Emergent Privacy

Ran Wolff

Yahoo Labs – Haifa, Israel

July 30, 2014

Abstract

Defining privacy is a long sought goal for philosophers and legal scholars alike. Current definitions lack mathematical rigor. They are therefore impracticable for domains such as economics and computer science in which privacy needs to be quantified and computed.

This paper describes a game theoretic framework in which privacy requires no definition per se. Rather, it is an emergent property of specific games, the strategy by which players maximize their reward. In this context, key activities related to privacy, such as methods for its protection and ways in which it is traded, are given concrete meaning.

Based in game theory, emergent privacy demonstrates that the right to privacy can be derived, at least in part, on a utilitarian philosophical basis.

1 Introduction

Understanding privacy is a problem which attracts immense attention from philosophers and legal scholars. While appearing in Western thought since antiquity, the modern discussion of privacy is usually considered to have begun in a seminal paper by Warren and Brandeis [1890], who claimed that a right to privacy should be recognized. However, the philosophical justification and the limits of that right have been a battleground of opposing visions ever since. Unlike other rights, (e.g., property rights), the right to privacy has no clear philosophy that serves as an accepted basis for legislation.

Because personal information has gained such enormous economic importance over the last few decades, both economists and computer scientists, especially data miners, have become intensely occupied with privacy. However, the lack of semantic clarity hinders efforts to model, quantify, trade and protect privacy. Furthermore, current trends in the philosophy of privacy seem to be making semantic clarity harder to achieve. Solove [2008] writes "This struggle ultimately made me recognize that privacy is a plurality of different things" and Nissenbaum

Electronic copy available at: http://ssrn.com/abstract=2193164

[2004] talks of privacy as a set of "norms of information flow" that are different from one context to another. To a data miner, if privacy is not one but a multiplicity of concepts, then the challenge of protecting each one becomes daunting.

However, the greatest challenge to information practitioners is not the possibility that privacy is complex, but rather the repeated claims that the right to privacy makes no sense. Posner [1978], a legal theorist, economist and judge on the United States Court of Appeals, examines privacy through the prism of market economy, in which information is an intermediate good. Posner forcefully argues that shielding personal information in privacy is illegitimate. Why, asks Posner, "... would someone want to conceal a fact, except to mislead others in transacting with him?"

Posner's argument runs deeper than mere market analysis. The philosophical position he takes is that of utilitarianism. According to this position, rights are desirable if they promote society's well-being at large. In today's world, where the vast majority of personal data is provided in exchange for services by companies such as Facebook and Google, utilitarianism must be given proper respect: Why should the legislator intervene and regulate a market in which information flows freely under the consent of the individuals who are the subjects of that information?

The challenge posed by Posner can be addressed by refuting it or by providing a positive example of privacy protections that are in the interest of all involved. As for refutations, Posner himself states that law does not necessarily follow economic rationale, which is another way to say that not all rights are derived on a utilitarian basis. Furthermore, one could argue that the economics of privacy, in today's conglomerated information technology arena, can hardly be represented as an efficient market. As for positive examples, Posner admits that sometimes the information provided can detract from some greater reward. For instance, providing the answers to an exam before the students are examined will do more harm than provide utility.

Recently, Kadane et al. [2008] extended this line of reasoning. First, they proved that under the Bayesian probability framework all information necessarily has positive value, just as Posner argued. However, they have shown that under three different accepted extensions of the Bayesian framework, information can indeed have negative value. Kadane et al. provide several examples in which cost-free information may have negative value. These includes when learning the information in a certain way has its own merit, as in a suspense movie where a spoiler has a negative value, and situations in which information would be share with all players at once, including some who may use that information to harm the decision maker. Kadane et al. do not give an example of a case where a decision maker would wish to forever avoid learning a piece of information which is known to another. Yet, as we show, this is the exact scenario at which privacy norms apply.

This paper extends the results of Kadane et al. [2008], using the minimax framework of game theory, which is one of the three they review. We begin by focusing on a specific piece of information, a secret, which is known to

one player and not to the other. The player who has the secret can use it to perform deterministic actions which seem random to another player. Then, we describe a simple game in which randomized (mixed) strategies offer a higher reward to all players. In such games, all of the players have an interest in maintaining the secret so that mixed strategies remain possible.

The second logical step this paper makes is interpreting this game theoretic framework as a model for the real-world behaviors of respect for privacy and protection of privacy: In everyday terms, an individual respects the privacy of another individual when the first intentionally avoids observing the actions of the second. Society can intervene to encourage individuals to be more respectful of privacy by implementing privacy protection mechanisms. These include legislation which punishes, for example, window peeking, but also mechanisms such as lockable bathroom doors, anonymous talk-back systems, encrypted communication, and privacy preserving data mining algorithms.

The rest of this paper is organized as follows. Section 2 provides a short introduction to game theory and pseudo-randomness. Section 3 clarifies the meaning of secrets in game theory. Section 4 describes a game in which respect for privacy emerges as a strategy. Section 5 follows with the limitations of self-motivated privacy-respectful strategies and Section 6 shows how privacy preserving mechanisms can compensate for those limitations. Finally, we describe the relation of this work to other works on privacy and its economics in Section 7 and discuss the possible implications of this work and directions for further research in Section 8.

