Published in Operations Research
Volume 64, number 1, pages 67–78, 2016.

The Empirical Implications of Privacy-Aware Choice

Rachel Cummings and Federico Echenique and Adam Wierman [*]

April 2015

Abstract

This paper initiates the study of the testable implications of choice data in settings where agents have privacy preferences. We adapt the standard conceptualization of consumer choice theory to a situation where the consumer is aware of, and has preferences over, the information revealed by her choices. The main message of the paper is that little can be inferred about consumers' preferences once we introduce the possibility that the consumer has concerns about privacy. This holds even when consumers' privacy preferences are assumed to be monotonic and separable. This motivates the consideration of stronger assumptions and, to that end, we introduce an additive model for privacy preferences that does have testable implications.

1 Introduction

We study what an observer can learn about a consumer's preferences and behavior when the consumer has concerns for her privacy and knows that she is being observed. The basic message of our results is that very little can be learned without strong assumptions on the form of the consumer's privacy preferences.

To motivate the problem under study, consider the following story. Alice makes choices on the Internet. She chooses which websites to visit, what books to buy, which hotels to reserve, and which newspapers to read. She knows, however, that she is being watched. An external agent, Big Brother (BB), monitors her choices. BB could be a private firm like Google, or a government agency like the National Security Agency (NSA). As a result of being watched, Alice is concerned for her privacy; this concern affects her behavior.

---

[*] Affiliations: Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA 91125; Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125; Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA 91125. We thank Jeff Ely, two anonymous referees, and especially Katrina Ligett, for useful comments. This work was supported by NSF grants 1101470, 1216006, and CNS-1254169, as well as BFS grant 2012348 and the Charles Lee Powell Foundation.

Alice has definitive preferences over the things she chooses among. For example, given three political blogs, a, b, and c, she may prefer to follow a. However, BB will observe such a choice, and infer that she prefers a over b and c. This is uncomfortable to Alice, because her preferences are shaped by her political views, and she does not like BB to know her views or her preferences. As a result, she may be reluctant to choose a. She may choose b instead because she is more comfortable with BB believing that she ranks b over a and c.[1]

Now, the question becomes, given observations of Alice's behavior, what can we learn about her preferences? We might conjecture that her behavior must satisfy some kind of rationality axiom, or that one could back out, or reverse-engineer, her preferences from her behavior. After all, Alice is a fully rational consumer (agent), meaning that she maximizes a utility function (or a transitive preference relation). She has a well-defined preference over the objects of choice, meaning that if she could fix what BB learns about her—if what BB learns about her were independent from her choices—then she would choose her favorite object. Further, Alice's preferences over privacy likely satisfy particular structural properties. For example, she has well-defined preferences over the objects of choice, and she cares about the preference revealed by her choices: she always prefers revealing less to revealing more. In economics, preferences of this form are called separable and monotone; such preferences normally place strong restrictions on agents' behavior.

However, contrary to the above discussion, the results in this paper prove that nothing can be inferred about Alice's preferences once we introduce the possibility that she has concerns about privacy. No matter what her behavior, it is compatible with some concerns over privacy, i.e., she always has an "alibi" that can explain her choices as a consequence of privacy concerns. The strongest version of this result is that *all possible behaviors on the part of Alice are compatible with all possible preferences that Alice may have over objects*: postulate some arbitrary behavior for Alice, and some arbitrary preference over objects, and the two will always be compatible.

This implies that BB's objective is hopeless. He can never learn anything about Alice's true preferences over political blogs, or over any other objects of choice. If BB tries to estimate preferences from some given choices by Alice, he finds that all preferences could be used to explain her choices. He cannot narrow down the set of preferences Alice might have, no matter what the observed behavior. The result continues to hold if BB adversarially sets up scenarios for Alice to choose from. That is, even if BB offers Alice menus of choices so as to maximize what he can learn from her behavior, the result is still that nothing can be learned.

The results in this paper have a variety of implications.

First, they motivate the use of specific parametric models of preferences over privacy. Our main result makes strong qualitative assumptions about preferences (separability, monotonicity). Given that such assumptions lack empirical bite, one should arguably turn to even stronger assumptions. The paper proposes an additive utility function that depends on the chosen object and on what is revealed by the consumer's choices. If Alice chooses x then she obtains a utility $u(x)$ and a penalty $v(x, y)$ for not choosing y, for all non-chosen y, as she reveals to BB that she ranks x over y. This additive model does have restrictions for the consumer's behavior, and could be estimated given data on Alice's choices. The model is methodologically close to models used in economics to explain individual choices, and could be econometrically estimated using standard techniques. The paper discusses a test for the additive model based on a linear program.

Second, while the paper's main motivation is consumers' behavior on the Internet, the results have implications for issues commonly discussed in behavioral economics. Certain behavioral anomalies could be the consequence of the presence of an outside observer. For example (elaborating on a laboratory experiment by Simonson and Tversky (1992)), consider a consumer who is going to buy a technical gadget, such as a phone or a camera. The consumer might prefer a simple camera over a more complex one which they do not know how to operate. However, when presented with a menu that has a simple, an intermediate, and an advanced camera, they might choose the intermediate one because they do not want to reveal to the world that they do not know how to use a complex camera. Of course, our results show that this line of reasoning may not be very useful, as anything can be explained in this fashion. The results do suggest, however, that a stronger parametric model may be useful to explain various behavioral phenomena.

Third, the results explain why BB may want to be hide the fact that consumer behavior is being observed. The NSA or Google seem to dislike openly discussing that they are monitoring consumers' online behavior. One could explain such a desire to hide by political issues, or because the observers wish to maintain a certain public image, but here we point to another reason. The observations simply become ineffective when the consumer is aware that she is being observed.

1.1 Related literature

The growing attention to privacy concerns has led to a growing literature studying privacy; see Heffetz and Ligett (2013) for a survey. Within this literature, an important question is how to model the preferences or utilities of privacy-aware agents in a way that describes their behavior in strategic settings.

One approach toward this goal, exemplified by Ghosh and Roth (2011), Nissim et al. (2012), Chen et al. (2013), Xiao (2013), and Nissim et al. (2014), is to use differential privacy in mechanism design as a way to quantify the privacy loss of an agent from participating the mechanism. Within this literature, each of Ghosh and Roth (2011), Nissim et al. (2012), and Xiao (2013) assume that the utility of a privacy-aware agent is her gain from the outcome of the interaction minus her loss from privacy leakage. This is a stronger condition than separability, as defined in Section 3.1, and a weaker condition than additivity, as defined in Section 3.2. In contrast, Chen et al. (2013) and Nissim et al. (2014) make the same separability assumption as used in this paper, but Chen et al. (2013) allows for non-monotone privacy preferences and Nissim et al. (2014)

uses a relaxed version of monotonicity.

Perhaps the model closest to ours is Gradwohl (2013), which also considers privacy-aware agents with preferences over outcome-privacy pairs. However, the technical quantification of privacy is different in the two models, as Gradwohl (2013) considers multiple agents engaging in a single interaction instead of multiple choices by a single agent as in our paper. Another related model is that of Gradwohl and Smorodinsky (2014), where a single privacy-aware agent must make decisions and knows that her choices are being observed. The model differs from ours in that the agent has a (cardinal) utility function over outcome-privacy pairs, rather than ordinal preferences as in this paper. Further, Gradwohl and Smorodinsky (2014) works in a Bayesian setting — the observer maintains a distribution over types that the agent may have, and performs a Bayesian update after each observed choice — whereas our results are in a prior-free setting. In addition, the nature of the results in both papers differ from ours. Gradwohl (2013) studies implementation from a mechanism design perspective, and Gradwohl and Smorodinsky (2014) studies the existence and uniqueness of equilibria between the agent and the observer. To contrast, we study the testable implications of privacy-aware preferences through the lens of revealed preference analysis.

