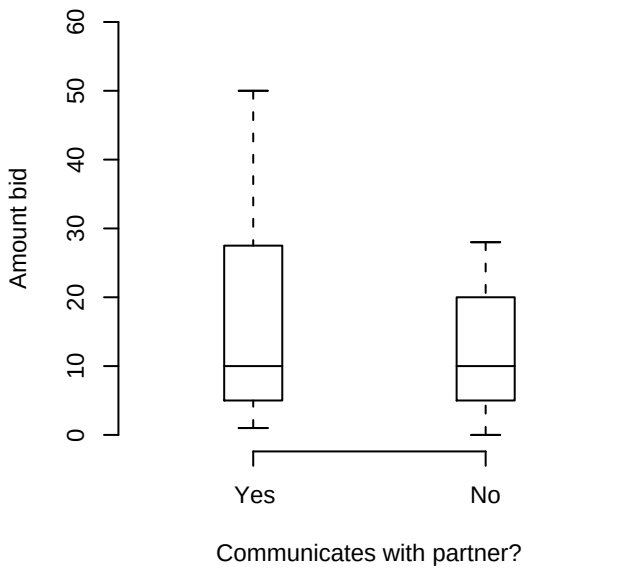

Fig. 4. Distributions of bids of those who communicate with a partner versus those who do not

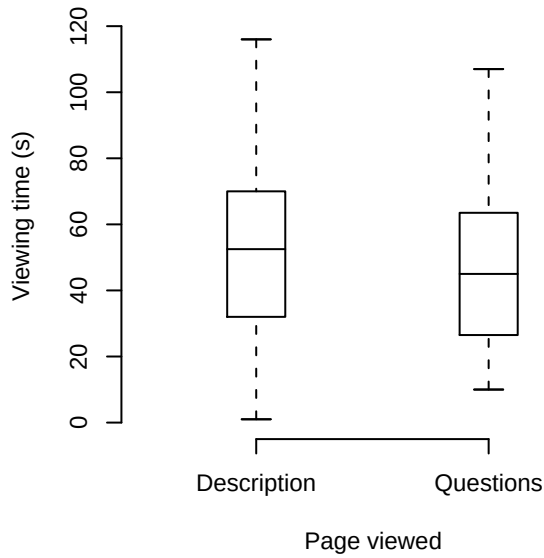

Fig. 5. Distributions of time spent looking at the experiment description and questions

The upper bound on the number of students who visited our site (by looking at the number of unique sessions recorded by our web application) is 216 (78% of the population). This number is an overestimate, because it is not corrected for students who visited the site more than once using different machines, or for those who did not allow the recording of cookies in their browsers.

Our participation statistics are more accurate once we consider those who actually logged into the site to receive more details. 111 students chose to log in to receive further details about our study. The dropout at this stage was before any detailed information about the nature of our study had been revealed, but the students who chose not to proceed could have been motivated by lack of time, lack of interest or privacy concerns.

Of the 111 students who authenticated, 74 filled in the questionnaire (67%), 5 said they had a phone but were not interested (5%) and 2 said they had no phone (2%). 13 students viewed the page of questions, and then chose not to fill them in.

We can also use the logs of page views to get some idea of how long the students took to make their decisions. The box plot in Figure 5 shows two distributions (both in seconds):

- how long the participants spent looking at the description of our study; and
- how long the participants spent filling in the questions.

The summary statistics for these distributions are:

Page view time (s)	Min.	1st Q	Med.	3rd Q	Max.
Description	1	32	53	70	213
Questions	10	27	45	64	(9285)

5 Conclusion

Previous studies on the economics of privacy have touched upon the fact that subjects of surveys may not report their preferences truthfully [14]. Financial incentives have been used before, but only as an enticement to complete a survey; and lab auctions have also been used [15], but as a means of valuing public goods rather than to ensure truth-telling. In this experiment we applied auction techniques in a context of experimental economics and psychology to give participants a motive to report truthfully the amount of compensation they would need to have their location monitored. Since the subjects should not have been aware of the natures of the experiment, and should have expected a low but nonzero probability of a real invasion of privacy, we expect more accurate results. This provides a sound theoretical basis for our work.

Looking at the initial bids of participants in our experiment, it is clear that the median bid of £10 is neither trivial nor particularly large. The distribution shows a wide spread: 11 students (15% of our sample) bid one pound or less, and the maximum was £400. When the possibility of commercial interest was mentioned, the median bid of those who were still willing to participate increased by £10.

The results suggest that those students who travel outside Cambridge more often value their location privacy more (the median bid of those who travelled outside Cambridge once a week or more was £20). It also appears that those who communicate with partners using their phones (presumably all of those who have a partner) have a higher valuation, although in this case the evidence from our study is less compelling.

It would be useful to extend this work to a larger sample, but we believe that in performing this study with Computer Science undergraduates we have established a good estimate of the location-privacy valuations of a subset of the population. The level of dropout at the authentication stage indicates that there was probably some selection against those with a very high privacy valuation, so the distribution of valuations may have an even longer tail than shown by our results.

Finally, students are likely to have lower privacy preferences than the general population, being generally unmarried, with few responsibilities, and in a highly tolerant environment. Thus our results may provide a lower bound on the location-privacy preference that would be stated by the population as a whole. If the preference revealed by the commercial failure of location-privacy services is substantially different, then this will highlight an interesting area for future study.

References

1. Stalder, F.: The voiding of privacy. Sociological Research Online **7** (2002)
2. Whitten, A., Tygar, J.D.: Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0. In: The 8th USENIX Security Symposium, Washington, D.C., USA (1999)
3. Odlyzko, A.M.: Privacy, economics, and price discrimination on the internet. In Sadeh, N., ed.: International Conference on Electronic Commerce (ICEC 2003), ACM Press (2003) 355–366
4. Acquisti, A., Dingledine, R., Syverson, P.F.: On the economics of anonymity. In Wright, R.N., ed.: Financial Cryptography. Volume 2742 of Lecture Notes in Computer Science., Springer (2003) 84–102
5. Acquisti, A., Grossklags, J.: Privacy and rationality in individual decision making. IEEE Security & Privacy (January/February 2005) 24–30
6. BBC News: Pay-as-you-go car cover launches (2005) `http://news.bbc.co.uk/1/hi/business/4167347.stm`.
7. Pascual, A.E.: Privacy in the next generation Internet: Data Protection in the context of the European Union Policy. PhD thesis, Royal Institute of Technology – KTH / IMIT (2004)
8. Eagle, N., Pentland, A.: Reality mining: Sensing complex social systems. In: Personal and Ubiquitous Computing. (2005)
9. Jackson, I.: Who goes here? – Confidentiality of location through anonymity. PhD thesis, Cambridge University (1998) `http://www.chiark.greenend.org.uk/~ijackson/thesis/`.
10. Vickrey, W.: Counterspeculation, auctions and competitive sealed tenders. The Journal of Finance **16(1)** (1961) 8–37
11. Warbrick, J.: The Raven web authentication service (2004) `http://www.cam.ac.uk/cs/raven/`.
12. Reips, U.D.: Standards for Internet-based experimenting. Experimental Psychology **49(4)** (2002) 243–256
13. British Psychological Society: Code of conduct, ethical principle & guidelines (2004) `http://www.bps.org.uk/`.
14. Camp, L.J., Lewis, S., eds.: What Price Privacy? In: Economics of Information Security. Kluwer (2004) 129–142
15. Mäler, K., Vincent, J., eds.: Experimental Methods and Valuation. In: Handbook of Environmental Economics. North-Holland (2001)