Uncertainty, Ambiguity and Privacy

Alessandro Acquisti
H. John Heinz III School of Public Policy and Management
Carnegie Mellon University
acquisti@andrew.cmu.edu

Jens Grossklags
School of Information Management and Systems
University of California at Berkeley
jensg@sims.berkeley.edu

March 6, 2005

Submitted to the 4th Annual Workshop on Economics and Information Security (WEIS 2005)

Abstract

In this paper we discuss the importance of ambiguity, uncertainty and limited information on individuals' decision making in situations that have an impact on their privacy. We present experimental evidence from a survey study that demonstrates the impact of framing a marketing offer on participants' willingness to accept when the consequences of the offer are uncertain and highly ambiguous.

1 Introduction

Incomplete information is relevant to privacy for two reasons. The first and perhaps most obvious reason is inherent to the very concept of privacy: an individual has some control on the level of access that other entities can gain on his personal sphere. For example, a subject's personal data may be concealed from other people's knowledge. Other people will thus rely only on incomplete information when interacting with the subject. This is the interpretation of privacy as "concealment" (of job-relevant skills, valuation for a product, creditworthiness, etc.) that [Pos78] and most subsequent formal economic models have recognized.

Secondly, incomplete information affects the individual herself whenever her control on her personal sphere is limited and in fact ambiguous. For example, a subject may not know when another entity has gained access to or used her personal information, nor may she be aware of the potential personal consequences of such intrusions. The ambiguity about the amount of control a person can exercise on her own privacy has become increasingly common in highly networked, digitized, and interconnected information societies. In this context, [Var96] noted that an individual has little or no control on the secondary use of her personal information, and hence may be subject to externalities whenever other parties transact her personal data.

This second sense in which incomplete information determines ambiguity that relates to privacy is, therefore, not new in the economic or legal literature on privacy. However, links between that literature and economic research on incomplete information (and risk, uncertainty, or ambiguity) have been surprisingly limited. So have also been formal or empirical analysis of the impact of ambiguity on privacy decision making.

Since [Kni21], economists have debated the differences between risk and uncertainty, and between the objective or subjective nature of probabilities. This research is instrumental in the understanding of decision making under uncertainty in both the descriptive and the normative sense. It is also important to privacy decision making, in that most privacy-sensitive decisions are taken by subjects under incomplete information and with significant uncertainties about the consequences of their actions.

In this paper, we discuss the link between the privacy and the uncertainty and ambiguity literature. In Section 2, we relate the analysis of the role of uncertainty over outcomes in human decision making (as discussed in the economic and, more recently, marketing literature) to the relevance of ambiguity about outcomes and probabilities in privacy decision making. In Section 4 we narrow the analysis to a simple scenario and present results from a survey study of individuals' behavior in a highly uncertain and ambigu-

ous privacy-relevant choice situations. In particular, we show that even small changes to the framing of a marketer's offer lead individuals to alter their choice behavior significantly even though the underlying consequences are not modified. We conclude the paper in Section 5.

2 Ambiguity and Privacy

The valuation of privacy is a complex and still little understood issue. Prior studies have shown that its interpretation varies not only across cultures (e.g., [Ser99]), but also from one individual to another (e.g., [Wes91]), and, for a given individual, according to the context in which privacy issues arise (e.g., [ACR99]).

Recent work on the economics of privacy has in fact highlighted a series of difficulties that hamper individual privacy decision making: incomplete information, bounded rationality, and systematic deviations from rationality seem to significantly affect privacy attitudes and privacy behavior. But those difficulties are actually larger categories that cover up a myriad of factors that are sometimes in contradiction with each other and hard to disentangle. While in prior works we proposed overviews of those factors that influence privacy decision making (see [AG03], [Acq04], and [AG05]), in this paper we restrict the discussion to the role of incomplete information about outcomes and probabilities associated with those outcomes. In particular, we relate the problem of privacy decision making to the research literature in the field of risk, uncertainty and ambiguity.[1]

The distinction between risk and uncertainty in economics dates back to [Kni21] (although earlier discussions of the relations between risk, uncertainty, and utility can be found with [Ber38] and then [Men71]). Knight proposed to distinguish situations characterized by risk (in which the possible random outcomes of a certain event have *known* associated probabilities) from those characterized by uncertainty or ambiguity (in which the randomness cannot be expressed in terms of mathematical probabilities, and the probabilities themselves are *unknown*). For example, the expected utility theory of [vNM44] is based on objectively knowable probabilities (what [Kni21] would have referred to as 'risk').

However, this distinction has not gone unchallenged by economic theorists and statisticians. A large body of literature suggests that individuals are always able to assign reasonable probabilities to random events. These probabilities could objectively exists in the world, as proposed by [vNM44], and could be used to calculated expected utilities. Or, these probabilities could be *subjective*, as proposed by [Sav54].

[1] Before we proceed we want to note that economists, psychologists, and marketers often use terms like risk and uncertainty in different ways. Even within the same discipline researchers disagree on the interpretation given to terms like 'uncertainty.'

Savage adapted expected utility theory into a theory of subjective expected utility, in which, under certain assumptions, people will have personal beliefs about the possible states of nature.

The concept of subjective probabilities establishes a bridge between the concept of risk and uncertainty, since the known probability (of a risk) is set on par with a subjective belief. Prescriptively, decision theory and mainstream economic theory of expected utility have incorporated the idea that knowledge (or subjective belief) of the actual risks associated with different events and decisions will drive the actions of an economic agent. An economic agent will consider a set of possible actions with different outcomes, probabilities over these outcomes, and associated utilities. He will then choose a strategy consisting of a series of actions leading to the highest expected utility.