Our paper is also related to the literature on psychological games. This line of work was initiated by Geanakoplos et al. (1989), and studies settings where players have preferences over their opponents' beliefs about their *actions*. This makes for an interesting interaction between the endogenous actions taken by players in a game, and each player's beliefs over those actions. In our paper, in contrast, there is a single agent making decisions (hence, we are not in a game theoretic setup), and this agent cares about what an outside observer infers about her *preferences*. The outside observer does not make decisions in our paper.

More broadly, there is a literature within economics that studies privacy in other game theoretic models. For example, Daughety and Reinganum (2010) study a signaling model where an agent's action may convey information about her type. In their setting, the agent may have privacy concerns, and therefore cares about what others learn about her type. Conitzer et al. (2012) study a setting in which consumers make repeat purchases from the same seller, and show that price discrimination based on past purchase decisions can harm consumer welfare. In a similar setting, Taylor (2004) shows that when consumers are privacy-oblivious, they suffer losses in welfare, but privacy-aware consumers may even cause firms to lower prices in equilibrium. Other studies on the relation between privacy and economics are surveyed by Acquisti et al. (2014).

Another related stream of work consists of empirical studies of people making privacy-aware choices in practice. For example, Goldfarb and Tucker (2012) examined the changing willingness of people to fill in personal data on income for market research surveys. The authors concluded that the recent decrease in willingness can be explained in part by an increased preference for privacy, and suggest that this desire for privacy may result from the now ubiquitous practice of data-driven price discrimination. In addition to pecuniary losses from price discrimination, simple embarrassment may also cause agents to become privacy-aware. Indeed, Goldfarb et al. (2013) demonstrates empirically that people behave differently when their choices are recorded by a human being, which the authors attribute to embarrassment. These studies suggest that people do indeed have privacy-aware preferences, and thus privacy concerns should be considered when analyzing empirical data.

2 Modeling privacy preferences

The goal of this paper is to study the testable implications of choice data in a context where agents have privacy preferences. To this end, we adapt the standard conceptualization of consumer choice theory in economics (see e.g. the textbook treatments in Mas-Colell et al. (1995) or Rubinstein (2012)) to a situation where the consumer is aware of, and has preferences over, the information revealed by her choices.

2.1 The setting

We focus on a situation where there is an outside observer (he), such as Google or the NSA, that is gathering data about the choices of a consumer (she) by observing her choices. We assume that the consumer is presented with a set of alternatives A and then makes a choice $c(A)$, which the outside observer sees. The observer then infers from this choice that $c(A)$ is preferred to all other alternatives in A.

The above parallels the classical revealed preference theory framework; however our model differs when it comes to the the behavior of the consumer, which we model as *privacy-aware*. We assume that the consumer is aware of the existence of an outside observer, so she may care about what her choices reveal about her. Specifically, her choices are motivated by two considerations. On the one hand, she cares about the actual chosen alternative. On the other hand, she cares about what her choices reveal about her preferences over alternatives, i.e., her revealed preferences. We capture this by assuming that the consumer has preferences over pairs (x, B), where x is the chosen object and B is the information revealed about the consumer's preferences.

An important point about the setting is that the inferences made by the observer do not recognize that the consumer is privacy aware. This assumption about the observer being naive is literally imposed on the behavior of the observer, but *it is really an assumption about how the agent thinks that the observer makes inferences.* The agent thinks that the observer naively uses revealed preference theory to make inferences about her preferences. The observer, however, could be as sophisticated as any reader of this paper in how they learn about the agent's preferences. The upshot of our results is that even such a sophisticated observer could not learn anything about the agent's behavior.

It is natural to go one step further and ask "What if the agent knows that the observer knows that the agent is privacy-aware?" Or, "what if the agent knows that the observer knows that the agent knows that the observer knows that the agent is privacy-aware?" This problem naturally lends itself to a discussion of the role of higher order beliefs. We formalize exactly this form of a cognitive hierarchy in Section 4, and we discuss how our results generalize.

2.2 Preliminaries

Before introducing our model formally, there are a few preliminaries that are important to discuss. Let $\mathbf{B}(X) = 2^{X \times X}$ denote the set of all binary preference relations on a set X and recall that a binary relation $\succeq$ is a weak order if it is complete (total) and transitive. We say that $x \succ y$ when $x \succeq y$ and it is not the case that $y \succeq x$. Finally, a linear order is a weak order such that if $x \neq y$ then $x \succ y$ or $y \succ x$.

We shall often interpret binary relations as graphs. For $B \in \mathbf{B}(X)$, define a graph by letting the vertex set of the graph be equal to X and the edge set be B. Thus for each element $(x, y) \in B$, we have a directed edge in the graph from x to y. We say that a binary relation B is acyclic if there does not exist a directed path that both originates and ends at x, for any $x \in X$. The following simple result, often called Szpilrajn's Lemma, is useful.

Lemma 1. *If $B \subseteq \mathbf{B}(X)$ is acyclic, then there is a linear order $\succeq$ such that $B \subseteq \succeq$.*

2.3 The model

Given the setting described above, our goal is to characterize the testable implications of choice data, and to understand how the testable implications change when consumers are privacy-aware as opposed to privacy-oblivious. To formalize this, we denote a *choice problem* by a tuple $(X, \mathcal{A}, c)$ where X is a finite set of alternatives, $\mathcal{A}$ is a collection of nonempty subsets of X, and $c : \mathcal{A} \to X$ such that $c(A) \in A$ for all $A \in \mathcal{A}$.

In choice problem $(X, \mathcal{A}, c)$, the consumer makes choices for each $A \in \mathcal{A}$ according to the function c. Further, given $A \in \mathcal{A}$ and $x = c(A)$, the observer infers that the consumer prefers x to any other alternative available in A. That is, he infers that the binary comparisons $(x, y)\ \forall\ y \in A \setminus \{x\}$ are part of the consumer's preferences over X. Such inferences lie at the heart of revealed preference theory (see e.g. Varian (1982) or Varian (2006)).

A *privacy preference* is a linear order $\succeq$ over $X \times 2^{X \times X}$. A privacy preference ranks objects of the form (x, B), where $x \in X$ and $B \in \mathbf{B}(X)$. If a consumer's choices are guided by a privacy preference, then she cares about two things: she cares about the choice made (i.e. x) and about what her choices reveal about her preference (i.e. B).

Our approach differs from the standard model in that the consumer has preferences not only over objects, but also over the choice data. Other papers have broken from the standard model to allow for preferences over menus (see e.g. Dekel et al. (2001, 2009)) or over beliefs (see Geanakoplos et al. (1989)).

Given the notions of a choice problem and privacy preferences defined above, we can now formally define the notion of rationalizability that we consider in this paper.

Definition 2. *A choice problem* $(X, \mathcal{A}, c)$ *is* ***rationalizable (via privacy preferences)*** *if there is a privacy preference* $\succeq$ *such that if* $x = c(A)$ *and* $y \in A \setminus \{x\}$ *then*

$$(x, \{(x, z) : z \in A \setminus \{x\}\}) \succ (y, \{(y, z) : z \in A \setminus \{y\}\}),$$

for all $A \in \mathcal{A}$*. In this case, we say that* $\succeq$ ***rationalizes*** $(X, \mathcal{A}, c)$*.*

This definition requires that for every observation of an element x chosen from a set A, and for every alternative $y \in A$ that was available but not chosen, the consumer prefers x *paired with the inferences* made from her choice of x, to the alternative y paired with the *counterfactual* inferences that would have been made if she had chosen y instead. We shall sometimes use the notation $A_x = \{(x, z) : z \in A \setminus \{x\}\}$ to denote the set of binary comparisons inferred by the observer from the consumer's choice of x from set A.

