HDSR
HARVARD DATA SCIENCE REVIEW

Volume 0 Issue 0
DOI: 00.000
ISSN: 2644-2353

A Refreshment Stirred, Not Shaken: Invariant-Preserving Deployments of Differential Privacy for the U.S. Decennial Census

James Bailie[†,*], Ruobin Gong[‡], Xiao-Li Meng[†]

[†] Department of Statistics, Harvard University

[‡] Department of Statistics, Rutgers University

Abstract. Protecting an individual's privacy when releasing their data is inherently an exercise in relativity, regardless of how privacy is qualified or quantified. This is because we can only limit the gain in information about an individual relative to what could be derived from other sources. This framing is the essence of differential privacy (DP), through which this article examines two statistical disclosure control (SDC) methods for the United States Decennial Census: the Permutation Swapping Algorithm (PSA), which resembles the 2010 Census's disclosure avoidance system (DAS), and the TopDown Algorithm (TDA), which was used in the 2020 DAS. To varying degrees, both methods leave unaltered certain statistics of the confidential data—their invariants—and hence neither can be readily reconciled with DP, at least as originally conceived. Nevertheless, we show how invariants can naturally be integrated into DP and use this to establish that the PSA satisfies pure DP subject to the invariants it necessarily induces, thereby proving that this traditional SDC method can, in fact, be understood from the perspective of DP. By a similar modification to zero-concentrated DP, we also provide a DP specification for the TDA. Finally, as a point of comparison, we consider a counterfactual scenario in which the PSA was adopted for the 2020 Census, resulting in a reduction in the nominal protection loss budget but at the cost of releasing many more invariants. This highlights the pervasive danger of comparing budgets without accounting for the other dimensions on which DP formulations vary (such as the invariants they permit). Therefore, while our results articulate the mathematical guarantees of SDC provided by the PSA, the TDA, and the 2020 DAS in general, care must be taken in translating these guarantees into actual privacy protection—just as is the case for any DP deployment.

Keywords: confidentiality, data swapping, TopDown Algorithm, invariant statistics, statistical disclosure control.

Media Summary

Preserving data privacy when publishing statistics is a permanent challenge for every organization that provides public use data files. This article tells two stories of how the U.S. Census Bureau dealt with this challenge in its 2010 and 2020 Decennial Censuses using two different systems. By incorporating both theses approaches into a broad mathematical framework called *differential privacy*, the authors show how they can be formally understood. This study clarifies what promises

[*] jamesbailie@g.harvard.edu

these systems truly make, and what they do not, reminding us that mathematical assurances do not automatically ensure real-world privacy.

1. Motivations and Contributions

1.1. **Privacy Protection Under the Constraint of Invariants.** In 2018, the United States Census Bureau (USCB) announced an overhaul of its disclosure avoidance system (DAS), retiring the data swapping methods that had been central to protecting the U.S. Decennial Census for the previous 30 years (Abowd, 2018; McKenna, 2018). While the privacy provided by these methods had been justified with intuitive arguments, the DAS for the 2020 Census would, in contrast, be redesigned from the ground up, with the primary goal of supplying a mathematical guarantee of protection. Moreover, this guarantee, the USCB decided, must be some type of *differential privacy* (DP) (Dwork et al., 2006)—a large family of technical standards (Desfontaines & Pejó, 2020) that conceptualize the 'privacy' of a statistical disclosure control (SDC) method in terms of its sensitivity to counterfactual changes in its input data (Bailie et al., 2026b).

The USCB's adoption of DP was driven in part by the desire for a formal, quantitative, and measurable characterization of privacy protection. The bureau was concerned that the existing, swapping-based DAS lacked the rigorous basis supplied by such a characterization. Indeed, data swapping had not been analyzed by a formal system like DP, and thus a theoretical account of its SDC protection was limited. At the same time, the bureau's empirical evaluation—that is, their reconstruction and reidentification attack (see Section 6.1 and Abowd et al., 2025)—concluded that the existing DAS was vulnerable, while in comparison, a DP-compliant DAS was expected to better protect against this and other emerging types of privacy attacks (Dwork et al., 2017).

Nevertheless, there were other priorities for the 2020 Census, many of which complicated a straightforward adoption of DP. In particular, state population counts are constitutionally mandated to be published exactly as counted, whereas DP—at least as originally defined in Dwork et al. (2006)—requires that such counts be infused with random noise.