Descriptively, however, in numerous situations (such as privacy decision making and other complex scenarios) it is unrealistic to assume existence of known or knowable probabilities or complete beliefs for probabilities over all possible outcomes. Additionally, even if individuals were served with sufficient data about outcomes and associated probabilities, they may still tend to use this data in ways which are different from that of expected utility maximization. Individuals tend to make sometimes paradoxical, surprising, and seemingly contradictory decisions (see, for example, [KT00] and [Ell01]). For example, individuals are often not only risk averse but also *ambiguity averse* (see [CW92]). Given the choice between a certain outcome (e.g., $10) and a lottery over outcomes (e.g., $0 with 50% likelihood and $X with 50% likelihood), individuals prefer the certain choice unless they are offered a premium in the lottery so that the expected value of the lottery is greater than the certain outcome (e.g., X strictly greater than $20).

Behavioral economists and psychologists have worked on modifications of the theories of risk and uncertainty to produce satisfactory descriptive models of human decision making under incomplete information.[2] For example, [HK92] suggests to focus on subjective *weights* associated to the various possible outcomes of a certain event - where the weights do not have the same mathematical properties as probabilities. [HK92] proposes that 'decision weights' may be obtained by the individual through a process of anchoring and adjustment. First, an individual may anchor her value on an initial estimate of probabilities over outcomes. Then, she would adjust such an estimate after mentally simulating alternatives values. This adjustment may be influenced by the degree of ambiguity and by the size of the outcome (e.g. whether the gain or loss is large or small). Alternative research has, however, shown that the process of initial anchoring can be subject to substantial manipulation (see, for example, the research on coherent arbitrariness by [ALP03]). Further-

[2] [CW92] reviews experimental evidence and formal modeling work on ambiguity in great detail.

more, there is evidence that competence and knowledge affect individuals' choices. People prefer to bet on events they know more about, even when their beliefs are held constant (see [HT91]).

Privacy decision making belongs to those complex scenarios in which information is incomplete and therefore risk, uncertainty, and ambiguity play a crucial role. When presented with a privacy-related problem, individuals often face two major unknowns: a) what privacy-relevant outcomes may occur under different contexts; and b) with what consequences [AG05]. Implicit in these two major unknowns there are, however, layers over layers of additional uncertainties.

First, an individual has often only vague and limited knowledge of the actions she can take to protect (or give away) her personal information. She has also limited knowledge of the possible actions of other entities (e.g., a marketer's purpose and means to collect information).

Second, actions taken by the individual (whether as an attempt to protect or trade information) or another party have often hardly predictable consequences.

Third, possible relevant states of nature (with associated additional actions and consequences) may be unknowable in advance, because they depend of future events and changes in the system (e.g., a technology development such as private information retrieval [CGKS95] or Google caching and making searchable old Usenet archives).

Fourth, certain desirable actions and information may not be available (see research on asymmetric information and hidden action). For example, an individual might have limited control and knowledge of what actions are actually taking place that relate to her privacy.

Fifth, we observed in prior work that individuals iteratively uncover additional layers of a choice situation revealing further actions and outcomes, with their sets of associated (possible) values and (possible) probabilities. For example, in [AG05] we describe how people change their perception on which parties have access to their credit card transactional data if they are prompted with this topic repeatedly. We show that individuals sometimes ignore both privacy risks and forms of protection, and when they are aware of them, often miscalculate their probability of occurrence and their numerical outcome in terms of financial magnitude. This carelessness or ignorance might be justifiable if one considers the effort needed to evaluate everyday privacy choices carefully.

Sixth, privacy protection or invasion are often by-products of unrelated transactions. The privacy 'good' is often attached to other goods in complex bundles - or, in other words, trade-offs involving privacy are often trade-offs involving heterogeneous goods and incommensurate values. For example, when an individual

purchases a *book* online (thus saving the *time* she would have spent going to the bookstore and paying in cash), she will often reveal her *credit card details* to the online merchant, which may lead to an increased risk of *identity theft*. Or, in order to receive a *monetary discount* from the *grocery store*, she will reveal her *buying patterns* by using a loyalty card, which may increase her probability of receiving *junk mail* or undesired commercial offers.

Comparisons between those different goods are difficult because of their combinatorial aspects, but may be further complicated if the offers are uncertain or ambiguous. The marketing literature has recently become interested in in scenarios where the underlying values are incommensurate. [NP03] define this as different forms of wealth which are difficult to "convert into any meaningful common unit of measurement." For example, they study a promotion that is presented in nonmonetary terms (e.g., an umbrella). Under these conditions, the marginal value of the nonmonetary, incremental benefits becomes difficult to evaluate for the individual, in relation to the focal product or its price. Similarly, [DN04] explore how individuals evaluate transactions that involve prices issued in multiple currencies and show that under certain conditions the price the individual is willing to pay when payments delivered in different currencies is superior to a standard, single-currency price. This is an interesting finding that challenges the common observation that individuals are predominantly ambiguity averse. If offers are formulated in incommensurate terms, individuals might in fact prefer uncertain choices.