Thus, a choice problem is rationalizable when there exists a privacy preference that "explains" the data, i.e., when there exists a privacy preference for which the observed choices are maximal.

3 The rationalizability of privacy-aware choice

In this section, we present our main results, which characterize when choice data from privacy-aware consumers is rationalizable. Our results focus on the testable implications of structural assumptions about the form of the privacy preferences of the consumer. While a consumer's preferences may, in general, be a complex combination of preferences over the choices and revealed preferences, there are some natural properties that one may expect to hold in many situations. In particular, we focus on three increasingly strong structural assumptions in the following three subsections: monotonicity, separability, and additivity.

3.1 Monotone and separable privacy preferences

A natural assumption on privacy preferences is "monotonicity," i.e., the idea that revealing less information is always better. Monotonicity of privacy preferences is a common assumption in the privacy literature, e.g., see Xiao (2013) and Nissim et al. (2012), but of course one can imagine situations where it may not hold, e.g., see Chen et al. (2013) and Nissim et al. (2014).

In our context, we formalize monotone privacy preferences as follows.

Definition 3. *A binary relation* $\succeq$ *over* $X \times 2^{X \times X}$ *is a* ***monotone privacy preference*** *when*

1. $\succeq$ *is a linear order, and*

2. *$B \subsetneq B'$ implies that $(x, B) \succ (x, B')$.*

This definition formalizes the idea that revealing less information is better. In particular, if $B \subsetneq B'$, then fewer comparisons are being made in B than in B', so (x, B) reveals less information to the observer than (x, B').

Given the above definition, the question we address is "what are the empirical implications of monotone privacy preferences?" That is, "Is monotonicity refutable via choice data?" The following proposition highlights that monotonicity is *not* refutable, so any observed choice data has a monotone privacy preference that explains it. Note that Proposition 4 follows from a more general result, Theorem 14, which is presented in Section 4.

Proposition 4. *Any choice problem is rationalizable via monotone privacy preferences.*

Proposition 4 provides a contrast to the context of classical revealed preference theory, when consumers are privacy-oblivious. In particular, in the classical setting, choice behavior that violates the strong axiom of revealed preferences (SARP) is not rationalizable, and thus refutes the consumer choice model. However, when privacy-aware consumers are considered, such a refutation of monotonic preferences is impossible. Interestingly, this means that while one may believe that preferences are non-monotonic, the form of data considered in this paper does not have the power to refute monotonicity.[2]

Note that the question addressed by Proposition 4 is only whether the consumer's choice behavior is consistent with rational behavior, and is not about whether the consumer's underlying preferences over outcomes in X can be learned. In particular, these underlying preferences may not even be well defined for the general model considered to this point.

Since the observer is trying to learn the agent's preferences *over objects*, it is natural to postulate that the consumer has some underlying, or intrinsic, preferences over possible options when her choices are not being observed. Such preferences should be well defined: if outcome x is preferred to outcome y when both are paired with the same privacy set B, then it is natural that x will always be preferred to y when both are paired with the same privacy set B', for any other B'. This property induces underlying preferences over items in X, as well as the agent's privacy-aware preferences.

We formalize the notion of separable privacy preferences as follows.

Definition 5. *A binary relation $\succeq$ over $X \times 2^{X \times X}$ is a **separable privacy preference** if it is a monotone privacy preference and additionally satisfies that for all $x, y \in X$ and $B \in \mathbf{B}(X)$,*

$$(x, B) \succeq (y, B) \Longrightarrow (x, B') \succeq (y, B') \; \forall B' \in \mathbf{B}(X)$$

That is, whenever (x, B) is preferred to (y, B) for some privacy set B, then also (x, B') is preferred to (y, B') for all other sets B'.[3]

Separable privacy preferences have an associated preference relation over X. If $\succeq$ is a separable privacy preference, then define $\succeq|_X$ as $x \succeq|_X y$ if and only if $(x, B) \succeq (y, B)$ for all $B \in \mathbf{B}(X)$. Note that $\succeq|_X$ is a linear order over X; we can interpret $\succeq|_X$ as the projection of $\succeq$ onto X.

There are two questions we seek to answer: "What are the empirical implications of separability?" and "When can an observer learn the underlying choice preferences of the consumer?" The following proposition addresses both of these questions. Proposition 6 also follows from the more general result of Theorem 14, presented in Section 4.

Proposition 6. *Let $(X, \mathcal{A}, c)$ be a choice problem and let $\succeq$ be any linear order over X. Then there is a separable privacy preference $\succeq^*$ that rationalizes $(X, \mathcal{A}, c)$ such that the projection of $\succeq^*$ onto X is well defined and coincides with $\succeq$, i.e., $\succeq^*|_X = \succeq$.*

Think of $\succeq$ as a conjecture that the observer has about the agent. Proposition 6 implies that *no matter the nature of such a conjecture, and no matter what choice behavior is observed, the two are compatible.*

This proposition carries considerably more weight than Proposition 4 since separability imposes a great deal more structure than monotonicity alone. Further, Proposition 6 says much more than just that separability has no testable implications, or that it is not refutable via choice data. Proposition 6 highlights that the task of the observer is hopeless in this case – regardless of the choice data, the observer cannot narrow his hypothesis about the consumer preferences at all.

In some sense, our result is consistent with the idea that secrecy is crucial for observers such as the NSA and Google. If the consumer is not aware of the fact that she is being observed then the observer can learn a considerable amount from choice data, while if the consumer is aware that she is being observed then the choice data has little power (unless more structure is assumed than separability).

One way out of the negative conclusions from our result is to impose additional structure on the consumer's preferences. For example, one could require that the consumer cares more about correct inferences than about incorrect ones. We look next at a specific family of privacy-aware utility functions with an additive structure, that do have empirical content. These functions also lend themselves nicely to imposing additional assumptions on the form of preferences (such as penalizing correct inferences more than incorrect ones).

3.2 Additive privacy preferences

So far, we have seen that monotonicity and separability do not provide enough structure to allow choice data to have testable implications or to allow the observer to learn *anything* about consumer preferences over choices. This implies that further structure must be imposed for choice data to have empirical power. To that end, we now give an example of a model for privacy preferences where choice data does have testable implications. The model we consider builds on the notion of separable privacy preferences and additionally imposes additivity.

Definition 7. *A binary relation $\succeq$ over $X \times 2^{X \times X}$ is an **additive privacy preference** if there are functions $u : X \to \mathbb{R}^+$ and $v : X \times X \to \mathbb{R}^+$ such that $(x, B) \succ (x', B')$ iff*

$$u(x) - \sum_{(z,z') \in B} v(z, z') > u(x') - \sum_{(z,z') \in B'} v(z, z').$$

While monotonicity and separability are general structural properties of privacy preferences, the definition of additivity is much more concrete. It specifies a particular functional form, albeit a simple and natural one. In this definition, the consumer experiences utility $u(x)$ from the choice made and disutility $v(x, y)$ from the privacy loss of revealing that $x \succ y$ for every pair $(x, y) \in X \times X$. Note that this form is an additive extension of the classical consumer choice model, which would include only u and not v.

Moreover, this definition also satisfies both monotonicity and separability, making it a strictly stronger restriction. Monotonicity is satisfied because the agent always experiences a *loss* from each preference inferred by the observer. Namely, the range of v is restricted to non-negative reals, so for a fixed choice element, the agent will always prefer fewer inferences to be made about her preferences.[4] Separability is satisfied because utilities u determine the linear ordering over X, so for a fixed set of inferences made by the observer, privacy preferences will correspond to the preferences determined by u.