Thus, even as DP offered the kind of formal guarantees the bureau sought, its implementation had to contend with the legal, operational, and statistical requirements of the Decennial Census—requirements that impose strict conditions on how SDC is applied. Indeed, the U.S. Decennial Census is a massive exercise in population enumeration, which is subject to numerous laws and regulations, as well as various pragmatic constraints and utility considerations. Taken together, these external criteria greatly complicated the design of the new 2020 DAS. A key challenge was the requirement that census publications must respect *invariants*—exact summaries of the confidential data that must be released without modification. The most notable invariants in 2020 are the state population totals mentioned previously, but for operational and data-quality reasons the USCB also incorporated additional invariants into the 2020 Census, including counts of housing units at the lowest level of census geography (blocks) and various other statistics (summarized in Table 4).

While we have so far focused exclusively on the 2020 Census, invariants were not a novel problem in 2020. To the contrary, although their exact composition and method of imposition has changed from decade to decade, invariants have been a mainstay of every Decennial Census due to their constitutional and regulatory significance. Moreover, many other types of data dissemination frequently feature key statistics that are not altered before publication—for example, the row and column margins of contingency tables are often published as is (see Example 1 in Section 2.3).

Thus, invariants are inherent not only to the dissemination of U.S. Census data, but are also a common property of other statistical data products.

Yet the presence of invariants complicates SDC. By definition, the values of any invariant statistics cannot be modified by a DAS and as such are exempt from any SDC protection. Naturally, this creates potential statistical disclosure risks because exact knowledge of certain features of the confidential data, as is provided by invariants, may aid an attacker in reconstructing and reidentifying these data.

Furthermore, standard formulations of DP—including those that the USCB has invoked (zero-concentrated DP), referenced (approximate DP), or at some point considered (pure DP)—do not allow for the specification of invariants. The issue is that DP, at least as it is typically understood, cannot measure the protection provided to the confidential data after taking into account the release of invariants. Even though DP is fundamentally an assessment of relative privacy—the privacy of an individual's unique information relative to knowing the rest of the population—it must be recast in a more general light in order to assess the privacy relative to knowing the invariants. Yet once one recognizes DP's relative nature, it becomes apparent that invariants do not contradict the fundamental essence of DP, but rather that their incorporation into DP greatly expand its applicability.

This is because no method can protect absolute privacy (Kifer & Machanavajjhala, 2011). Hence any data protection standard, including DP, must take into account information that cannot be protected, whether that be invariants or other public information. As has largely gone unnoted, to spell this out is not only natural, but key to articulating the actual DP guarantee of any invariant-preserving SDC method, including those used for the 2020 Census. One thing that is well understood in the literature, however, is that invariants nullify the DP protection guarantees along the dimensions of the confidential data that the invariants implicate (see, e.g., Gong and Meng, 2020 and references discussed in Appendix A).

Nevertheless, the compromising effect of invariants on the DP guarantees of the Decennial Census is a central motivation of this article. Of particular importance in this regard is a DP analysis of the TopDown Algorithm (TDA), which the USCB created for disseminating key 2020 Census data products. This algorithm runs in two steps: It first adds DP-calibrated noise to all of the 2020 Census data, then it removes this noise from the invariants via a complex optimization procedure (Abowd et al., 2022). While a DP analysis of the first step is easy due to its use of an established procedure (Canonne et al., 2022), the second step, as the bureau's own assessment makes clear (Ashmead et al., 2019), is particularly challenging to analyze from the perspective of DP because of the invariants it enforces. Yet a complete and rigorous assessment of the TDA's mathematical guarantee of protection must address the entire procedure—both the noise infusion in the first step and the noise removal in the second step. This assessment is, to the best of the authors' knowledge, missing from the literature thus far.

As we have mentioned, invariants were not just a requirement in the 2020 DAS but were also present in all other Decennial Censuses. Most relevant to a discussion of the USCB's overhaul of their DAS is the 2010 Census. Since it also respected a set of invariants, the SDC protection provided by the 2010 DAS inherits all the same complications outlined in the previous section. This makes it even more important to handle invariants in a unified way, especially if one wants to compare different invariant-preserving SDC methods from within the same theoretical system—as is one of the main goals of this article.