Because of these various layers of complexity, we conclude that an individual who is facing privacy sensitive scenarios may be uncertain about the values of possible outcomes and their probability of occurrence, and that sometimes she may not even be able to form any beliefs about those values and those probabilities. In fact, she may have no knowledge of the possible *outcomes* of a certain situation - the states of nature may be not known or knowable in advance (in [AG05] we show that individuals sometimes ignore both privacy risks and forms of protection).

By considering these complexities and the ramifications of privacy decision making, it becomes difficult not to embrace [Kni21]'s separation of risk and uncertainty and favor a model of individuals' decision making with respect to privacy where decisions are at least partly based on unknown outcomes (uncertainty) with unknown probabilities over outcomes (ambiguity).[3]

As a first step in ongoing analysis, our interest is how individuals evaluate offers that combine certain payments (such as monetary discounts on a product) with uncertain and ambiguous outcomes (such as the

[3]We will use these definitions for the remainder of the paper.

annoyance of telemarketing based on the release of personal information).

3 Motivation for Survey Research Questions

In Section 2 we presented related research with respect to limited information and discussed its applicability to privacy choice behavior. We want to highlight two aspects of this research summary. On the one hand, there is ample evidence that individuals attempt to avoid ambiguity and need to be paid an ambiguity premium to participate in lotteries that are otherwise comparable to a certain payment (see, for example, [CW92]). On the other hand, the marginal value of nonmonetary, incremental benefits (such as discounts in a purchase scenario) becomes difficult to evaluate for the individual, in relation to a focal product and its price if the benefit is presented in an incommensurate form [NP03]. In fact, if a premium is presented as an incommensurate resource individuals might even prefer an option that represents an ambiguity penalty.

In Section 4 we do not directly attempt to verify which of these two hypothesis is more relevant to privacy choice behavior. Rather we are interested in the following two related questions. First, how do individuals value personal information if presented with a highly ambiguous and uncertain marketer's offer? Second, can we manipulate this valuation (to either increase or decrease) by reframing the scenario to involve a discount relatively to a product's purchase value (that is, we present a framing that stimulates relativistic processing [KT84])?

To further motivate our analysis in Section 4 consider the following sweepstakes advertisement that we found at San Francisco's Pier 39 (see Figure 1). The advertisement offers passers-by the opportunity to participate in a sweepstakes organized by an entity that is likely unknown to potential participants (Grand Pacific Resorts; Promotion Department). On the day this advertisement was collected a car was showcased. The offer's small-print gives an approximate winning probability for the 'grand price' of 1 out of 700000 for a value between $15000 and $25000. Each participant is asked to provide contact information (that obviously needs to be accurate) and further data about the individual and the spouse such as age, job description, combined income, house ownership, basic credit card information and an email address (for which no further verification was required). The advertisement also gives the opportunity to opt in for further Pier 39 marketing offers.

The small-print further informs the reader that all data will be owned by a second entity (PNR Marketing Inc.). Additionally, participants of the offer may be contacted to participate in sales promotions of a third entity (about Red Wolf Lodge at Squaw Valley). No further information is provided about the entities that

are acknowledged to use the information. Further trade to other parties is not excluded.

Free Giveaway!

Name: _____

Address: _____

City: _____ State: _____ Zip: _____

Home Phone: _____

Work Phone: _____

❑ Single ❑ Married

Age: _____ Occupation: _____

Spouses Age: _____ Occupation: _____

Combined Income:
❑ Under $30,000 ❑ Over $30,000 ❑ Over $50,000

DO YOU: ❑ RENT OR ❑ OWN YOUR HOME?

❑ VISA ❑ MASTERCARD ❑ AMERICAN EXPRESS

Would you like info on special events & promotions at Pier 39?

❑ Yes ❑ No

E-mail address: _____

Details of Participation and Eligibility Requirements

- Only one Entry per Family.
- Winner allows the use of his or her name, photo, and statements for future promotional use without further compensation.
- Winner must be 18 or over. I.D. required. Winner must provide all necessary federal and state tax reporting information before receiving prizes.
- Drawing held February 23, 2003. Last date to enter drawing is February 16, 2003.
- Winner need not be present to win. Winner will be notified by phone.
- Drawing will be conducted by a Certified Public Accounting Firm at the corporate office of Grand Pacific Resorts, 5900 Pasteur Ct., #200, Carlsbad, CA 92008. To request winner information, correspondence may be forwarded to Grand Pacific Resorts, Promotions Dept. P.O. Box 4068, Carlsbad, CA 92018.
- All local, state, and federal taxes, fees and licenses are the winner's responsibility. Acceptance of the prizes constitutes a release of Facility Management, it's agents and employees from all responsibility to the winner.
- Odds are based on number of entries received, approximately 1 in 700,000.
- No purchase or attendance is necessary to be entered into the drawing. Entrants may be invited to attend a sales presentation about the Red Wolf Lodge at Squaw Valley.
- Entries become the property of PNR Marketing Inc.
- The annual "Grand Prize" Giveaway consists of any vehicle with a retail value not to exceed $25,000 or a three year lease (value to $25,000) on a luxury car; or any prize (or similar) displayed in a Grand Pacific Resorts Promotion February 25, 2002 - February 23, 2003 (valued up to $15,000), or the winner may choose cash in the amount of $15,000.