Of course there are a number of variations of this form that could also make sense, e.g., if the disutility from a revealed preference (x, y) was only counted once instead of (possibly) multiple times due to multiple revelations in the choice data. This would correspond to a consumer minimizing a global privacy loss rather than optimizing online for each menu. However, this modeling choice requires the agent to know ex ante the set $\mathcal{A}$ of menus from which she will choose, and additional assumptions about the order in which she faces these menus. For our analysis we restrict to additive preferences as defined above.

Rationalizability of additive privacy preferences corresponds to the existence of functions u and v, such that the observed choice behavior maximizes the consumer's utility under these functions. Here, it turns out the imposed structure on privacy preferences is enough to allow the model to have testable implications, as shown in the following proposition.

Proposition 8. *There exists a choice problem $(X, \mathcal{A}, c)$ that is not rationalizable with additive privacy preferences.*

Proposition 8 highlights that, while monotonicity and separability cannot be refuted with choice data, additivity can be refuted. To show this, we construct a simple example of choice data that cannot be explained with any functions u and v.

Proof of Proposition 8. To construct an example that is not rationalizable via additive privacy preferences, we begin by defining the set of alternatives as $X = \{x, y, z, w\}$ and the choice data to include six observations as follows:

$z = c(\{x,z\})$, $x = c(\{x,y,z\})$, $w = c(\{w,z\})$, $z = c(\{w,y,z\})$, $x = c(\{x,w\})$, $w = c(\{x,y,w\})$.

To see that this choice data is not rationalizable suppose towards a contradiction that the pair (u,v) rationalizes $(X,\mathcal{A},c)$. Then $z = c(\{x,z\})$ implies that

$$u(z) - v(z,x) > u(x) - v(x,z),$$

while $x = c(\{x,y,z\})$ implies that

$$u(z) - v(z,x) - v(z,y) < u(x) - v(x,z) - v(x,y).$$

Therefore $v(z,y) > v(x,y)$.

Similarly, we can argue that $w = c(\{w,z\})$ and $z = c(\{w,y,z\})$ together imply that $v(w,y) > v(z,y)$, and $x = c(\{x,w\})$ and $w = c(\{x,y,w\})$ together imply that $v(x,y) > v(w,y)$. This gives us a contradiction, so the choice data is not rationalizable. ☐

Given that the structure imposed by additive privacy preferences is testable, the next task is to characterize data sets that are consistent with (or refute) the additive privacy preference model. The example given in the proof of Proposition 8 already suggests an important feature of choice data that must hold for it to be rationalizable.

Given a choice problem $(X,\mathcal{A},c)$ and an element $y \in X$, define the binary relation R^y by $x \; R^y \; z$ if there is $A \in \mathcal{A}$ with $z = c(A)$ and $x = c(A \cup \{y\})$. Our next result gives a test for additively rational preferences. It says that if there exist cycles in the binary relation R^y, then the choice data cannot be rationalized by additive privacy preferences.

Proposition 9. *A choice problem $(X,\mathcal{A},c)$ can be rationalized by additive privacy preferences only if R^y is acyclic for all $y \in X$.*

Proof. Let c be rationalizable by the additive privacy preferences characterized by (u,v). For each $x,z \in X$ such that $x \; R^y \; z$, there exists some $A \in \mathcal{A}$ such that $z = c(A)$ and $x \in A$, so

$$u(z) - \sum_{t\in A} v(z,t) > u(x) - \sum_{t \in A} v(x,t).$$

Similarly, $x = c(A \cup \{y\})$ and $z \in A \cup \{y\}$, so

$$u(z) - \sum_{t\in A} v(z,t) - v(z,y) > u(x) - \sum_{t \in A} v(x,t) - v(x,y).$$

For both inequalities to be true simultaneously, we need $v(z,y) > v(x,y)$. Thus,

$$x \; R^y \; z \Longrightarrow v(z,y) > v(x,y). \tag{1}$$

Now assume there exists a cycle in binary relation R^y: $a_1 \; R^y \; a_2 \; R^y \cdots R^y \; a_k \; R^y \; a_1$. Then by Equation (1), it must be that $v(a_1,y) > v(a_2,y) > \cdots > v(a_k,y) > v(a_1,y)$. In particular, $v(a_1,y) > v(a_1,y)$ which is a contradiction. Thus for choices to be rationalized, acyclicity of R^y for all $y \in X$ is a necessary condition. ☐

Of course, one would like to develop a test for rationalizability that is both necessary and sufficient. We do this next. Unfortunately, the test we develop takes super-exponential time to even write down. This suggests that acyclicity of R^y, despite being only a necessary condition, is likely a more practical condition to use when testing for rationalizability.

To describe the test for rationalizability, first observe that when an object x is chosen from a set, the observer infers that x (paired with A_x) is preferred to y (paired with A_y), for all $y \in A \setminus \{x\}$. Since we have assumed these preferences take a specific functional form as in Definition 7, the observer can also infer the corresponding inequality in terms of functions u and v. We initialize a large matrix to record the inequalities that are inferred from choice behavior, and ask if there exist values of $u(x)$ and $v(x, x')$ for all $x, x' \in X$ for which all inferred inequalities hold. If so, these values of u and v form additive privacy preferences that rationalize choices. If not, then no such preferences exist and the observed choice behavior is not rationalizable.

Remark 10. A choice problem $(X, \mathcal{A}, c)$ is rationalizable if and only if there exists functions $u : X \to \mathbb{R}^+$ and $v : X \times X \to \mathbb{R}^+$ satisfying the matrix inequality given by Equation (4), below.

To explicitly state the matrix inequality, let us index the elements of $X = \{x_1, \ldots, x_n\}$. Then for each $A \in \mathcal{A}$, the agent chooses some $x_i = c(A) \in A$. By the definition of additive preferences, every $x_j \in A$ for $j \neq i$ was *not* chosen because

$$u(x_i) - \sum_{z \in A \setminus \{x_i\}} v(x_i, z) > u(x_j) - \sum_{z \in A \setminus \{x_j\}} v(x_j, z)$$

Rearranging terms gives the following inequality:

$$u(x_i) - u(x_j) + \sum_{z \in A \setminus \{x_j\}} v(x_j, z) - \sum_{z \in A \setminus \{x_i\}} v(x_i, z) > 0 \tag{2}$$

To record all inequalities implied by observed choices, we instantiate a matrix T with n^2 columns, where the first n columns correspond to elements $x_1, \ldots, x_n \in X$, and the remaining $n^2 - n$ columns correspond to ordered pairs (x_i, x_j) of elements in X, for $i \neq j$.[5] T will have a row for each triple (A, x_i, x_j), where $A \in \mathcal{A}$, and $x_i, x_j \in A$. If the agent is observed to choose $x_i = c(A)$, then Equation (2) must be true for each $x_j \in A$ for $j \neq i$. To encode this inequality for each such x_j, we fill in the row corresponding to (A, x_i, x_j) as follows: enter $+1$ in the i^{th} column, -1 in the j^{th} column, $+1$ in columns corresponding to pairs (x_j, z) where $z \in A$, -1 in columns corresponding to pairs (x_i, z) where $z \in A$, and zeros elsewhere.

To complete the encoding, we also instantiate a vector $\vec{u}$, which represents the values of $u(\cdot)$ and $v(\cdot, \cdot)$ evaluated on all elements of X. The first n entries of $\vec{u}$ will contain variables for $u(x_1), \ldots, u(x_n)$, and the remaining $n^2 - n$ entries will contain variables for $v(x_i, x_j)$ for $i \neq j$, in the same order in which the pairs appear in the columns of T.