Like the 1990 and 2000 Censuses, the 2010 DAS primarily consisted of a *data swapping* method (Dalenius & Reiss, 1982; Fienberg & McIntyre, 2004). While data swapping refers to a large class of methods and is widely employed by statistical offices around the world, in the context of the U.S. Decennial Census, swapping works by permuting the geographical data of a randomly selected subset of households (McKenna, 2018). By definition, swapping keeps invariant all counts that are unaffected by this permutation operation. More specifically, as Section 3 spells out in detail, the invariants induced by swapping are population totals over various geographic and demographic stratification, whose exact composition are determined by the particular swapping parameters used. These invariants are inevitably much more numerous than the TDA's—an important observation when comparing data swapping with the TDA because, as the number of invariants increases, their impact ranges from negligible to completely nullifying any supposed guarantee of protection.

Be that as it may, swapping does still provide some protection since it can alter any record's geographic information. According to the standard argument of how swapping provides SDC, it gives plausible deniability as to the reported location of any household, because those households that do in fact have their location changed are randomly sampled and kept secret. In this way, so the argument goes, swapping obfuscates the location, and hence the identity, of all households, thereby making it difficult for an attacker to learn any personal information. (Appendix B will critically examine this argument in detail.) While swapping has this ad hoc and intuitive justification, it has not been given a rigorous theoretical foundation—that is, it has lacked the kind of formal, mathematical guarantees of protection provided by DP (Abowd, 2017; Christ et al., 2022; Slavković & Seeman, 2023).

1.2. **Our Main Contributions.** We have thus come to a puzzling revelation. On the one hand, the new DAS in 2020 was designed according to the rigorous mathematical principle of DP; but it also accommodates invariants, compromising its DP guarantee in ways that are poorly understood and have not been properly mathematically characterized. On the other hand, data swapping—upon which the old DAS was based—was abrogated in part because it had not been studied from the perspective of DP and accordingly lacked a DP guarantee. It would be prudent, therefore, to rectify the deficiency in our understanding of the actual DP guarantees for both the 2020 DAS as well as for data swapping.

The current article does exactly this. Casting both the new SDC methods employed in the 2020 DAS and the traditional SDC technique of data swapping within what we call our system of *differential privacy specifications*, we show how each can be formally understood through the lens of DP. In this regard, this article makes four main contributions.

First, this work presents a formal treatment of invariants in DP (Section 2). Acknowledging invariants as essential to the validity, viability, and utility of the Decennial Census and other statistical disseminations, we describe why a data protection standard based only on the realized values of the invariants is not feasible, and why, in fact, a recognition of all potential and counterfactual values is a necessary part of any invariant-accommodating standard, including DP (Section 2.1). From this basis, we demonstrate how invariants can naturally be integrated into our system of DP specifications (Sections 2.2 and 2.3) and provide a preliminary, theoretical investigation into the impact of invariants on DP (Section 2.4).

Second, this article shows that, contrary to previous views, it is possible to supply data swapping with rigorous guarantees of protection based on DP (Section 3). To do this, we prove a formal description of the SDC protection—that is, a DP specification—for the Permutation Swapping

Algorithm (PSA), which is a data swapping method with similarities to the 2010 DAS (Section 3.3). Intuitively, this specification can be understood as stating that the PSA satisfies pure DP (Dwork et al., 2006) subject to the invariants it induces. While this means the PSA's specification differs from conventional formulations of DP, as we have already argued such a departure is necessary to analyze any invariant-preserving DP method and does not contradict the essentially relative nature of DP.

We note at the outset that it is impossible to properly assess the actual swapping algorithm used in the 2010 Census from the perspective of DP because the details of that algorithm are not entirely public. Yet we do know that the actual 2010 swapping procedure differs from the PSA in a number of ways (see Appendix C.1)—ways that make it difficult to characterize that procedure using DP. Nevertheless, the PSA captures the 2010 DAS in its essence by mimicking its principal design features, and therefore, its DP specification is still informative for understanding the disclosure risk of the 2010 Census. Indeed, by making the assumption that the PSA—with parameter settings chosen to resemble the 2010 data swapping algorithm—was used as the 2010 DAS, we can (and do) provide a reasonable estimate for the DP specification associated with the 2010 Census (Section 3.5).