Figure 1: **Sweepstakes offer from Pier 39 in San Francisco**

It is difficult to exactly predict the expected benefits associated with this offer (given the information provided we estimate them to be at most 3.5 cents). A further nonmonetary benefit could be the thrill to participate in sweepstakes (that, however, would have to be weighted with the possible disappointment if the individual does not win). The negative consequences are telemarketing and other forms of contacts regarding sales promotions. Furthermore, nothing is known about the further use of the collected data and the resulting consequences.

We do not want to make a qualitative statement whether this offer should or should not be accepted by an individual. We, however, are interested in testing how people evaluate such an offer and react to subtle changes in its presentation.

4 Analysis of Survey data

We first present the experimental methodology we used in our survey before discussing the outcome of the survey.

4.1 Procedure

In May 2004, we contacted potential subjects who had shown interest in participating in economic studies at Carnegie Mellon University. We offered participants a lump sum payment of $16 to fill out an online, anonymous survey about e-commerce preferences and gathered 119 responses. We used the term "e-commerce preferences" to mitigate self-selection bias from pre-existing privacy beliefs. The survey contained several questions organized around various categories: demographics, a set of behavioral economic characteristics (such as risk and discounting attitudes), past behavior with respect to protection or release of personal information, knowledge of privacy risks and protection against them, and attitudes toward privacy. We discuss only a subset of questions in this article; the full survey is available online at `http:\\www.heinz.cmu.edu/~acquisti/survey/page1.htm`. This survey was funded by the Berkman Faculty Development Fund.

4.2 Basic Demographics

Participants ranged from 19 to 55 years old (with the mean age of 24). Eighty-three percent were US citizens; the remainder were from heterogeneous backgrounds. More than half of our subjects worked full or part time or were unemployed at the time of the survey, although students represented the largest group (41.3 percent). All participants had studied or were studying at a higher education institution. While our population of relatively sophisticated individuals is not an representative sample of the U.S. population., we are able to exhibit trends we believe are quite general.

Most participants had personal and household incomes below $60,000. Approximately 16.5 percent reported household income above that level, including 6.6 percent with an income greater than $120,000. Most respondents are also frequent computer users (62.0 percent spend more than 20 hours per week) and Internet browsers (69.4 percent spend more than 10 hours per week) and access computers both at home and work (76.0 percent).

4.3 Privacy and Independent Private Values

Individuals have different monetary and non-monetary values associated with the personal information. In this section, we attempt to explore the differences in monetary valuation when individuals are faced with a nonspecific 'information purchase' offer by a marketer similar to the offer in Figure 1. We asked individuals the following 'Question 1':

> "Suppose a marketing company wants to buy your personal information. You do not know and you cannot control how the company will use that information. You know that the company will effectively own that information and that information can be linked to your identity. For how much money (in U.S. dollars) would you reveal the following data items to this company: (if you would never reveal that information, write 'never')."

We asked individuals to reveal their valuation for 13 data categories: a) Full name, b) Social Security Number (SSN), c) Favorite online user name, d) Physical home address, e) Phone number, f) Email address, g) Job title and job description, h) Interests outside work or study, i) Previous and current health history, j) Statistics about personal email received and sent (e.g., keywords, names, places), k) Actual content of personal emails, l) Rights to future health history, and m) Description of sexual fantasies. People were asked to indicate a dollar amount they would accept or write down 'never' to categorically refuse such an offer.

In Question 1 we avoided giving specific details about the purpose for which the information is collected or the positive or negative consequences the survey participants could expect.[4]

We expected to find substantial differences for the data categories and between survey participants. Table 1 presents a classification of the response data into 3 categories: values requested below $500, values above $500, and the number that responded with 'never'.[5] By observing the number of participants that responded with 'never,' we can distinguish different degrees of sensitivity for the collection of the 13 data categories. First, people are most often reluctant to divulge their SSN, which is likely a result from recent extensive media coverage of the frequency and consequences of identity theft and financial fraud. Second, health information and personal email are considered precious and less tradable by participants. Third, the least rejection frequency we observe for data that is often publicly available such as the full name, email address, interests and basic job information. Still, given our scenario in Question 1, some participants appear very hesitant to release information that could be part of almost every day 'small-talk' with strangers.

[4]We believe the exact phrasing of our question weakened the likelihood of assigning a positive benefit to the information collected. As an alternative example consider 'whistle-blowing' where an individual has a preference or incentive for divulging some information at a personal cost. That is, a participant would be willing to pay to communicate this information. In our survey, no such responses were recorded.

[5]We chose this cutoff level since it enables us to compare responses for Question 1 with those reported in later parts of this section.

Data	Valuation <= 500	Valuation > 500	'Never' released	Total
a) Full name	78	20	21	119
b) SSN	4	16	97	118
c) Online name	74	18	26	118
d) Home address	60	24	34	118
e) Phone number	56	28	34	118
f) Email address	78	24	16	118
g) Job description	92	15	11	118
h) Interests	98	11	9	118
i) Previous health	38	32	48	118
j) Email statistics	34	31	52	117
k) Email contents	22	29	67	118
l) Future Health	7	33	78	118
m) Sexual Fantasies	37	32	49	118

Table 1: **Question 1 - Response frequencies for the different data categories** Frequencies are reported for individuals that would 'never' release personal information, those that ask for a compensation of less than $500, and those that request a compensation higher than $500