Each row of the matrix product $T\vec{u}$ would equal

$$u(x_i) - u(x_j) + \sum_{z \in A \setminus \{x_j\}} v(x_j, z) - \sum_{z \in A \setminus \{x_i\}} v(x_i, z) \tag{3}$$

for some set $A \in \mathcal{A}$, observed choice $x_i = c(A)$, and not-chosen element $x_j \in A$. Note that Equations (2) and (3) are identical, so the observed choices are rationalizable if and only if there exists an assignment of the variables in $\vec{u}$ such that each row of $T\vec{u}$ is greater than zero. That is,

$$T\vec{u} > \vec{0} \tag{4}$$

Any such $\vec{u}$ would specify functions $u : X \to \mathbb{R}^+$ and $v : X \times X \to \mathbb{R}^+$ which correspond to additive privacy preferences that are optimized by the observed choices.

4 Higher order privacy preferences

The results we have discussed so far are predicated on the notion that the agent thinks that the observer is naive. We shall now relax the assumption of naiveté. We are going to allow the agent to believe that the observer thinks that she is privacy-aware.

Going back to Alice choosing among political blogs, suppose that she reasons as follows. Alice may realize that her observed choices violate the strong axiom of revealed preference and therefore cannot correspond to the choices of a rational agent. This could tip off the observer to the fact that she is privacy-aware, as we have seen that privacy-awareness is a plausible explanation for violations of the revealed preference axioms. Alice could now be concerned about the observer's inference about her preferences over objects *and* over revealed preferences. Perhaps she thinks that the observer will infer that she is avoiding blog a because of what it reveals about her, and that fact itself is something she does not wish be known. After all, if Alice has a preference for privacy, perhaps she has something to hide.

More generally, an agent may be concerned not only about what her behavior reveals about her preferences over X, but also about what her behavior reveals of her preferences for privacy. She may then make choices to minimize inferences the observer is able to make about her preferences for privacy, as well as her preferences over X.

To provide a model that incorporates such issues, we define a hierarchy of higher order preferences, called *level-k preferences*, where a level-k consumer is aware that the observer may make inferences about her level-$(k-1)$ privacy preferences, and has preferences over the information the observer can infer. In our construction, level-0 corresponds to the classical privacy-oblivious setting, and the setting we have considered to this point is that of a level-1 consumer (Sections 2 and 3).

The meaning of such levels should be clear. If Alice is concerned about facing an observer who makes level k inferences, then her behavior will be dictated by the level $k+1$ model. To emphasize a point we have made repeatedly, *the real observer may be as sophisticated as one wants,* but Alice thinks that the observer thinks that Alice thinks that the observer thinks that Alice thinks ... that the observer makes inferences based on revealed preferences.

4.1 Level-k privacy preferences

To formally define a "cognitive hierarchy" for privacy-aware consumers we use the following sequence of sets, $\mathcal{Y}^k$ for $k \geq 0$. $\mathcal{Y}^0 = X$, $\mathcal{Y}^1 = X \times \mathbf{B}(\mathcal{Y}^0)$, and let $\mathcal{Y}^k = X \times \mathbf{B}(\mathcal{Y}^{k-1})$. A level-$k$ privacy preference can then be defined as a binary relation $\succeq^k$ over $\mathcal{Y}^k = X \times \mathbf{B}(\mathcal{Y}^{k-1})$. That is, $\succeq^k$ describes preferences over pairs of objects $x \in X$ and the set of level-$(k-1)$ preferences that are revealed from the choice of x.

Given the results in Section 3, our focus is on monotone, separable privacy preferences. As such, we can extend the notion of monotonicity discussed in Section 3 to level-k privacy preferences as follows.

Definition 11. *A* ***monotone level-k privacy preference*** *is a binary relation $\succeq^k$ over $\mathcal{Y}^k = X \times \mathbf{B}(\mathcal{Y}^{k-1})$ such that*

1. *$\succeq^k$ is a linear order, and*
2. *$B \subsetneq B'$ implies that $(x, B) \succ (x, B')$, for all $B, B' \in \mathbf{B}(\mathcal{Y}^{k-1})$.*

For this definition to hold for level-0, we define $\mathcal{Y}^{-1}$ to be the empty set.

Similarly, we extend the notion of separability to level-k privacy preferences as follows.

Definition 12. *A* ***separable level-k privacy preference*** *is a binary relation $\succeq^k$ over $\mathcal{Y}^k = X \times \mathbf{B}(\mathcal{Y}^{k-1})$ that is monotone and additionally satisfies for any $B \in \mathbf{B}(\mathcal{Y}^{k-1})$,*

$$(x, B) \succeq^k (y, B) \Longrightarrow (x, B') \succeq^k (y, B') \; \forall B' \in \mathbf{B}(\mathcal{Y}^{k-1})$$

Given the notion of level-k privacy preferences, we need to characterize how an observer will make inferences from observed choices. Naturally, the exact information inferred will depend on the level which the observer believes the privacy preferences to be. For example, if the observer believes the consumer to have level-0 preferences, the information inferred by the observer is the set

$$A_x = \{(x, y) : y \in A \setminus \{x\}\},$$

which is a binary relation over X. So $A_x \in \mathbf{B}(\mathcal{Y}^0)$. However, if the observer believes the consumer to have level-1 preferences, the information inferred by the observer is the set

$$\{((x, A_x), (y, A_y)) : y \in A \setminus \{x\}\} \in \mathbf{B}(\mathcal{Y}^1).$$

More generally, to describe the observer's inferences under the the belief that the consumer is level-k, we introduce the following notation. Consider the functions $T^k : \mathcal{A} \times X \to \mathbf{B}(\mathcal{Y}^k)$, for $k \geq 0$. Let

$$\begin{aligned}
T^0(A, x) &= \{(x, y) : y \in A \setminus \{x\}\} \in \mathbf{B}(\mathcal{Y}^0) \\
T^1(A, x) &= \{\big((x, T^0(A, x)), (y, T^0(A, y))\big) : y \in A \setminus \{x\}\} \in \mathbf{B}(\mathcal{Y}^1) \\
\vdots \quad &\quad \vdots \\
T^k(A, x) &= \{\big((x, T^{k-1}(A, x)), (y, T^{k-1}(A, y))\big) : y \in A \setminus \{x\}\} \in \mathbf{B}(\mathcal{Y}^k).
\end{aligned}$$

In words, $T^k(A, x)$ are the level-k preferences (over alternatives in A and set of level-$(k-1)$ preferences that will be inferred from each choice) that would cause the agent to choose x from the set A. Then generally, a level-k agent making choice $x = c(A)$ must have $T^k(A, x)$ as a subset of her level-k preferences.

Example: Level-2 privacy preferences.

In order to illustrate the cognitive hierarchy more concretely it is useful to describe the case of level-2 privacy preferences in detail. Recall that the level-0 privacy preferences are the classical setting of privacy-oblivious consumers and level-1 privacy preferences are the case we study in Sections 2 and 3. As we shall see, there is a sense in which level-2 is all that is needed.

Continuing with the story about Alice, we remarked how she could come to question her level-1 behavior because she should realize that there is something suspicious about her choices violating the revealed preference axioms. As the result of such a realization, she might entertain level-2 behavior. She might think that the observer thinks that she is level-1. Now, there is no reason for her to go any further because, in contrast with level-1, *nothing could give her away.*

While her violations of the revealed preference axioms indicate that she cannot be level-0, given our Proposition 4, nothing about her behavior could contradict that she is level-1. She has no reason to think that reasoning beyond level-2 will afford her more privacy — we have already seen that nothing in her behavior that could prove to the observer that she is not level-1.