Third, this article conducts a comparative analysis of the protection afforded to the 2020 Census with the counterfactual scenario in which the PSA was used to publish the 2020 data (Section 5). To support this endeavor, we formulate and prove the first DP characterization of the TDA (Section 4). Because, much like data swapping, the TDA can only satisfy DP subject to its invariants, this result requires the use of our system of DP specifications to formally incorporate the TDA's invariants into its DP guarantee. In addition to the TDA's specification, we also compile DP specifications for all the primary 2020 Census data products (Section 5.2). Doing so requires identifying the *protection* (or '*privacy*') *units* for the 2020 Census, which we determine to be 'post-imputation persons'—a result that is important because, like invariants, post-imputation units negatively impact the actual SDC protection afforded by a DP specification.

We aggregate the DP specifications for the various 2020 publications into a single specification describing the SDC protection afforded to the 2020 Census across all its major publications. This DP specification is then compared to that of a hypothetical application of the PSA to the 2020 Census with various parameter settings (Sections 5.3 and 5.4). This comparison is informative because it places mathematical summaries of the PSA's and the 2020 Census's SDC protection side-by-side. However, as these two specifications differ on several dimensions, drawing an overall conclusion about the relative strengths of their SDC protection remains difficult. Even so, just articulating the DP specifications for both systems is an important step forward because it clarifies and organizes the specific ways in which their SDC protection diverge.

Fourth, as part of this article's discussion (Section 6), we put forth a set of open issues with respect to the development of DP and the work needed going forward. These issues center on the difficulty just mentioned: While our system of DP specifications is a useful toolkit for describing the SDC protection of DP implementations, there are currently few tools to effectively compare two DP specifications that differ on multiple dimensions. To address this, future research is needed on exploring the trade-off between different dimensions of a DP specification—for example, trading off the presence of additional invariants with a reduced protection loss budget. We propose that such trade-offs might be assessed via their impact on disclosure risk, or via their efficiency against a privacy attack (Section 5.4). Additionally, as we do not attempt in this article to evaluate what substantive SDC protection an invariant-induced DP specification actually provides—this being

a difficult question deserving its own dedicated study—we also outline in Section 6 some future directions for understanding and mitigating the impact of invariants on disclosure risk.

Background on data swapping and other related work are provided in Appendices B and A respectively. Appendix C provides the most comprehensive (to the authors' knowledge) publicly available description of the 2010 DAS, along with a comparison between the 2010 DAS and the PSA (Appendix C.1) and a discussion of ways the PSA could be modified to further align with the 2010 DAS while still preserving its DP flavor (Appendix C.2).

2. Invariants and Differential Privacy

2.1. **Integrating Invariants Into Differential Privacy—Intuitions and Subtleties.** Intuitively speaking, the impact of invariants on SDC is similar to conditioning in statistical inference, that is, constraining the possible states of the (confidential) data by known or assumed information. In other words, any invariant-respecting SDC criterion should only consider the counterfactual data sets that share the same values for the invariants as the confidential data set—just as any conditional probability only considers outcomes that agree with the conditioning information. In that sense, the procedure for infusing invariants into a DP formulation parallels the process of disintegrating a probability into a collection of conditional distributions. The overall mathematical notion of a probability—or of DP—remains the same; the difference lies in the state spaces (i.e., the set of possible outcomes, or the set of possible data sets) to which it is applied.

At the same time, just as defining conditional distributions brings complications and subtleties (such as conditioning on probability-zero events), defining an invariant-accommodating formulation of DP requires addressing a variety of nuances, many of which are implicit in conventional DP definitions. Without making them explicit, comparing protection loss budgets across different forms of DP—especially those with different invariants—would be as meaningless as comparing the face values of different currencies without considering their conversion rates. Indeed, the problem is worse: the issue is not merely the conversion rates but, more importantly, the realizable purchasing power—that is, how much data privacy each method actually affords in terms of reducing statistical disclosure risk.

As an example of the subtleties in incorporating invariants into a DP formulation, consider the scenario in which the U.S. population size N is enumerated by the Decennial Census to be exactly 330 million. Once this value is made public, any counterfactual data set with a different value for N becomes immediately distinguishable from the actual confidential census data. Hence, such data sets must be excluded from consideration under the DP paradigm, which relies on the notion of distinguishability, as we shall review later.