We subjected this data to a statistical test to evaluate whether the rejection frequencies differ significantly for the different data categories. We applied McNemar's test; a non-parametric test that uses matched-pairs of labels (A, B; here: 'reject offer', 'accept offer') and determines whether the proportion of A- and B-labels is equal for both labels (labels are the thirteen data categories). It is a very good test when only nominal data are available. In Table 2 we list significance levels for those combinations that are significantly different from each other. This data enables us to rank the information categories on the basis of their rejection frequency ('never') and to form groups of data that are treated alike by our participants (usage of brackets denotes membership to a group). The only data that is not unequivocally categorizable is 'favorite online name'. The resulting rank order (lowest rejection frequencies first) is presented below:

(interests and job [and favorite online name]) < ([favorite online name and] email and full name) < (home address and phone number) < (Previous health history, sexual fantasies, and

Email statistics) < (Email contents) < (Future health history) < (SSN)

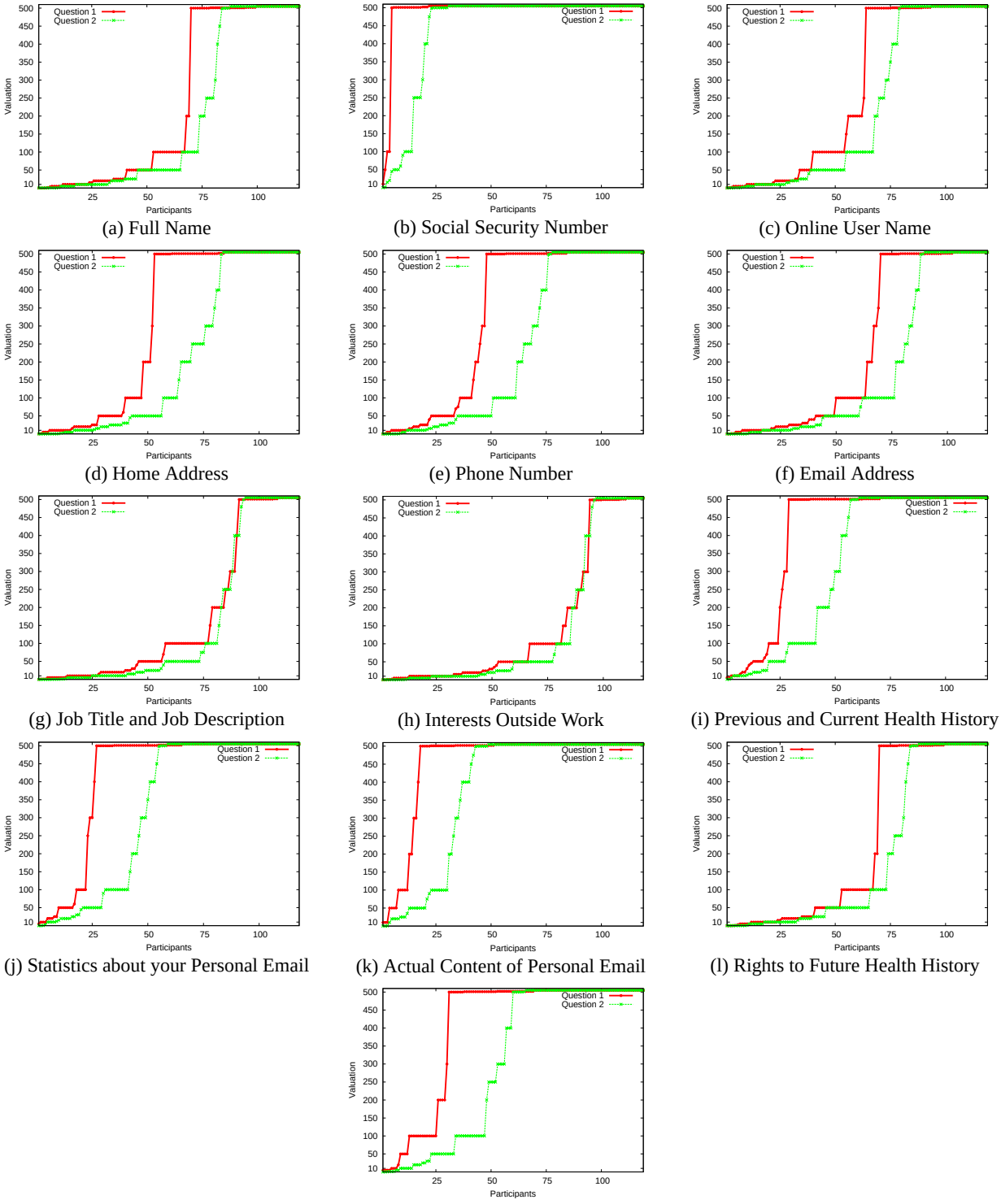
This ranking shows that participants treat their physical home address and their phone number similarly. Otherwise, despite being very sensitive with respect to the release of SSN, future health history and email contents our participants value these information significantly different from one another.

	b)	c)	d)	e)	f)	g)	h)	i)	j)	k)	l)	m)
a)	.00001		.0006	.0027		.029	.0116	.00001	.00001	.00001	.00001	.00001
b)		.00001	.00001	.00001	.00001	.00001	.00001	.00001	.00001	.00001	.00001	.00001
c)				.0947	.0389	.0011	.00001	.0002	.00001	.00001	.00001	.0001
d)					.0001	.00001	.00001	.0196	.0044	.00001	.00001	.0137
e)					.0001	.00001	.00001	.0164	.0035	.00001	.00001	.0071
f)							.0455	.00001	.00001	.00001	.00001	.00001
g)								.00001	.00001	.00001	.00001	.00001
h)								.00001	.00001	.00001	.00001	.00001
i)										.0001	.00001	
j)										.0011	.00001	
k)											.0076	.001
l)												.00001

Table 2: **Question 1 - Significance levels for McNemar's chi-square test** Data categories: a- Full name, b- SSN, c- Favorite online user name, d- Physical home address, e- Phone number, f- Email address, g- Job title and job description, h- Interests outside work or study, i- Previous and current health history, j- Statistics about personal email, k- Content of personal email, l- Rights to future health history, and m- Description of sexual fantasies. Insignificant values are omitted.