More concretely, suppose that x is chosen from set A. The observer, who thinks the consumer is at level-1, infers the level-1 preferences

$$(x, A_x) \succ (z, A_z) \;\; \forall\, z \in A \setminus \{x\},$$

or, more specifically, that her level-1 privacy preferences correspond to the binary relation,

$$\bigcup \{[(x, A_x), (z, A_z)] : z \in A \setminus \{x\}\}. \tag{5}$$

Now, the agent who believes that the observer will make such an inference will only choose x when this choice *together with inferences revealed by the choice* is better than the choice of another alternative in A with its accompanying inferences. That is, she will choose x over y in A whenever the choice of x together with the release of the information in Equation (5) is preferred to the choice of y together with the information,

$$\bigcup \{[(y, A_y), (z, A_z)] : z \in A \setminus \{y\}\}\,.$$

If a level 2 agent chooses x from menu A, she knows that observer will make inferences according to Equation (5). Then her choice of x must maximize her preferences over outcomes *and* these known inferences that will be made. Specifically, she will choose x if her level-2 preferences are, for all available $y \in A$,

$$(x, \cup\{[(x, A_x), (z, A_z)] : z \in A \setminus \{x\}\}) \succ (y, \cup\{[(y, A_y), (z, A_z)] : z \in A \setminus \{z\}\}) \tag{6}$$

Using the notation defined earlier in this section, we can re-write Equation (6) as a binary relation,

$$[(x, T^1(A, x)), (y, T^1(A, y))].$$

Since the same can be said for *every* available alternative $y \in A$ that was not chosen, the following must be a part of the agent's level-2 preferences

$$T^2(A, x) = \{\big[(x, T^1(A, x)), (y, T^1(A, y))\big] : y \in A \setminus \{x\}\}$$

Note, however, that the observer does not get to infer $T^2(A, x)$. He believes the agent to have level-1 preferences, and upon seeing $x = c(A)$, he infers $T^1(A, x)$. This is why the agent chooses $x \in A$ to optimize her preferences over X and sets of the form $T^1(A, \cdot)$.

4.2 The rationalizability of level-k preferences

Given the notion of a privacy-aware cognitive hierarchy formalized by level-k privacy preferences, we are now ready to move on to the task of understanding the empirical implications of higher order reasoning by privacy-aware consumers. To do this, we must first adapt the notion of rationalizability to level-k reasoning. For this, the natural generalization of Definition 2 to higher order reasoning is as follows. This definition reduces to Definition 2 when level-1 is considered, and to the classical definition of rationalizability in the privacy-oblivious case when level-0 is considered.

Definition 13. *A choice $(X, \mathcal{A}, c)$ is **level-k rationalizable** if there is a level-k privacy preference $\succeq^k \in \mathbf{B}(\mathcal{Y}^k)$ such that for all $A \in \mathcal{A}$, $T^k(A, c(A)) \subseteq \succeq^k$.*

Given this definition, we can now ask the same two questions we considered in Section 3 about level-k privacy preferences: "What are the empirical implications of level-k privacy preferences?" and "When can the observer learn the underlying choice preferences of consumers?" Our main result is the following theorem, which answer these questions.

Theorem 14. *Let $(X, \mathcal{A}, c)$ be a choice problem. Let $k > 0$ and $\succeq$ be any linear order over X. Then there is a monotone, separable level-k privacy preference $\succeq^*$ that level-k rationalizes $(X, \mathcal{A}, c)$ and such that:*

$$x \succeq y \text{ iff } (x, B) \succeq^* (y, B) \text{ for all } B \in \mathbf{B}(\mathcal{Y}^{k-1}).$$

This result implies that the conclusions of Proposition 6 are not just an anomaly due to restrictions on Alice's reasoning. Theorem 14 says that for *any* level-k at which Alice chooses to reason, Big Brother cannot test if she is behaving rationally, and cannot learn anything about her preferences over X.

Another interpretation is that Alice always has an "alibi," in terms of other preference orderings over objects that are also consistent with her choices. In the case where Big Brother deems one particular preference ordering to be the most desirable, Alice's choices can never reveal that her preferences differ from Big Brother's desired ordering, for any level of reasoning that she may use.

Proof of Theorem 14. Let $T^{k-1} : \mathcal{A} \times X \to \mathbf{B}(\mathcal{Y}^{k-1})$ be as defined in Section 4.1. For shorthand, write $\mathcal{Y}$ for $\mathcal{Y}^{k-1}$ and T for T^{k-1}. Then T describes the set of level-$(k-1)$ preferences inferred by the observer as a result of the agent's choice behavior. That is, when the agent chooses $x = c(A)$, the observer will infer all preferences in the set $T(A, x)$. Note that T is one-to-one and satisfies the following property: for all $A \in \mathcal{A}$ and all $x, x' \in A$,

$$|T(A, x)| = |T(A, x')|. \tag{7}$$

Property (7) follows because the number of pairs $\{((x, T(A, x)), (y, T(A, y)))$ $y \in A \setminus \{x\}\}$ is the same for any $x \in A$.

We now construct a binary relation E over $X \times \mathbf{B}(\mathcal{Y})$. It will be useful for the proof to think of E as the edges of a directed graph $G = (V, E)$, where $V = X \times \mathbf{B}(\mathcal{Y})$. We create edges in E according to the desiderata of our privacy-aware preferences: monotone, separable, and rationalizing choice behavior. Define E as follows: $(x, B)\ E\ (x', B')$ if either (1) $x = x'$ and $B \subsetneq B'$, (2) $B = B'$ and $x \succ x'$ according to linear order $\succeq$, or (3) $x \neq x'$ and there is $A \in \mathcal{A}$ with $x = c(A)$, $x' \in A$ and $B = T(A, x)$ while $B' = T(A, x')$. We will call these edge types respectively "monotone," "separable," and "rationalizing," as a reference to the property they are meant to impose. By Lemma 1, we are done if we show that E is acyclic.

Assume towards a contradiction that there is a cycle in this graph. Then there exists a sequence $j = 1, \ldots, K$ such that

$$(x^1, B^1)\ E\ (x^2, B^2)\ E \cdots E\ (x^K, B^K) \text{ and } (x^K, B^K)\ E\ (x^1, B^1).$$

For any monotone edge $(x^i, B^i)\ E\ (x^{i+1}, B^{i+1})$, it must be the case that $|B^i| < |B^{i+1}|$ since $B^i \subset B^{i+1}$. If this were a separable edge, then $B^i = B^{i+1}$, so $|B^i| = |B^{i+1}|$. Similarly, for any rationalizing edge, $|B^i| = |T(A, x)| = |T(A, x')| = |B^{i+1}|$. Thus as we traverse any path in this graph, the size of the second component is non-increasing along all edges, and strictly decreasing along monotone edges. This implies that there can be no cycles containing monotone edges, and our assumed cycle must consist entirely of rationalizing and separable edges.

If there are two sequential rationalizing edges in this cycle, then there exists a j and some $A \in \mathcal{A}$ such that

$$(x^j, T(A, x^j)) \, E \, (x^{j+1}, T(A, x^{j+1})) \, E \, (x^{j+2}, T(A, x^{j+2})),$$

where $x^j, x^{j+1}, x^{j+2} \in A$. From the first edge, $x^j = c(A)$ in some observation, and from the second edge, $x^{j+1} = c(A)$ in another observation. If $x^j \neq x^{j+1}$ then $c(A) = x^j \neq x^{j+1} = c(A)$ which contradicts the uniqueness of choice imposed by the linear ordering. If $x^j = x^{j+1}$, then $(x^j, T(A, x^j)) = (x^{j+1}, T(A, x^{j+1}))$, which implies that an element of $X \times \mathbf{B}(\mathcal{Y})$ is strictly preferred to itself, which is a contradiction. Thus no cycle can contain two sequential rationalizing edges.