2 Preliminaries

In the context of game theory, a game is fully defined by a set of players, the choices each player can make, and the reward each player receives for every combination of choices made by all players. Choices are defined in terms of their relative order and in terms of the information every player has, when making a choice, about previous choices.

When two players make their choices simultaneously it is customary to summarize the game in a table. In such a table the columns correspond to the choices of one player – the column player – and the rows to the choices of the other – the row player. In every cell of the table, two numbers are depicted; these numbers designate the utility of the column and the row player. To simplify the discussion of the game, we denote the players by real names: Alice, Bob, and David.

2.1 Strategies and Nash equilibrium

The combination of choices a player makes is denoted a strategy. A strategy can be pure (i.e., deterministic) or mixed (i.e., involve randomness). We expand upon the relation between the two below. One of the most

important properties a game can have is a Nash equilibrium – a combination of strategies which are stable in the sense that, given the strategies of other players, no player can increase his utility by changing his own strategy.

In games which have just one Nash equilibrium, the combination of choices leading to that equilibrium can be thought of as a solution for the game. Players would consider any other combination of strategies irrational for at least one player, and therefore unachievable. It is important to note that a Nash equilibrium is not part of the definition of a game but rather an emergent property. The topic of this paper is games in which a specific kind of Nash equilibrium emerges: one in which players choose to respect the privacy of other players.

2.2 Randomness and pseudo-randomness

In some games, the utility of a player can be increased if he makes some of his choices at random. Randomness provides that the other players cannot anticipate that choice when they calculate their own choices. When random choices are made, the strategy is denoted *mixed* and the expected utility replaces the deterministic utility which is used in games with a pure strategy. Mixed strategies are sometimes explained as choices made by a special player, Chance (a.k.a. Nature), whose behavior s governed by a probability distribution over the possible choices.

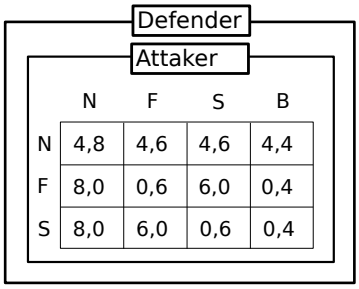
In this paper, we prefer discussing pseudo-random choices rather than random ones. A pseudo-random generator is an algorithm which is given an input, often denoted a *secret* or a *seed*, and provides a sequence of bits, or coin flips. Pseudo-random generators have the following key property: Given the secret, they are deterministic algorithms. However, to anyone who is oblivious of the secret, their output is *indistinguishable*[1] from a random sequence of bits.

For the purpose of game playing, there is no difference between a player who bases his choices on randomness, and a player who bases his choices on pseudo-randomness[2], so long as the input to the pseudo-random generator remains a *secret*. However, limiting the discussion to pseudo-randomness has three main benefits: First, it narrows the possible strategies of the players to pure ones (which may still use the deterministic pseudo-random generator.) Second, it allows interpreting specific choices related to keeping or revealing secrets as choices leading to pure or pseudo-mixed strategies. An example of such a choice follows in the next section. Third, when more than two players are involved, pseudo-randomness allows a player to maintain a single strategy which appears to a second player – who does not know the first player's strategy – to be mixed. Yet that same strategy can be pure in the eyes of a third player who does know the first player's secret input.

[1] Indistinguishability is defined in the strong, cryptographic, sense on which we do not extend.

[2] This observation is not very original. Rasmusen [1994] states that mixed strategies are a "... good description of the world is that the actions appear random to observers". I.e., the importance is on an appearance of randomness rather than on true randomness.

Figure 3.1: Game of defense in pure strategy



3 The value of secrets

We begin the discussion of game theoretic privacy by using a known game, the game of defense Dighe et al. [2009], to exemplify how properties such as secrecy, which closely relate to privacy, can be laden with economic meaning. In the context of game theory, a secret has value and meaning only if a game can be described wherein the possession of a secret improves the player's outcome.

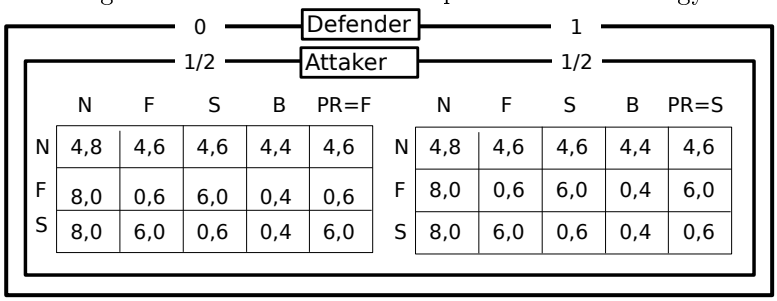
Consider a game between a defender, David, and an attacker, Alice. David has a small fortune and two houses. As depicted in Fig. 3.1, Davids choices are to spend part of his fortune hiring a guard to protect either the first house (F), the second house (S), both houses (B), or not to hire any guard (N). Alice has her own fortune. Her choices are to do nothing (N), or to invest her fortune in buying a gun and attacking either the first house (F), the second one (S). If Alice manages to capture any of the houses, then she can demand ransom and David would have to give her all of his fortune to get the house back. However, if the house is guarded, then Alice will lose her investment in the gun.