Another observation from Table 1 is that the number of participants that requests high dollar amounts ($>$ \$500) for a data item is fairly constant. There are between 11 (for Interests) and 33 (for Future Health) of those individuals for the different categories. In contrast, the number of participants that rejects the marketer's offer varies more drastically between 9 (for Interests) and 97 (for SSN). Rather than seeking money as compensation for divulging personal information those participants reject data collection categorically in the setting of Question 1.

Next, we discern the degree of variation between individuals. Figure 2 displays the requested amounts in U.S. dollars for the different data categories. For readability purposes, for each category we plotted the



(a) Full Name (b) Social Security Number (c) Online User Name

(d) Home Address (e) Phone Number (f) Email Address

(g) Job Title and Job Description (h) Interests Outside Work (i) Previous and Current Health History

(j) Statistics about your Personal Email (k) Actual Content of Personal Email (l) Rights to Future Health History

(m) Description of Sexual Fantasies

Figure 2: **Valuation for Data Items for Question 1 and Question 2** (Valuations are depicted in increasing order for both questions; Participant ID's are not matched pairs; Valuations higher than 500 are shown as 501; Valuations of "no" and "never" are shown as 505)

requested amounts in increasing order (thereby reordering the participants' IDs) and printing values larger than 500 and 'never' as a constant (500 + small epsilon). We observe that the data exhibits a high variability in the outcomes. For all categories the minimum amount requested is between zero (for Full name, Phone number, Email address, Interests and Job description) and ten (for SSN). The maximum numerical amount ranges from \$100000 (for Home address, Email address, Interests and Job description) to $\$10^{21}$ (for Email contents).[6]

In general, we believe that the data we gathered follows the independent private value model [Vic61] where the value of a data item is dependent on (multiple) private signals. Such signals could be a) evidence on an expected financial loss, b)fairness considerations, c) previous experiences with data trades etc. The dispersion arises also because no unified resale value is known. Anecdotal evidence suggests that, for example, email addresses can be bought as bulk data for limited amounts of money. However, a valuation of the contents of personal email is type dependent and potentially immeasurable. That does not mean, however, that a marketer would not be willing to pay a certain price for this information. Google Gmail, for example, aims to receive revenue from advertisements based on statistical information about individuals' personal email. Google's ability to execute this practice might also be dependent on general consumer trust towards this brand, however, we can cite their business proposal as an example to indicate that for most of the information categories in this study there exists indeed a market. Our participants were also generally worried (as remarked in free text responses) that health information can be used for the variation of insurance premiums. Individuals can base their evaluation on their private signals and potentially on public information about marketers' demands and willingness-to-pay, however, the formulation of a fair price for her information will likely remain uncertain and ambiguous.

Likewise, because of the general concern that individuals have for information collection and information use outside their control that is ambiguous and uncertain marketers could find it difficult to gather such information at a reasonable price. In the next section, we test whether individuals' perception for the valuation of information can be influenced. That is, can we tempt individuals to accept an offer more often even though the degree of ambiguity and uncertainty is similar.

[6] A very small number of individuals asked for very high compensations in exchange for their personal information.

4.4 Valuation of Personal Information in a Purchase Scenario

We asked individuals a modified version of Question 1 where they were put into a different frame. We asked participants to specify a discount for which they were willing to accept a marketer's offer. The discount was on the purchase of a product with a fixed value ($500). In exchange for the discount the marketer asked for the same information as in Question 1. We asked individuals the following (we refer to this as 'Question 2'):

> "Would you provide this information for a discount on an item you want to purchase or service you want to use? The items value is $500. If yes, what discount (in US dollars) would you expect? If you would not provide this information please enter 'no'."

We expected that those participants that were asking for an amount larger than $500 in the scenario of Question 1 would now reject the marketer's offer and would respond 'no'. Otherwise everything we believed would stay unchanged. In particular, we believed that having the questions located close to one another in the questionnaire would support consistency of the subjects' responses.

In Figure 2 we display the data for Question 2 in similar fashion as for Question 1. We represent 'no' with a value marginally above 500 (500 plus small epsilon). At first glance we observe that *more individuals seem to accept the marketer's offer in Question 2 compared to those in Question 1 that are satisfied with a compensation below $500*. In this section we test this observation statistically.

First, we subject our observation to a McNemar non-parametric test for matched cases. Question 1 is coded as follows: 'accept = 1' for all those individuals that requested less than $500; 'reject = 0' for all others including those that said never. Question 2: 'accept = 1' for all individuals that accepted the offer and specified a discount; 'reject = 0' for all individuals that wrote 'no'. Table 3 demonstrates that in 10 out of 13 data categories significantly more individuals are willing to accept the offer of the marketer in Question 2.[7]

We also test whether individuals' monetary valuation differs between the two offers. In Table 4 we examine those individuals that reported a monetary valuation for both Question 1 and 2. For some data categories the frequency numbers are very low (see, for example, SSN and Future health history). However, for the majority of data categories the valuation in Question 2 is lower than for Question 1. Of the ten data categories for which we have at least 20 observations, seven categories are shown to be significant for the

[7] Two of those categories have the lowest rejection rates for both marketer's offers (Job description and Interests). We believe that the few participants that refused to trade this personal information are privacy fundamentalists that are difficult to influence by our offer variation.