If there are two sequential separable edges in our cycle, then there exists a j such that

$$(x^j, B^j) \, E \, (x^{j+1}, B^{j+1}) \, E \, (x^{j+2}, B^{j+2}),$$

where $B^j = B^{j+1} = B^{j+2}$ and $x^j \succ x^{j+1} \succ x^{j+2}$. By transitivity, $x^j \succ x^{j+2}$, so there must also be a separable edge in the graph $(x^j, B^j) \, E \, (x^{j+2}, B^{j+2})$. If the cycle we have selected contains two sequential separable edges, then there must exist another cycle that is identical to the original cycle, except with the two sequential separable edges replaced by a single separable edge. Thus we can assume without loss of generality that the cycle we have selected does not contain two sequential separable edges. Note that the only time this is with some loss of generality is when there is a cycle containing only separable edges. By assumption, $\succeq$ is a linear order over X and must be acyclic, so this is not possible.

Given the previous two observations, we can assume without loss of generality that the cycle $j = 1, \ldots, K$ contains alternating rationalizing and separable edges. This includes the endpoints $j = K$ and $j = 1$ since they too are connected by an edge.

If there is a path in the graph containing sequential rationalizing, separable, and rationalizing edges, then there exists $A, A' \in \mathcal{A}$ and a j such that

$$(x^j, T(A, x^j)) \, E \, (x^{j+1}, T(A, x^{j+1})) \, E \, (x^{j+2}, T(A', x^{j+2})) \, E \, (x^{j+3}, T(A', x^{j+3})),$$

where $x^j, x^{j+1} \in A$, $x^{j+2}, x^{j+3} \in A'$, and choices $x^j = c(A)$ and $x^{j+2} = c(A')$ are observed. Since the middle edge is separable, it must be that $T(A, x^{j+1}) = T(A', x^{j+2})$, so $x^{j+1} = x^{j+2}$ and $A = A'$. However, this means that $(x^{j+1}, T(A, x^{j+1}))$ is strictly preferred to itself, which is a contradiction, so no such path in the graph can exist.

Since edges must alternate between rationalizing and separable, this leaves the only possible cycles to be of length two, containing one rationalizing edge and one separable edge. However, if such a cycle existed, then traversing the cycle twice would yield a path containing sequential rationalizing, separable, and rationalizing edges, which has been shown to not exist in this graph.

Thus we can conclude that E must be acyclic, which completes the proof. $\square$

5 Concluding remarks

We conclude by describing what our results mean for Alice's story and for future research on privacy.

Alice makes choices knowing that she is being observed. She thinks that the observer uses revealed preference theory to infer her preferences. She might think that the observer is not sophisticated and uses classical revealed preferences naively to infer her preferences over objects. Alternatively, she might think that the observer is sophisticated and knows that she has preferences for privacy; in this case, the observer tries to infer (again using revealed preferences) Alice's preferences for objects and privacy.

The story of Alice, however, is more "The Matrix" than "in Wonderland." Alice believes that she is one step ahead of the observer, and makes choices taking into account what he learns about her from her choices. In reality, however, the observer is us: the readers and writers of this paper.

We are trying to understand Alice's behavior, and to infer what her preferences over objects might be. The main result of our work is that such a task is hopeless. Any behavior by Alice is consistent with any preferences over objects one might conjecture. This fact remains is true for any degree of sophistication that Alice may have in her model of how the observer makes inferences. Other observers on the internet, such as Google or the NSA, would have to reach the same conclusion.

One way out is to impose additional structure on Alice's preferences. Our main result imposes separability and monotonicity, which are strong assumptions in many other environments, but are not enough here. We have suggested additive privacy preferences (Section 3.2) as a potentially useful model to follow. Additive preferences do impose observable restrictions on choice, and its parameters could be learned or estimated from choice data. Privacy researchers looking to model a utility for privacy should consider the additive model as a promising candidate.

Another source of structure is the observer's possible objectives. Our model is one of *intrinsic* preferences for privacy; Alice has an innate, exogenous desire for privacy. One could instead imagine an *instrumental* preference for privacy, which arises from Alice's prediction of how the (adversarial) observer will use the information he collects about her preferences. By imposing assumptions on the observer's possible actions and objectives, one can restrict the universe of possible privacy-aware preferences. If one does not impose any assumptions on the observer's future actions, one is back in the framework of our paper, even if privacy concerns are instrumental.[6]

6 APPENDIX

An alternative notion of separability.

Here we consider a different notion of separability and show that our main results continue to hold. This alternative definition was suggested to us by an anonymous referee, and captures the idea that inferred preferences should be consistent with revealed preferences.

Definition 15. *A binary relation $\succeq$ over $X \times 2^{X \times X}$ is a* ***separable privacy preference*** *if it is a monotone privacy preference and additionally satisfies that for all $x, y \in X$ and for all nonempty $B, B' \in \mathbf{B}(X)$ with the property that $\{(x,y),(y,x)\} \cap B = \emptyset$, and $\{(x,y),(y,x)\} \cap B' = \emptyset$,*

$$(x, B \cup \{(x,y)\}) \succeq (y, B \cup \{(y,x)\}) \Longrightarrow (x, B' \cup \{(x,y)\}) \succeq (y, B' \cup \{(y,x)\}).$$

Proposition 16. *Let $(X, \mathcal{A}, c)$ be any choice problem. Then there is a monotone and separable privacy preference $\succeq^*$ that rationalizes $(X, \mathcal{A}, c)$.*

We note here that inspection of the proof of Proposition 16 reveals that we actually prove a stronger statement that is similar to Theorem 14.

Proof of Proposition 16. Let $\succeq$ be any given linear order on X. Let $T : \mathcal{A} \times X \to \mathbf{B}(X)$ equal T^1 as defined in Section 4.1. As established in the proof of Theorem 14, T is one-to-one and satisfies the property that $|T(A,x)| = |T(A,x')|$ for all $A \in \mathcal{A}$ and all $x, x' \in A$.

Let $V = X \times \mathbf{B}(X)$. Define a binary relation E over V as follows: $(x, B)\ E\ (x', B')$ if either (1) $x = x'$ and $B \subsetneq B'$; (2) there is a non-empty $C \in \mathbf{B}(X)$ such that $\{(x,y),(y,x)\} \cap C = \emptyset$, $B = C \cup \{(x,x')\}$ and $B' = C \cup \{(x',x)\}$, and $x \succ x'$ according to linear order $\succeq$; or (3) $x \neq x'$ and there is $A \in \mathcal{A}$ with $x = c(A)$, $x' \in A$ and $B = T(A,x)$ while $B' = T(A,x')$. We will call these edge types respectively "monotone," "separable," and "rationalizing." By Szpilrajn's Lemma, it will suffice to show that there are no cycles in the binary relation E.

Observe:

1. $|B| = |B'|$ if $(x,B)\ E\ (y,B')$ and E is either a separable edge or a rationalizing edge.

2. $|B| < |B'|$ if $(x,B)\ E\ (y,B')$ and E is a monotone edge.

3. $B \setminus \{(x,y)\} \subseteq B' \cup \{(y,x)\}$ if $(x,B)\ E\ (y,B')$ is a separable edge.

Assume towards a contradiction that there is a cycle in this graph. Then there exists a sequence $i = 1, \ldots, K$ such that

$$(x^1, B^1)\ E\ (x^2, B^2)\ E \cdots E\ (x^K, B^K) \text{ and } (x^K, B^K)\ E\ (x^1, B^1).$$

Observations 1 and 2 imply that there cannot be any monotone edges in the cycle because the size of B^i is non-increasing along all edges and strictly decreasing along monotone edges. By reasoning exactly as in the proof of Theorem 14, we conclude that there cannot be two sequential rationalizing edges.