The game can be analyzed with pure or pseudo-mixed strategies: If David is only allowed pure strategies, then Alice can calculate where he will place a guard before she chooses to make her investment. Otherwise, David can employ pseudo-randomness and Alice has no way to tell whether a house is guarded other than by trying to break in.

Figure 3.1 depicts the first game when David's initial fortune is eight and Alice's is four, the cost of each guard is two, and the cost of a gun is four. Since David does not have a secret, his strategy is to hire two guards. Any other choice would lead to Alice attacking his unguarded house and David losing the rest of his fortune. Alice's response is a strategy of not buying the gun and not attacking, retaining her original four.

Figure 3.2 depicts the second game with the same initial fortunes and costs. This time, it is assumed David does have a secret. With that secret, David can use a pseudo-random generator to make choices which will appear random to Alice. In other words, David can now select the pseudo-mixed strategy of placing a single guard at the house chosen by the pseudo-random generator (PR). To David, this is a deterministic choice because the probability of placing the guard in the first house is zero and the probability of placing the guard in the second

Figure 3.2: Game of defense in pseudo-mixed strategy



house is one. However, when David chooses to follow the pseudo-random generator, Alice has no way to tell if the guard was placed in the first or in the second house. Her expected utility if she chooses either F or S is just three. Therefore, she is better off retaining her original fortune. The Nash-equilibrium in this game is therefore (PR=S,N) and it yields David six and Alice four.

The difference between David's utility in the game described in Figure 3.2 and his utility in the game described in Figure 3.1 is a direct result of his possessing a secret. If David is given the choice between playing the defense game with or without a secret, he will choose to play with a secret. Thus, secrecy – choosing to keep secrets – emerges as a concept which can be derived mathematically from the game.

4 The value of respecting privacy

It is not too surprising, for anyone observing the real world, to find that players can sometimes increase their utility by holding on to their secret. Showing that respect for privacy has an economic justification requires a different game, one in which a player chooses not to obtain another player's secret. Intuitively, this should be a game in which one player's ability to use a pseudo-mixed strategy benefits all players. If that is the case, then the other players should preserve this ability by not observing the secret.

4.1 Safe fall-back game

We begin by introducing the safe fall-back game in which it is mutually beneficial that players can use a pseudo-mixed strategy. In this game, if either Alice or Bob is restricted to pure strategies, as described in Fig. 4.1, then the Nash-equilibrium is that Alice chooses C and Bob M and their expected utility is two for each. Note that the choice of C guarantees Alice's utility of two regardless of Bob's choice, and vice-verse. On the other hand, no other choice is a good deterministic choice for Alice because Bob always has a choice which would increase his utility to eight while Alice's utility drops to zero. In the same why, Bob cannot deterministically choose anything but M.

Figure 4.1: Safe fall-back game in pure strategy

Bob \ Alice	L	LC	C	RC	R
T	5,5	0,8	0,2	8,0	5,5
TM	8,0	0,0	1,2	0,0	0,8
M	2,0	2,1	2,2	2,1	2,0
BM	0,8	0,0	1,2	0,0	8,0
B	5,5	8,0	0,2	0,8	5,5

Contrary to their pure strategies, if Alice and Bob each has a secret then Alice can choose L or R based on a pseudo-random generator and Bob can choose T or B based on his own pseudo-random generator. This results in the game depicted in Fig. 4.2. If Alice follows this pseudo-random strategy, then Bob's expected utility from selecting either T or B is higher than the expected utility of choosing TM, M, or BM. So following the strategy of choosing one of T and B as dictated by the pseudo-random generator is rational to Bob as well. The result of this strategy is the one of the Nash-equilibria (L,T), (L,B), (R,T) or (R,B) depending on the secrets. The utility for both Alice and Bob is therefore five, which is higher than they can expect when they do not have a secret.

4.2 Respect other's secret

As we have seen, it can sometimes be better for both players if each of them has a secret. However, we should not prepossess that each player controls his own secret. It may well be that Alice can decide if she obtains Bob's secret and the vice-verse. Consider two step game, in which Alice and Bob first make a choice regarding each other's secret, and then proceed to play a safe fall-back game. The choice they make on the first step is whether to obtain the other player's secret. In other words, the choice is whether to respect the other's privacy or to invade it.

For the time being, we assume that the safe fall-back game is played in full knowledge of the choice made in the first step. Thus, both Alice and Bob know by this stage which of them still has a secret. If both still have their secret, then the game they play in the second step is the one depicted in Fig. 4.2, else, they play the game depicted in Fig. 4.1. Since each of these games has a known Nash-equilibrium, its utility to both players is known. We summarize the 24 by 24 state table corresponding to all of the choices Alice and Bob have to Figure 4.3. If both Alice and Bob choose to respect (R) each other's privacy, then they both have an expected utility of five. If either one invades (I) the other's privacy, then that other player will opt for the safe fall-back of choosing C or M, respectively, resulting in a Nash equilibrium in which each has a utility of two.

We conclude that the strategy which leads to a Nash Equilibrium in the full game, which we henceforth denote

Figure 4.2: Safe fall-back – pseudo-mixed strategy

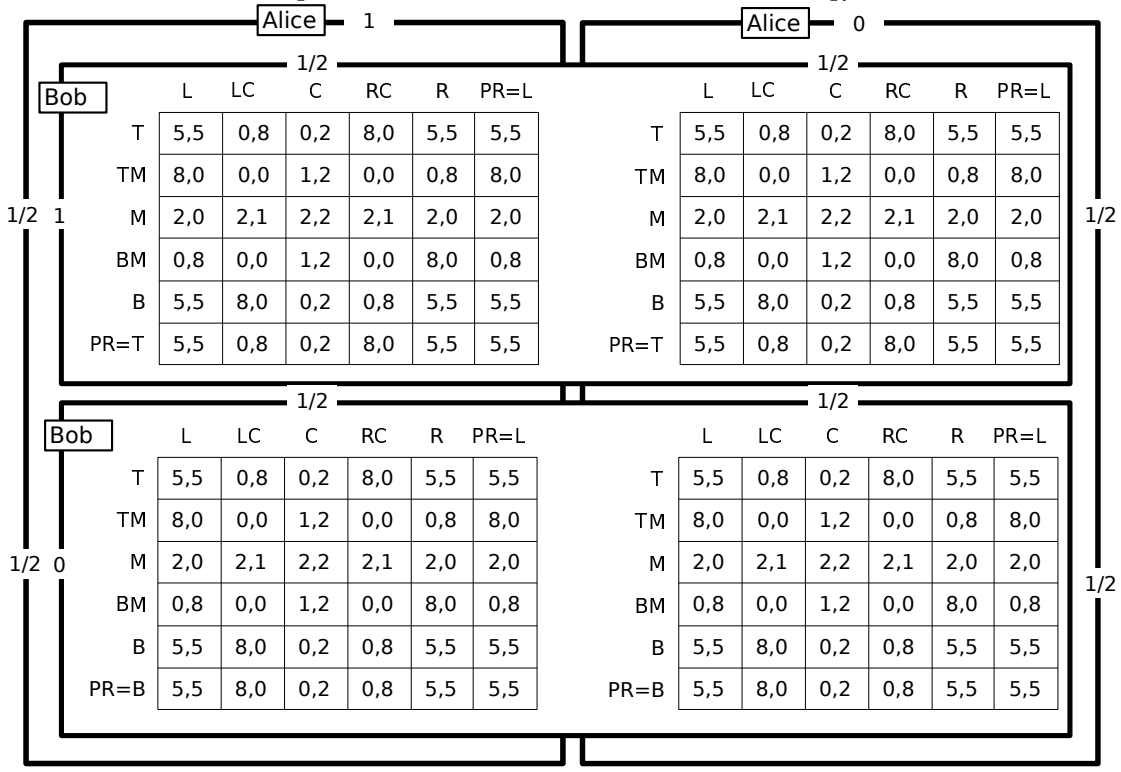
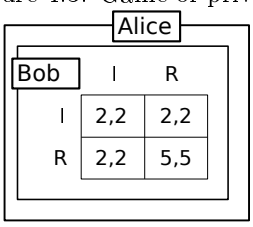


Figure 4.3: Game of privacy



the *game of privacy*, is for both players to make the following choices: First, each deterministically chooses to respect the other player's privacy. Then, If Alice invaded Bob's privacy then he deterministically chooses M and she chooses C. Likewise, if Bob invades Alice's privacy then she chooses C and he chooses M. Only if both chose to respect each other's privacy then Alice chooses pseudo-randomly L or R and Bob chooses pseudo-randomly T or B.

This strategy depends on the assumption, which is removed in the next section, that each player has full information about the other players choice of respecting or invading his or her privacy. Still, the important feature of this game is that the strategy of each player is to respect the privacy of the other player. Respect for privacy is dictated by the definitions of the game of privacy: players, choices, and utility. The concept of privacy per se does not require any definition.

5 Limitations of a privacy-respectful strategy

The examples in the previous section have shown that in some games respect for privacy is the rational behavior of players. However, they rely on one main assumption, which is that players have full information about the choices other players make with regard to secrets. This section shows that when this assumption is dropped, players often fail to achieve the benefits of a strategy that respects privacy.

We choose to relax the assumption regarding full information in two ways: One, denoted *possible intrusion*, is that players know that with probability p the other player respects their privacy. The other, denoted *partial intrusion*, is that players know that the others gained some knowledge of their secret, which allows them to guess the binary outcome of the pseudo-random generator with probability $\frac{1}{2} + \lambda$.

5.1 Possible intrusion

Consider what would happen if the Alice's strategy were to choose pseudo-randomly, with probability p_A, whether to violate Bob's privacy. Likewise, Bob would choose pseudo-randomly, with probability p_B, whether to invade Alice's privacy. How would the game of privacy proceed?