However, it would be rather unwise—regardless of which DP specification we decide to adopt—to implement an SDC method that will satisfy the specification only when $N = 330,000,000$. This is because the particular value of the enumerated national total is *accidental*, in the sense that this specific total population count is, but need not be, a property of the Decennial Census (see e.g. Robertson Ishii & Atkins, 2023). Even if it were completely accurate (which it never is), it reflects only the count at the time of the census. A DP method needs to work irrespective of the enumerated value of the population total. Likewise, a DP specification should guarantee protection irrespective of this value, or the realized value of any other invariant.

On the other hand, the fact that the national population count and other invariants were published exactly as enumerated is *essential* to the Decennial Census because this occurs by design.

This distinction between the essential invariant statistics and their accidental values leads respectively to the notions of the *multiverse* and the *universe*, which will be defined mathematically in Section 2.2. For the current example, a universe is the collection of all plausible census data sets that share the same specific value for N. The multiverse is then the collection of these universes as N varies within a specified range. (Not incidentally, specifying this range for N—or more generally, the admissible values of any invariant statistic—is yet another component that a rigorous theoretical formulation of DP must make explicit.)

Invoking the analogy to conditional distributions, one may view the distinction between the multiverse and a universe as analogous to the difference between the conditional probability $\mathsf{P}(\cdot \mid Y)$, conditioning on a random variable Y, and the conditional probability $\mathsf{P}(\cdot \mid Y = y)$, conditioning on the event $\{Y = y\}$. The former is a collection of probability distributions as Y takes on different values, and the latter is a single probability distribution determined by the particular value y. Here the random variable Y is the essential quality, and the event $\{Y = y\}$ is an accidental realization. Similarly, a DP specification should concern the essential nature of a data release mechanism, rather than its properties within some particular, but accidental, universe.

Yet even setting aside the subtleties of properly accounting for invariants, the plethora of existing DP formulations differ along several other dimensions (Desfontaines & Pejó, 2020). As such, DP can vary widely in form and spirit (Dwork et al., 2019), making it difficult to (1) understand what it means for an SDC method to be DP and (2) objectively compare different DP deployments in a systematic way—two tasks central to the goals of this article.

Therefore, to properly integrate invariants into DP, we must first be transparent about which formulation of DP the invariants will be incorporated into. This need ultimately led us to articulate a system of DP specifications, which is presented in the next section. The phrase "a refreshment stirred, not shaken" in our article title is intended to emphasize that this system is not new, but simply a synthesis of the existing literature on the many variations of DP. Indeed, this system is in essence the formalization of three principles that we believe are widely accepted in the DP community, as discussed below.

2.2. **Invariants in Our System of Differential Privacy Specifications.** The first principle states that a DP formulation is a technical standard that requires the rate of change—or 'derivative'—of an SDC method to be controlled (hence the epithet 'differential'). The second principle asserts that the rate of change of an SDC method is defined as the change in the *probability distribution* of the method's output per unit change in its input data. And the third principle observes that different DP formulations correspond to different choices for how and where to measure these changes, as well as how much to control the associated rate of change.

We call these choices the *building blocks* of DP because each of them formalizes a different dimension of DP and all of them are required to fully define a DP formulation. Since they are essential for establishing our theoretical results concerning the PSA's and the TDA's DP guarantees, we will describe the five building blocks in two ways—mathematically and intuitively:

- The *domain* $\mathcal{X}$: a set of data sets $\boldsymbol{x}$
  — *Who* is eligible for protection?
- The *multiverse* $\mathscr{D} \subset 2^{\mathcal{X}}$: a set of universes $\mathcal{D} \subset \mathcal{X}$
  — *Where* does the protection extend to?
- The *input premetric* $d_{\mathcal{X}}$: a 'distance' between any two data sets $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$
  — *What* is the granularity of protection?

- The *output premetric* D_{Pr}: a 'distance' between probability distributions
  — *How* are changes in output variations measured?
- The *protection loss budget* (PLB) $\varepsilon_{\mathcal{D}}$: a function $\mathscr{D} \to [0, \infty]$
  — *How much* protection is afforded?