Assume first that there is a rationalizing edge in the cycle, and suppose without loss of generality that it is $(x^1, B^1)\,E\,(x^2, B^2)$. Then the edge $(x^2, B^2)\,E\,(x^3, B^3)$ must be a separable edge. Since the first edge is a rationalizing edge, there must exist a menu $A \in \mathcal{A}$ such that $B^2 = T(A, x^2)$. Since the second edge is a separable edge, there must be a non-empty set $C \in \mathbf{B}(X)$ such that $B^2 = C \cup \{(x^2, x^3)\}$. Together, these imply that $T(A, x^2) = C \cup \{(x^2, x^3)\}$. Since C is required to be non-empty, there exists a $w \in A$ such that $w \neq x^3$ and $(x^2, w) \in C$.

Let i be the smallest integer such that $(x^i, B^i)E(x^{i+1}, B^{i+1})$ is a rationalizing edge. Then the sequence

$$(x^2, B^2)\, E\, (x^3, B^3)\, E \cdots E\, (\, x^{i-1}, B^{i-1})\, E\, (x^i, B^i)$$

consists entirely of separable edges. By a repeated application of Observation 3, we obtain that

$$B^2 \setminus \{(x^2, x^3)\} \subseteq B^i \cup \left(\bigcup_{\ell=3}^{i-1} \{(x^\ell, x^{\ell+1}\} \right).$$

Note that $x^2 \succ x^3 \succ \cdots x^i$, so $x^2 \neq x^\ell$ for any $\ell = 3, \ldots, i$ Then since $(x^2, w) \in B^2 \setminus \{(x^2, x^3)\}$, it must be that $(x^2, w) \in B^i$.

The edge $(x^i, B^i)\ E\ (x^{i+1}, B^{i+1})$ is a rationalizing edge, so there exists a menu $A' \in \mathcal{A}$ such that $B^i = T(A', x^i)$. Since $(x^2, w) \in B^i$, then it must be that $x^2 = x^i$, which is a contradiction since $x^2 \succ x^3 \succ \cdots \succ x^i$. Thus there cannot be a rationalizing edge followed by any number of separable edges, followed by another rationalizing edge.

This leaves that the cycle must consist either entirely of rationalizing edges or entirely of separable edges. It cannot consist of only rationalizing edges, as we have established previously that the cycle cannot contain two sequential rationalizing edges. The cycle cannot consist of only separable edges, since that would imply a cycle in the graph $(X, \succ)$. This is not possible because $\succeq$ is a linear order. Thus no cycles can exist in this graph. $\square$

Endnotes

1. Like Alice, 85% of adult internet users have take steps to avoid surveillance by other people or organizations, see Rainie et al. (2013) (and the papers by Goldfarb and Tucker (2012) and Goldfarb et al. (2013) for related evidence).

2. This phenomenon is common in the consumer choice formulation of the revealed preference problem, but it comes about for completely different reasons.

3. Our definition of separability is standard: see for example Varian (1984). We also consider an alternative notion of separability in the appendix.

4. Monotonicity restricts to the case where people want to keep their preferences private. It may be interesting to explore in future work, the case where people are happy to reveal their information, e.g., conspicuous consumption. Under additive preferences, this would correspond to allowing the range of v to be all of $\mathbb{R}$.

5. There are only $n^2 - n$ columns because we do not compare elements to themselves

6. It is worth mentioning that the literature on privacy in computer science is essentially motivated by intrinsic preference for privacy, and has not tried to model the observer as we suggest here. To contrast, much of the philosophy literature on privacy assumes an instrumental preference for privacy. See Moor (1990) for a comparison of various privacy definitions from a philosophical perspective.

References

Acquisti, Alessandro, Curtis Taylor, Liad Wagman. 2014. The economics of privacy. Forthcoming: Journal of Economic Literature.

Chen, Yiling, Stephen Chong, Ian A. Kash, Tal Moran, Salil Vadhan. 2013. Truthful mechanisms for agents that value privacy. *Proceedings of the 14th ACM Conference on Electronic Commerce*. EC '13, 215–232.

Conitzer, Vincent, Curtis R. Taylor, Liad Wagman. 2012. Hide and seek: Costly consumer privacy in a market with repeat purchases. *Marketing Science* **31**(2) 277–292.

Daughety, Andrew F, Jennifer F Reinganum. 2010. Public goods, social pressure, and the choice between privacy and publicity. *American Economic Journal: Microeconomics* **2**(2) 191–221.

Dekel, Eddie, Barton L Lipman, Aldo Rustichini. 2001. Representing preferences with a unique subjective state space. *Econometrica* **69**(4) 891–934.

Dekel, Eddie, Barton L. Lipman, Aldo Rustichini. 2009. Temptation-driven preferences. *The Review of Economic Studies* **76**(3) 937–971.

Geanakoplos, John, David Pearce, Ennio Stacchetti. 1989. Psychological games and sequential rationality. *Games and Economic Behavior* **1**(1) 60–79.

Ghosh, Arpita, Aaron Roth. 2011. Selling privacy at auction. *Proceedings of the 12th ACM Conference on Electronic Commerce*. EC '11, 199–208.

Goldfarb, Avi, Ryan C. Mcdevitt, Sampsa Samila, Brian Silverman. 2013. The effect of social interaction on economic transactions: An embarrassment of niches?

Goldfarb, Avi, Catherine Tucker. 2012. Shifts in privacy concerns. *American Economic Review* **102**(3) 349–353.

Gradwohl, Ronen. 2013. Privacy in implementation. *CMS-EMS Discussion Paper 1561*.

Gradwohl, Ronen, Rann Smorodinsky. 2014. Subjective perception games and privacy. *arXiv preprint 1409.1487* .

Heffetz, Ori, Katrina Ligett. 2013. Privacy and data-based research. Tech. rep., National Bureau of Economic Research.

Mas-Colell, Andreu, Michael D. Whinston, Jerry R Green. 1995. *Microeconomic theory*. Oxford University Press.

Moor, James H. 1990. The ethics of privacy protection. *Library Trends* **39**(1-2) 69–82.

Nissim, Kobbi, Claudio Orlandi, Rann Smorodinsky. 2012. Privacy-aware mechanism design. *Proceedings of the 13th ACM Conference on Electronic Commerce*. EC '12, 774–789.

Nissim, Kobbi, Salil Vadhan, David Xiao. 2014. Redrawing the boundaries on purchasing data from privacy-sensitive individuals. *Proceedings of the 5th Conference on Innovations in Theoretical Computer Science*. ITCS '14, 411–422.

Rainie, Lee, Sara Kiesler, Ruogu Kang, Mary Madden. 2013. Anonymity, privacy and security online. Tech. rep., Pew Research Center.

Rubinstein, Ariel. 2012. *Lecture notes in microeconomic theory: the economic agent*. Princeton University Press.

Simonson, Itamar, Amos Tversky. 1992. Choice in context: tradeoff contrast and extremeness aversion. *Journal of Marketing Research* **29**(3) 281–285.

Taylor, Curtis R. 2004. Consumer privacy and the market for customer information. *The RAND Journal of Economics* **35**(4) 631–650.

Varian, Hal R. 1982. The nonparametric approach to demand analysis. *Econometrica* **50**(4) 945–974.

Varian, Hal R. 1984. *Microeconomic analysis*. W. W. Norton & Company.

Varian, Hal R. 2006. Revealed preference. Michael Szenberg, Lall Ramrattan, Aron A. Gottesman, eds., *Samuelsonian economics and the twenty-first century*. Oxford University Press, 99–115.

Xiao, David. 2013. Is privacy compatible with truthfulness? *Proceedings of the 4th Conference on Innovations in Theoretical Computer Science*. ITCS '13, 67–86.