Making the choice pseudo-randomly adds the following ingredient to the game: If Alice (respectively, Bob) chooses not to intrude on Bob's privacy then she will never know if Bob has intruded on her privacy. On the other hand, if Alice does intrude on Bob's privacy then she can recover his decision on whether to intrude on her privacy (knowing his secret, that decision becomes deterministic.) The game is therefore played in one of four variants, depending on the pseudo-random decision each player makes regarding the other player's privacy. If Alice intrudes on Bob's privacy then she will know if they both intruded on each other privacy (I-I variant) or did he respect her privacy (I-R variant). If Alice respects Bob's privacy then she would not know if Bob chose to

intrude on her privacy (R-I variant) or whether he intruded her privacy (R-I variant).

In the I-I variant, the game simply becomes the one depicted in figure 4.1. The Nash equilibrium here is (C,M) and the utility of each player is two. In the I-R variant, Alice's choice is also simple, if Bob chooses M then she will choose C and both their utility will be two. For any other choice Bob makes, Alice has an option which will guarantee her a utility of eight and him zero.

Alice's choice is the most complex in the R-I and the R-R variants. Alice can tell the probability p_B with which Bob will intrude on her privacy. But since she chose not to intrude on his privacy, Alice cannot tell if she is playing the R-I or the R-R variant. With probability p_B, Alice is in the R-I variant where choosing C has a utility of two and choosing otherwise has a utility of zero. With probability $1 - p_B$, she is in the R-R variant where choosing C has a utility of two and choosing to follow the pseudo-random generator, as she did in the game in Fig. 4.2, has a utility of five.

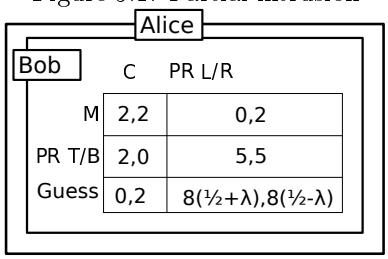
The Nash-equilibria of this version of the privacy game are, therefore, as follows: If the chances that Bob respects Alice's privacy are below $\frac{2}{5}$ then Alice's expected utility from the second strategy is lower than her utility for always choosing C. Since the game is symmetric we can say that for $p_A > \frac{3}{5}$ or $p_B > \frac{3}{5}$ the Nash-equilibrium is (C,M) and the expected utility is two for each. For $p_A, p_B < \frac{3}{5}$ the second strategy offer's a higher expected utility:

$$\begin{aligned} E\left[A|p_B < \frac{3}{5}\right] &= 2p_Ap_B + 8p_A\left(1-p_B\right) + 5\left(1-p_A\right)\left(1-p_B\right) \\ &= 5 - 5p_B + 3p_A - p_Ap_B \\ E\left[B|p_A < \frac{3}{5}\right] &= 5 - 5p_A + 3p_B - p_Ap_B \end{aligned}$$

Clearly, both $E\left[A|p_B < \frac{3}{5}\right]$ and $E\left[B|p_A < \frac{3}{5}\right]$ are greater than two. Therefore, the Nash-equilibrium for $p_A, p_B < \frac{3}{5}$ is the strategy of pseudo-randomly choosing whether to intrude on the other's privacy. Then, if the choice is to respect privacy, choose pseudo-randomly L or R and T or B.

However, we must not neglect another strategic choice, which is the choice of p_A and p_B. This choice makes the game one in which there is a continuum of choices. To analyze this type of game we need to look at the derivatives $\frac{\partial E[A]}{\partial p_A} = 3 - p_B$ and $\frac{\partial E[B]}{\partial p_B} = 3 - p_A$. The important characteristic of the derivatives is that they are both positive. I.e., increasing p_A is always the better strategic choice for Alice as is increasing p_B for Bob. This leads us to the sobering conclusion that if Alice and Bob can choose the chances of intruding on each other's privacy, then they will do this at the maximal probability which does not reach the tipping point. I.e., p_A and p_B will both tend infinitesimally to $\frac{3}{5}$. This makes the game a continuous choice version of the nutritious prisoners' dilemma, because increasing the probability of intrusion decreases the expected utility.

Figure 5.1: Partial intrusion



5.2 Partial intrusion

Consider what would happen in the privacy game if Alice's pseudo-random generator was imbalanced. I.e., if Bob could guess Alice's pseudo-random decisions in probability $\frac{1}{2}+\lambda$ for $\lambda \in \left(0, \frac{1}{2}\right)$. This would allow Bob a new strategic choice of trying to follow a guess of Alice's choice. If Alice's choice is not C, then Bob has a utility of eight for a correct guess and zero for an incorrect choice and Alice has the opposite. In expectancy $E\left[B|guess\right] = 8\left(\frac{1}{2}+\lambda\right)$ and $E\left[A|guess\right] = 8\left(\frac{1}{2}-\lambda\right)$.

Figure 5.1 summarizes the expected utility for Alice and for Bob in this version of the game. For $\frac{1}{4} > \lambda > \frac{1}{8}$, this game has a Nash-equilibrium at which Alice pseudo-randomly chooses L or R and Bob follows his best guess of what Alice's choice was.