We term a collection of choices for all five of the building blocks a *DP specification*, while we call a collection of choices for the first four a *DP flavor*. The last building block, the PLB, is more commonly known as the 'privacy' loss budget, although we eschew this term to avoid implying that the PLB fully captures the complex concept of privacy (Benthall & Cummings, 2024; Nissenbaum, 2010; Seeman & Susser, 2024). There is a subscript $\mathcal{D}$ in $\varepsilon_{\mathcal{D}}$ because adopting the same DP flavor for different universes does not mean that we must maintain the same PLB across these universes (for example, the USCB could decide that more protection is required for small values of N than large values). Consequently, we allow the value the PLB takes to vary across universes. The subscript $_{\mathcal{D}}$ therefore indicates the value of the PLB for the particular universe $\mathcal{D}$. (However, we may drop this subscript in some places for reasons explained later in Remarks 1 and 2.) An extensive treatment of these notions and their nuances, implications, and applications is presented in Bailie et al. (2026b). Here, we only focus on aspects essential for stating and proving our main results regarding the PSA and the TDA.

Through our system of DP specifications, we can unify many (though not all) DP formulations in the literature via the following formalism.

Definition 1. A *data release mechanism* (or, synonymously, an SDC method) T is a function that takes as input a (confidential) data set $\boldsymbol{x} \in \mathcal{X}$ and a random seed $U \in \mathcal{U}$, and outputs some noisy statistics $T(\boldsymbol{x}, U)$ based on $\boldsymbol{x}$.

A mechanism T *satisfies the DP specification* $\varepsilon_{\mathcal{D}}$-DP$(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, D_{\mathrm{Pr}})$ if, for all universes $\mathcal{D} \in \mathscr{D}$ and all pairs of data sets $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}$,

$$D_{\mathrm{Pr}}\Big[\mathsf{P}\big(T(\boldsymbol{x}, U) \in \cdot\,\big), \mathsf{P}\big(T(\boldsymbol{x}', U) \in \cdot\,\big)\Big] \leq \varepsilon_{\mathcal{D}} d_{\mathcal{X}}\big(\boldsymbol{x}, \boldsymbol{x}'\big), \tag{2.1}$$

where P is the probability distribution induced by the random seed $U \in \mathcal{U}$, taking the input data set ($\boldsymbol{x}$ or $\boldsymbol{x}'$) as fixed.

Under this system, there are two ways that invariants can naturally be integrated with DP. First, one can set $d_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}') = \infty$ whenever $\boldsymbol{x}$ and $\boldsymbol{x}'$ disagree on the invariants. Second, invariants may be encoded through the multiverse $\mathscr{D}$. We prefer the second approach for its better interpretability. The input premetric $d_{\mathcal{X}}$ is already used to specify the *units* (for example, people, households, or businesses) to which a DP specification provides protection. The units of a DP specification are those entities whose records differ between data sets $\boldsymbol{x}$ and $\boldsymbol{x}'$ with $d_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}') = 1$ (see Bailie et al., 2026b). Overloading $d_{\mathcal{X}}$ to also encode the invariants breaks the connection between a DP specification's units and its input premetric and leads to confusion when both the units and the invariants are nontrivial—for example, as may occur in survey statistics (Bailie & Drechsler, 2024).

Nevertheless, these two approaches—incorporating invariants via $d_{\mathcal{X}}$ or via $\mathscr{D}$—have the same result: Both ensure that the DP condition (Equation 2.1) only bounds the 'distance' between the distributions $\mathsf{P}\big(T(\boldsymbol{x}, U) \in \cdot\big)$ and $\mathsf{P}\big(T(\boldsymbol{x}', U) \in \cdot\big)$ when the two counterfactual data sets $\boldsymbol{x}$ and $\boldsymbol{x}'$ agree on the invariants (see Section 2.3). Intuitively, this corresponds to conditioning on the invariants, except that a priori we do not know the realized value of the invariants. As discussed in Section 2.1, the DP specification must therefore account for all possible values through the

multiverse $\mathscr{D}$, rather than conditioning on any single particular value of the invariants (which would correspond to using a single universe $\mathcal{D}$).

We show in Section 2.4 that weakening the DP condition via a nonvacuous multiverse $\mathscr{D}$—as happens whenever there are invariants—leads to a reduction in the actual protection guaranteed by a DP specification. (By 'nonvacuous multiverse,' we mean one that is not equivalent to the multiverse $\mathscr{D} = \{\mathcal{X}\}$.) However, this complication is necessary in many real-world applications of DP. In addition to examples from the literature outlined in Bailie et al. (2026b), we prove in Section 4 that a nonvacuous multiverse is required to describe the DP protection provided to the 2020 U.S. Census. