

# Differential Privacy as a Tool for Truthfulness in Games

Differential privacy guarantees the input data from a single individual has a very small impact on the output of a computation. Tools from privacy can also be used in game theory and economics to incentivize people to truthfully reveal their data.

*By Rachel Cummings*

DOI: 10.1145/3123742

**A**lgorithmic game theory considers settings in which self-interested agents interact and make choices to maximize their payoff, or utility. Often this payoff also depends on the joint actions of other agents. An equilibrium of a game is an action profile where each player is maximizing their utility, given the actions of others. Equilibria are desirable because they are steady states of the game, where no player wishes to change their action. In some cases, the game designer—also known as mechanism designer—may have the power to determine the rules of the game or the information available to the players. This mechanism designer may also have their own objective, such as maximizing the total utility of all players or maximizing the profit of a seller (in

the case of an auction). In settings where players' utility functions contain sensitive information, the designer should also worry about privacy concerns when communicating information to players.

Differential privacy is a notion of database privacy that provides a mathematically rigorous worst-case bound on the maximum amount of information that can be learned about any one individual's input from the output of a computation. This "maximum amount" can be tuned using a

parameter that balances the trade-off between accuracy of the computation and privacy of the individuals. The mathematical rigor of differential privacy ensures the security of the output, as it is robust to post-processing as well as the presence of external information about the data. These formal guarantees give a sharp contrast to ad hoc privacy measures such as anonymization, which has led to infamous privacy violations such as the Netflix challenge or AOL search logs release. (See Dwork and Roth for a survey of the growing

literature on differentially private algorithms applied to a wide variety of computational settings [1].)

Interest in the connection between algorithmic game theory and differential privacy is fueled by a tension between privacy of the inputs to the computation, and usefulness of the output (usually measured in terms of the players' utility or the mechanism designer's objective). Privacy alone can be achieved by outputting pure noise, but this is likely to lead to a poor outcome of the game; utility alone can be



achieved by selecting the best outcome of the game, but this may be very revealing of a particular player's input. In this article, we will explore privacy as both a requirement and a tool. It is a necessary objective when players of a game explicitly care about their privacy and when their utility depends on the privacy of their data. Privacy-preserving algorithms can also serve as a powerful tool to incentivize good behavior, even in settings where players are not concerned with the privacy of their data.

### FORMALIZING DIFFERENTIAL PRIVACY

Differential privacy guarantees if a single entry in the database were to be changed, then the algorithm would still have approximately the same distribution over outputs, where the probability of producing any output can change by at most a multiplicative  $(1+\epsilon)$  factor. This means an adversary would have limited ability to detect small changes in the database from observ-

ing the mechanism's output. This privacy guarantee is typically achieved by injecting carefully calibrated noise into the algorithm to randomize the output without significantly compromising usefulness of the computation. The privacy parameter  $\epsilon$  balances this trade-off between accuracy of the computation and privacy of the individuals:  $\epsilon=0$  provides perfect privacy because the output must be independent of the input. At the other extreme,  $\epsilon=\infty$  offers no privacy guarantees because the differential privacy constraint does not bind, and the mechanism's output is allowed to depend arbitrarily on a single entry in the database.

Differential privacy has three key properties that make it desirable for use as a privacy notion. Firstly, differentially private algorithms are robust to post-processing, making it hard for an adversary to learn additional information about the database by performing further computations on a differentially private output. Secondly, the privacy

guarantee extends to arbitrary groups of players, where the level of privacy decreases linearly with the size of the group. This property is known as group privacy. Finally, differentially private algorithms compose, meaning the privacy guarantee degrades gracefully as multiple computations are performed on the same database. This allows for the modular design of differentially private mechanisms; an algorithm designer can combine several simple differentially private mechanisms as building blocks in a larger, more complicated algorithm. Then can reason about the overall privacy guarantee of their mechanism by reasoning only about these simple mechanisms.

### GAME THEORY TERMINOLOGY AND EXAMPLES

In the field of economics, a game consists of several players who strategically choose their actions, which jointly determine everyone's utility from the game. Throughout this article, we

will use three illustrative examples of games. The first is an auction, where each player's action is their bid for the good being sold, and their utility is their value for the good (if the player wins the auction) minus the price paid. We also consider a facility location game, where a central designer, such as a government, wishes to build a hospital. The action of each player is the preferences reported to the designer about where they would like the hospital to be located, and their utility is a function of one's convenience reaching the hospital. Finally, we use a traffic routing game, where each player simultaneously wants to drive from home to work. Their action is their chosen route, and their utility is decreasing commute time, given the traffic conditions created by the routes of other players.