We conclude that partial intrusion, in the context of the privacy game, does not necessarily reduce utility and can be used to shift the utility from one player to the other. This can make partial intrusion a good model for cases in which players trade their utility in the privacy game for some other utility, possibly accounted for in another game.

6 Emergent privacy protection

In the two examples provided in the previous section, control of privacy invasion was critical to materializing the potential gains from a privacy-respectful strategy. In potential intrusion, if Alice and Bob cannot be certain their privacy will be respected they end up invading each other's privacy with a probability of $\frac{3}{5}$ and with an expected utility of $3\frac{9}{25}$ rather than the optimum of 5. Likewise, in partial intrusion, Alice can only continue using the pseudo-random generator if she is sure that Bob can only guess the result with probability less than $\frac{3}{4}$.

This section reviews two mechanisms which can be used to establish such certainty: punishment and k-anonymity. Thematically, the discussion of mechanisms which assure greater utility belongs to the area of mechanism design. However, the intention here is not to discuss the design problem but rather to express some well-known privacy protection mechanisms in the language of game theory.

6.1 Punishment

In real life, the most well-accepted mechanism for preserving privacy is punishment in the form of social condemnation or legal consequences. Punishment mechanisms assume that either the player whose privacy is invaded, or a third player (the government) would, with some probability p_e, know of the privacy invasion. It is further assumed that the discoverer has both the capability and the commitment to levy a fine f on the invader.

Consider the effect of punishment on the game of privacy with possible intrusion (Section 5.1). The expected punishment changes the utility for any combination of choices which involves invading privacy. Taking Alice for instance, her expected utility in the I-I variation (Invade-Invade) would no longer be 2 but $2 - p_e f$. In the I-R variation she will gain $8 - p_e f$. In the R-I and R-R variation her expected utility would not change, because being respectful of privacy she will not be punished, and will still be $5\,(1 - p_B)$, so long as $p_B < \frac{3}{5}$. In summary, her expected utility for given probabilities of intrusion, p_A and p_B, is $E\,[A] = 5 - 5p_B + 3p_A - p_A p_B - p_A p_e f$. Likewise, Bob's utility is expected to be $E\,[B] = 5 - 5p_A + 3p_B - p_A p_B - p_B p_e f$.

Importantly, the addition of a potential punishment changes the derivative of the utility. $\frac{\partial E[A]}{\partial p_A} = 3 - p_B - p_e f$ and $\frac{\partial E[B]}{\partial p_B} = 3 - p_A - p_e f$. If $p_e f$ is sufficiently large then both Alice and Bob would strategically choose to decrease the probability at which they intrude on each other's privacy. I.e., a sufficiently large punishment on the discovery of intrusion is effective in resolving the prisoners' dilemma. However, unless $p_e f$ is greater than three, a new Nash equilibrium will be reached for p_B and p_A that are still greater than zero. Thus, a punishment can encourage respect for privacy, but the expected fine must exceed a threshold (three, in this case) to bring the probability of privacy intrusion to zero, and the expected utility of the game to the optimum.

6.2 k-Anonymity

One of the oldest methods of preventing loss of privacy is providing anonymity[3]. The modern use of anonymity in terms of computerized data is to release the secret data of groups of at least k individuals who are indistinguishable from one another. Without going into detail about the potential usefulness of such secrets (see Samarati and Sweeney [1998], Friedman et al. [2008], Ohm [2010]) emergent privacy can be used to explain how k-anonymity provides privacy protection.

Consider k players who take the roles of Alice, A_1 through A_k, and k who take the role of Bob, B_1 through B_k. Each Alice A_i plays the game of privacy with the Bob B_i. Alice can access the secrets of all Bobs. However, her only way of accessing the list of secrets is via the government, which collects all secrets and randomly permutes the list before it is released.

[3] The Talmud, a third century Jewish scripture, advises that "If a man sees that his [evil] desire is conquering him, let him go to a place where he is unknown, don black and cover himself with black, and do as his heart desires, but let him not publicly profane God's name." In modern language, the Talmud prescribes anonymity.

Hence, each Alice A_i is given k secrets and she knows one of these secrets belongs to the Bob B_i whose choices determine her utility. In lack of further information, A_i's best guess of Bob's pseudo-random choice is the choice most frequently dictated by the pseudo-random generator when used with the secrets of each of the Bobs. The probability that she invades B_i's privacy is the probability that his pseudo-random choice is the same as that of the relative majority.

Assume the pseudo-random choices are binary and have equal probability, as is the case in the game of privacy. For B_i to be among the majority, $\frac{k-1}{2}$ or more out of the other $k-1$ choices should be the same as his. The probability of that converges to $\frac{1}{2}$ from above as k grows. It follows that for any desired $\lambda \in \left(0, \frac{1}{2}\right)$, k-anonymity can be used to assure that Alice cannot invade Bob's privacy to a degree higher than $\frac{1}{2} + \lambda$. k-anonymity is therefore a mechanism which guarantees partial intrusion as discussed in Sec. 5.1.