Data	Number of Observations	Significance for McNemar chi-square	Exact McNemar
a) Full name	119	0.0956	0.1325
b) SSN	119	0.00001	0.00001
c) Online name	119	0.4497	0.5716
d) Home address	118	0.0003	0.0004
e) Phone number	119	0.0010	0.0015
f) Email address	118	0.0396	0.0576
g) Job description	118	0.8618	1.0000
h) Interests	118	0.7150	0.8555
i) Previous health	119	0.0005	0.0007
j) Email statistics	118	0.0003	0.0004
k) Email contents	118	0.00001	0.00001
l) Future Health	119	0.00001	0.00001
m) Sexual Fantasies	119	0.00001	0.00001

Table 3: **Question 2 - Significance levels for McNemar's chi-square test**

Wilcoxon Match-Pairs Signed Ranks Test, and six for the more general Signtest for matched pairs. We can also firmly reject the hypothesis that the valuation for Question 2 is larger for any of the data categories (see second column for the signtest).

In Table 5 we report the same statistics but include all observations. The coding for Question 1 is as follows: actual valuation for all reported amounts $<=$ \$500, all values above \$500 and those reported as 'never' are included as 505; Question 2: all reported values reported as 'no' are coded 505. One limitation of our data is that we are not having a numerical valuation for all participants. Some responded with 'no' or 'never'. Our general results do not hinge on the analysis reported in Table 5, however, we attempt to make all data usable by coding the data as described. We believe this to be a fairly conservative, nevertheless imprecise analysis. The underlying assumption is that people who categorically refused to participate in the marketer's offers in either Question 1 and 2 would value the information at least with \$500. We find in the Wilcoxon Match-Pairs Signed Ranks Test 12 out of 13 categories to have a significant effect between Question 1 and Question 2; for the Signtest we must reject our observation that the discount in Question 2

Data	Number of Observations	Signtest Q1 - Q2 > 0	Signtest Q1 - Q2 < 0	Wilcoxon test
a) Full name	64	0.1659	0.8942	0.1646
b) SSN	3	0.75	0.75	0.78
c) Online name	63	0.28	0.80	0.2
d) Home address	51	0.0298	0.9862	0.0118
e) Phone number	46	0.00001	1	0.00001
f) Email address	66	0.0314	0.9834	0.0212
g) Job description	76	0.0038	0.9984	0.0018
h) Interests	82	0.0124	0.9937	0.0035
i) Previous health	29	0.1050	0.9534	0.0305
j) Email statistics	24	0.3318	0.8083	0.3303
k) Email contents	17	0.1938	0.9270	0.0971
l) Future Health	4	0.6875	0.6875	0.7150
m) Sexual Fantasies	34	0.0012	0.9997	0.0009

Table 4: **Question 2 - Significance levels for Signtest and Wilcoxon Match-Pairs Signed Ranks Test** Only observations with numerical values for both Question 1 and 2

is lower than the valuation in Question 1 only 2 times (Full name and Favorite online name).

We suggest that there is strong support for the fact that individuals react to the framing of Question 2 by lowering their demands significantly and reject the marketer's offer less often.

5 Discussion and concluding remarks

In this paper we discussed the importance of ambiguity, uncertainty and limited information on individuals privacy decision making. We also show experimental evidence from a survey study that demonstrates the impact of framing a marketing offer on participants' willingness to accept when the consequences of the offer are uncertain and highly ambiguous.

In our first offer (Question 1) subjects were asked to specify an amount at which they would be willing to sell certain information or to reject the offer entirely. We did not specify who exactly is collecting this

Data	Number of Observations	Signtest Q1 - Q2 > 0	Signtest Q1 - Q2 < 0	Wilcoxon test
a) Full name	118	0.1066	0.9269	0.0446
b) SSN	118	0.0003	0.9999	0.0002
c) Online name	118	0.4561	0.6297	0.2384
d) Home address	117	0.0004	0.9998	0.00001
e) Phone number	118	0.00001	1	0.00001
f) Email address	117	0.0117	0.9933	0.0028
g) Job description	117	0.0235	0.9860	0.0367
h) Interests	117	0.0501	0.9679	0.0666
i) Previous health	118	0.0320	0.9815	0.0045
j) Email statistics	117	0.0550	0.9662	0.0037
k) Email contents	118	0.0137	0.9933	0.0027
l) Future Health	118	0.0290	0.9850	0.0100
m) Sexual Fantasies	118	0.00001	1	0.00001

Table 5: **Question 2 - Significance levels for Signtest and Wilcoxon Match-Pairs Signed Ranks Test** All observations; Valuations larger than 500 and reported as 'never' and 'no' are included as 505

information and for what purpose. Our second offer changed the frame to a decision over joint outcomes. Participants were asked whether they would accept or reject the offer. If they would accept they again had to specify the minimum discount on a $500 product they would expect in exchange for personal information. We observed that participants requested lower monetary amounts under the condition of the second offer, and rejected the offer less often.