One primary objective of game theory is to help players arrive at an equilibrium of the game. Two different equilibrium notions are used in this article. In a Nash equilibrium, every player chooses one's best action, given the actions of others. In a dominant strategy equilibrium, every player chooses one's best action, regardless of the actions of others. Dominant strategy is a stronger equilibrium notion because it requires a player's action to be optimal for any behavior of their opponents, rather than just one fixed action profile. As such, Nash equilibria are generally guaranteed to exist, while many games do not have a dominant strategy equilibrium.

Within game theory, the field of mechanism design studies the problem(s) faced by a central designer, an entity who can influence the behavior of players by changing their utility functions, often through the use of payments that depend on a player's action. Formally, the designer must choose a mapping from the actions of all players to both the outcome of the game, and the utilities to all players. In the case of auctions, the mechanism designer can be the auctioneer selling the item and deciding how much the winning bidder will pay. In the facility location game, the government plays the role of mechanism designer; it can ask players where they live and work, and then decide where the hospital will be located. Additionally, the designer

may wish to optimize their own goals; the auctioneer may want to maximize his or her revenue, and the government may want to serve the maximum number of people with the hospital.

Another property, known as truthfulness or incentive compatibility, requires that players maximize their expected utility by reporting their true data. Games with this property incentivize players to truthfully report their data to the mechanism designer. In an auction, a player's data may be their value for the good being sold; in the traffic routing game, the player's data may be the location of their home and work.

### DIFFERENTIAL PRIVACY THROUGH THE LENS OF GAME THEORY

To better match the language of game theory and mechanism design, we rephrase the definition of differential privacy in terms of player utility functions. Now differential privacy guarantees that every player's expected utility does not change more than a multiplicative  $(1 + \epsilon)$  factor from a change in a single player's action or data report. This prohibits the utility of any player from changing significantly due to the choice to either participate in the mechanism or opt-out, or even decide to report truthfully or lie. It also promises that if a player, called Alice, changes her type report, then the utility of another player, called Bob, does not change drastically either.

This interpretation automatically implies an approximate truthfulness guarantee for any differentially private mechanism. Regardless of the actions chosen by other players, Alice can always maximize her expected utility (up to the multiplicative  $(1 + \epsilon)$  factor) by truthfully reporting her data to the mechanism. If Alice cannot gain anything from misreporting, she might as well be truthful. However, misreporting is also an approximate dominant strategy. Differential privacy promises a player's utility will be insensitive to their own report, so they could achieve approximately the same utility with non-truthful reporting, which intuitively cannot strongly incentivize truthful reporting.

### STRICT INCENTIVE COMPATIBILITY VIA PRIVACY

Stronger truthfulness guarantees can be achieved by combining differen-

tially private mechanisms with other algorithms designed to enforce strict truthfulness. One such mechanism is the punishing mechanism developed in [2], which works in environments where players can react to the mechanism's output and the mechanism designer has the power to restrict reactions. Examples of reactions to the mechanism's output include buying a good at the price selected by the mechanism, or visiting one of the several hospitals with locations selected by the mechanism.

The punishing mechanism is a convex combination of two other mechanisms. The first is the exponential mechanism, which satisfies differential privacy. It samples an output according to a probability distribution that is exponentially weighted toward the socially optimal outcomes, such as hospitals with locations in the center of the city or price points at which players can purchase their favorite good. Once the outcome is sampled, players can react freely, regardless of their data reports. The second is the commitment mechanism. It samples an outcome uniformly at random, and then requires that each player's reaction is the best response to the outcome, according to their reported data. For example, if this mechanism was used to determine the price of a good, players would be forced to buy the good if their reported value was above the chosen price and not allowed to buy it otherwise.

The exponential mechanism is known to select a good outcome with high probability, but suffers from the weak truthfulness guarantees of differential privacy. The commitment mechanism, on the other hand, is easily seen to be truthful, but is likely to choose a low-quality outcome. By randomizing between the two, the punishing mechanism enjoys both strict truthfulness and a high-quality outcome; truthfulness is enforced by the threat of the commitment mechanism, and choosing the exponential mechanism with high probability preserves the quality of the outcome.

### JOINT DIFFERENTIAL PRIVACY, AN ALTERNATIVE APPROACH TO TRUTHFULNESS

Another approach for strengthening the truthfulness guarantees of pri-

vate mechanisms is through a relaxed privacy notion, known as “joint differential privacy.” As previously discussed, differential privacy is often too strong of a notion in the context of mechanism design. Consider again the traffic routing game, where each player’s data is the location of their home and work, and their action is the route driven between these two locations. While we may want Alice’s route to work to be differentially private in the input of the other players, it is natural to allow it to be more sensitive to changes in Alice’s own input. If she changes jobs, she should certainly drive a different route.

Joint differential privacy was defined precisely for such a setting, where the mechanism’s output is partitioned among all players, where each player can only see their portion of the output. It guarantees the output for all other players besides Alice will be insensitive to Alice’s input. This protects Alice’s privacy from arbitrary coalitions of other players. Even if all other players shared their portions of the output, they would still not be able to infer much about her data. This allows the mechanism designer more power to enforce truthfulness, while still preserving the desirable properties of differentially private mechanisms.

One useful tool for proving algorithms satisfy joint differential privacy is the billboard lemma. Imagine an algorithm first computes a differentially private signal of the data, and then displays that signal publicly to all players, as if posted on a billboard. If each player’s portion of the output is computable from only this public signal and their own data, then the mechanism is jointly differentially private. In our traffic routing game, the billboard may contain the amount of traffic on all major freeways, and each player could use this information to decide their fastest route to work. This lemma allows a designer to use existing tools from differential privacy to design jointly differentially private algorithms.

## MEDIATED GAMES

Joint differential privacy can be brought to bear for implementing equilibrium behavior with the help of a mediator. The mediator collects data

## Privacy-preserving algorithms can also serve as a powerful tool to incentivize good behavior, even in settings where players are not concerned with the privacy of their data.

from all players, computes an equilibrium of the game, and suggests back to each player an action corresponding to their part of the equilibrium. The mediator is weak because it does not have the power to enforce actions or outcomes. In the running example of traffic routing, the mediator can be thought of as a navigation app—everyone enters a home address and work address into the mobile app, and the app suggests back a route for each person to drive. Each player is free to opt-out of the mediator (i.e., navigate to work without using the app), to misreport their type (i.e., lie to the app about one’s work address), or to deviate from the mediator’s suggested action (i.e., drive a different route than the suggested one).

In a mediated game, the designer would like players to truthfully report their type, and then faithfully follow the suggested action of the mediator. We call this strategy good behavior. Rogers and Roth showed if the mediator computes an equilibrium of the game under the constraint of differential privacy, then the suggested actions are jointly differentially private (by the billboard lemma), and good behavior is a Nash equilibrium [3].

The intuition behind this result lies in the usage of joint differential privacy, which ensures if Alice changes her type report the joint actions suggested to other players will remain approximately unchanged. The mediator will suggest Alice’s own best action, given this fixed action profile of the other players, if Alice reports truthfully. However, if she misreports, she

may receive an arbitrarily worse suggestion. Further, when other players follow the good behavior strategy, then deviating from the mediator’s suggestion is equivalent to deviating from a Nash equilibrium, which can only decrease Alice’s utility. Thus, each player can maximize their expected utility by truthfully reporting their type to the mediator and faithfully following the suggested action.

## CONCLUSIONS

The main results presented here did not require caring about data privacy; privacy is a tool, not necessarily an objective. A related body of work considers game theory and mechanism design for players with privacy concerns. In addition to incentivizing truthfulness, these mechanisms must also ensure every player is fairly compensated for their loss in utility incurred by providing their data to the mechanism.

The truthfulness guarantees can be achieved by interpreting differential privacy not as a privacy guarantee, but rather as a stability notion; the algorithm’s output is robust to small changes in the input. Similar interpretations of differential privacy have borne fruit in other research areas as well, such as machine learning, statistics, and optimization. I encourage the reader to look for other places in their own research where the stability properties of differential privacy may prove useful.

---

## References

- [1] Dwork, C. and Roth, A. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science* 9, 3–4 (2014), 211–407.
- [2] Nissim, K., Smorodinsky, R., and Tennenholtz, M. Approximately optimal mechanism design via differential privacy. In *Proceedings of the Third Innovations in Theoretical Computer Science Conference*. ACM, New York, 2012, 203–211.
- [3] Rogers, R. and Roth, A. Asymptotically truthful equilibrium selection in large congestion games. In *Proceedings of the 15th ACM Conference on Economics and Computation*. ACM, New York, 2014, 771–782.

---

## Biography

Rachel Cummings is a recent Ph.D. graduate in Computing and Mathematical Sciences at the California Institute of Technology. Her research interests lie at the intersection of computer science and economics, specifically problems that explore the interwoven threads of data, incentives, and privacy. She has recently joined the faculty at Georgia Institute of Technology as an assistant professor in the School of Industrial and Systems Engineering.