7 Related work

One of the important tests for any new model of privacy is how well it integrates existing models. This section investigates the relation of emergent privacy and two existing models of privacy: Solove's pluralistic concept of privacy and Nissenbaum's privacy as contextual integrity. Additionally, we remark on some related work in economy and specifically on related results in game theory.

7.1 Pluralistic privacy

In his excellent book on privacy, Solove [2008] thoroughly reviews a list of conceptions about privacy as well as their critique. Solove summarizes these by stating that existing concepts of privacy are either too narrow, too broad, or too vague. He then concludes that privacy

> "involves a cluster of protections against a group of different but related problems. These problems impede valuable activities that society wants to protect, and therefore society devises ways to address these problems."

Solove then lists several of the possible harms which are prevented via privacy protection.

Emergent privacy fully agrees with Solove's description of privacy. The difference between emergent privacy and Solove's description is that we present the "problems" and the "ways" as games, strategies and mechanisms. These, we express with mathematical rigor. While any social "problem" can be expressed as a slightly different game, they are related to one another and can be clustered. Also, while there are different "ways" in which society addresses problems, many of them map to one of the mechanisms described in Section 6. Society often demotivates intrusion of privacy by legislating punishment. Otherwise, society often guarantee intrusion limitations

by assuring a certain level of anonymity or by intentionally creating ambiguity.

7.2 Contextual integrity

In an influential paper turned book, Nissenbaum [2004] discusses privacy in terms of contexts. Nissenbaum observes that in any social context there are norms which govern the flow of information. Specifically, she points to a norm of appropriateness, which dictates what is acceptable for an individual in a context, and a norm of information distribution, which states what information should be transferred.

Importantly, those norms are different, perhaps even contending, from one context to the other. Therefore, as an individual moves from one social context to the next, there is a need to limit the flow of information. Otherwise, the individual's behavior in one context might seem inappropriate in another.

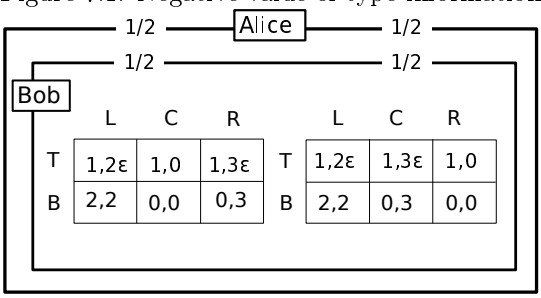
Without doubt, contextual integrity is an accurate description of the experience of privacy. However, Nissenbaum does not provide any specific reasoning for the specific choice of norms, quoting various possible reasons, including history. Thus, while contextual integrity improves our understanding of what people desire in privacy, the model does not provide any reason why such rights should be granted.

Emergent privacy complements contextual integrity in two ways. Firstly, it allows understanding the central concept of a social context as a game. A player participating in more than one game has to act in multiple contexts. In one game (context) the strategy may dictate keeping a secret, which in Nissenbaum's terminology would constitute an information limitation norm. Yet, in a different game (context) the same secret had better be shared, which in Nissenbaum's terminology constitutes an information distribution norm. Emergent privacy therefore provides the reason such norms exist, as well as an explanation for the difference between the norms of different social contexts. Secondly, because it is able to explain context and privacy norms in a utilitarian framework, emergent privacy provides argumentation for fulfillment of the desire for privacy. Additionally, emergent privacy provides the mathematical rigor which contextual integrity lacks.

7.3 The economy of privacy

The economy of privacy and of private information has attracted much interest in the past few decades. However, very little of that work have ventured into discussing what privacy is in the first place. Most of the work reviewed (with the noted exception of Posner [1978] and Kadane et al. [2008]) adopts one of two approaches: Either assume individuals place value on privacy as given, or assume any information which relates to an individual and can be harmful in a business transaction is private. Listings of research in the intersection of privacy and economics can be found on the Web (see, e.g., Acquisti). Below, two papers are outlined which deal with concepts bordering that which is discussed in this paper.

Figure 7.1: Negative value of type information



Calzolari and Pavan [2006] investigates the conditions under which an upstream seller would withhold information about an agent from downstream sellers. Thus, providing the agent privacy. These conditions relate to a situation in which the downstream seller would benefit from the information, at the agent's expense. Emergent privacy, instead, is interested in cases in which no-one would benefit from passing the information to the downstream seller.

Huang [1998] discusses the economics of privacy regulation in the context of data mining. Huang lists several reasons why individuals should be concerned with their data being mined. He also applies game theory to the analysis of the conditions necessary for adoption of self regulation by data mining companies, as well as mechanisms which enhance privacy – including penalties (i.e., fines). However, Posner [1978] specifically rejects the argumentation that the knowledge obtained by the company allows it to benefit at the expense of the individual as a legitimate basis for privacy rights. Emergent privacy overcomes this rejection because it deems private only information whose sharing is harmful to both the company and the individual.