If we want to explain these results we have to take different effects into consideration. Firstly, when presented in a purchase situation with an additional opportunity for a discount individuals will value this discount relatively to a product's purchase value (an observation known as relativistic processing [KT84]). In contrast, if consumers would have a stable valuation for private information we should not observe behavioral differences between the two offers that are discussed in this paper. In fact, individuals that reported in Question 1 a valuation for their personal information of more than $500 should not accept the discount. This hypothesis is rejected by our analysis in Table 5. Individuals, in fact, lower their demands significantly

if presented with a discount.

Secondly, limited information, ambiguity and uncertainty increase the difficulty to predict consequences of releasing personal information. As a result, individuals are hampered in their relativistic processing. We observe that the valuation of personal information is reduced when compared with a certain benefit (as given by the dollar discount). This contradicts the general belief that individuals attempt to avoid ambiguity.

Privacy as a good differs from monetary resources or tangible goods in the sense that its valuation is based on multiple factors. For example, we expect the value of privacy to be influenced by the subjective evaluation of negative consequences of trading personal information, fairness considerations and further nonmonetary costs (such as annoyances). At another part in our survey we asked participants to specify what they would expect to loose (in U.S. dollars) if they release personal information. We found those expected losses to be smaller than the values requested by the participants in the two offer conditions. This suggests in fact that participants expect a premium (in addition to the monetary amount they expect to loose) to cover, for example, nonmonetary losses, but they will reduce their premium if they are offered a certain benefit.

References

[Acq04] Alessandro Acquisti. Privacy in electronic commerce and the economics of immediate gratification. In *Proceedings of the ACM Conference on Electronic Commerce (EC '04)*, pages 21–29, 2004.

[ACR99] Mark Ackerman, Lorrie Cranor, and Joseph Reagle. Privacy in e-commerce: Examining user scenarios and privacy preferences. In *Proceedings of the ACM Conference on Electronic Commerce (EC '99)*, pages 1–8, 1999.

[AG03] Alessandro Acquisti and Jens Grossklags. Losses, gains, and hyperbolic discounting: An experimental approach to information security attitudes and behavior. In *2nd Annual Workshop on Economics and Information Security - WEIS '03*, 2003.

[AG05] Alessandro Acquisti and Jens Grossklags. Privacy and rationality in decision making. *IEEE Security & Privacy*, January-February:24–30, 2005.

[ALP03] Dan Ariely, George Loewenstein, and Drazen Prelec. Coherent arbitrariness: Stable demand curves without stable preferences. *Quarterly Journal of Economics*, 118:73–106, 2003.

[Ber38] Daniel Bernoulli. Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientiarum Imperialis Petropolitannae*, 1738. Translated and published as "Exposition of a New Theory on the Measurement of Risk" in Econometrica 22(1): 23-36, January 1954.

[CGKS95] Benny Chor, Oded Goldreich, Eyal Kushilevitz, and Madhu Sudan. Private information retrieval. In *IEEE Symposium on Foundations of Computer Science*, pages 41–50, 1995.

[CW92] Colin Camerer and Martin Weber. Recent developments in modeling preferences: Uncertainty and ambiguity. *The Journal of Risk and Uncertainty*, 5:325–370, 1992.

[DN04] Xavier Drèze and Joseph C. Nunes. Using combined-currency prices to lower consumers perceived cost. *Journal of Marketing Research*, 41:59–72, 2004.

[Ell01] Daniel Ellsberg. *Risk, Ambiguity, and Decision*. Garland Publishing, New York and London, 2001.

[HK92] Robin M. Hogarth and Howard Kunreuther. Decision making under uncertainty: The effects of role and ambiguity. In F. Heller, editor, *Decision making and leadership*, pages 189–212, Cambridge, England: Cambridge University Press, 1992.

[HT91] Chip Heath and Amos Tversky. Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of Risk and Uncertainty*, 4:5–28, 1991.

[Kni21] Frank H. Knight. *Risk, Uncertainty, and Profit*. Hart, Schaffner & Marx; Houghton Mifflin Company, Boston, MA, 1921.

[KT84] Daniel Kahneman and Amos Tversky. Choices, values, and frames. *American Psychologist*, 39(4):341–350, 1984.

[KT00] Daniel Kahneman and Amon Tversky. *Choices, Values, and Frames*. University Press, Cambridge, 2000.

[Men71] Carl Menger. *Principles of Economics*. New York University Press (1981 edition), New York, 1871.

[NP03] Joseph C. Nunes and C. Whan Park. Incommensurate resources: Not just more of the same. *Journal of Marketing Research*, 40:26–38, 2003.

[Pos78] Richard A. Posner. An economic theory of privacy. *Regulation*, May-June:19–26, 1978.

[Sav54] Leonard J. Savage. *The Foundations of Statistics*. Dover (1972 edition), New York, 1954.

[Ser99] IBM Global Services. Multi-national consumer privacy survey, 1999.

[Var96] Hal R. Varian. Economic aspects of personal privacy. In *Privacy and Self-Regulation in the Information Age*. National Telecommunications and Information Administration, 1996.

[Vic61] W. Vickrey. Counter-speculation, auctions, and competitive sealed tenders. *Journal of Finance*, 16:8–37, 1961.

[vNM44] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press (1953 edition), Princeton, NJ, 1944.

[Wes91] Alan F. Westin. *Harris-Equifax Consumer Privacy Survey 1991*. Equifax Inc., Atlanta, GA, 1991.