7.4 Negative value of information

That information may have negative value is by no means new. Osborne Osborne [2003] describes a game in which some information has a negative value. As shown in Fig. 7.1, the game consists of players, Alice and Bob, who are both oblivious as to which of the two sub-games they are playing (their *type*). For $0 < \epsilon < \frac{1}{2}$ the Nash-equilibrium in this game is (L, B): If Bob selects T then Alice prefer the 2ϵ utility of L to a 50% chance of a 3ϵ which is her utility in selecting M or R and if Bob chooses B then Alice prefers utility 2 over a 50% chance of 3. This allows Bob to rationally prefer B.

However, it is irrational for Bob to choose B if Alice knows which sub-game she plays. If she does know the type and Bob chooses B then Alice would choose R in the left sub-game and M in the right sub-game yielding him a utility of zero whereas Bob can guarantee a utility of 1 by choosing T. It turns out that Alice's additional knowledge degrades her utility from 2 to 3ϵ and Bob's utility from 2 to 1. The value of information on the type is therefore negative.

The main difference between the game presented by Osborne and the kind of games discussed in this paper is information which is hidden or revealed. Our work focuses on secrets in their cryptographic interpretation – pieces of information which allow players to act in a pseudo-random way. In games such as the one presented in Fig. 4.2 the types of the players are the same. While it may be possible to emulate an n-bit secret with 2^n different types of games we feel that this would misrepresent both the nature of a secret and the intention behind player types.

The second difference, which is still important, is that in games such as those discussed in this paper the information whose value to the player is negative is information about another player. This is in line with the intuition that Alice respects Bob's privacy by avoiding to learn something about Bob, and not something about herself.

8 Discussion

This paper presents emergent privacy – a concept of privacy which relies on utilitarian arguments and can thus be shown to benefit those who respect it. Utilitarianism is by far not the only possible philosophical grounds for privacy rights. For instance, it would be hard to justify privacy in extreme settings, such as in jail, based on strictly utilitarian arguments. It is therefore only reasonable to assume that there are cases in which the right for privacy is claimed, and perhaps granted, even if the benefits from establishing such a right cannot be spelled out. Emergent privacy should be considered a minimalistic definition, rather than a complete one.

Nevertheless, emergent privacy is by no means a powerless concept. Since its definition is simple and rigorous, it is a powerful tool with which related concepts such as privacy protection and privacy breach can be understood. It can be used to infer the consequence of choosing the parameters of privacy protection mechanisms such as punishment and anonymity. Through the lens of emergent privacy, the diversity of privacy concepts can be understood as different strategies in different games which all aim to protect the same type of behavior.

A mathematical model of privacy is very limited unless it can be applied to large populations as well as address the inherent differences between abstract players and human beings. In upcoming work, we intend to place emergent privacy in the context of collaborative action. Additionally, we intend to explain, using the psychological model of social representations, why setups which can be represented as games played with a privacy-respectful strategy really do occur in society.

The main technology that disrupts privacy today is data mining. It is therefore a major open research question whether existing methods for privacy preserving data mining can be described as privacy protection mechanisms. It is equally interesting to see whether standard applications Web search, fraud analysis, and pharmaceutical research – can be modeled as games. If we can answer both questions in the positive, then emergent privacy can

provide meaningful recipes for privacy preserving data mining in these applications.

References

Alessandro Acquisti. The Economics of Privacy. http://www.heinz.cmu.edu/ acquisti/economics-privacy.htm.

Giacomo Calzolari and Alessandro Pavan. On the optimality of privacy in sequential contracting. *Journal of Economic Theory*, 130(1):168–204, 2006.

Nikhil S. Dighe, Jun Zhuang, and Vicki M. Bier. Secrecy in Defensive Allocations as a Strategy for achieving more Cost-effective Attacker Deterrence. *International Journal of Performability Engineering*, 5(1):31 – 43, 2009.

Arik Friedman, Ran Wolff, and Assaf Schuster. Providing k-Anonymity in data mining. *VLDB Journal*, 17(4): 789–804, 2008.

Peter H. Huang. The Law and Economics of Consumer Privacy Versus Data Mining. Social Science Research Network (SSRN), 1998.

Joseph B. Kadane, Mark Schervish, and Teddy Seidenfield. Is Ignorance Bliss? *Journal of Philosophy*, 105(1): 5–36, 2008.

Helen Nissenbaum. Privacy as Contextual Integrity. *Washington Law Review*, 79(4):119 – 158, 2004.

Paul Ohm. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review*, 57:1701–1777, 2010.

Martin J. Osborne. *An introduction to game theory*. Oxford University Press, August 2003. p. 283.

Richard A. Posner. The right to privacy. *Georgia Law review*, 12(3):393–422, 1978.

E. Rasmusen. *Games and information*. Cambridge, 1994.

Pierangela Samarati and Latanya Sweeney. Protecting Privacy when Disclosing Information: k-Anonymity and Its Enforcement through Generalization and Suppression. Technical Report SRI-CSL-98-04, 1998.

Daniel J. Solove. *Understanding privacy*. Harvard University Press, Cambridge, Mass, 2008.

Samuel D. Warren and Louis D. Brandeis. The Right to Privacy. *Harvard Law Review*, IV(5), 1890.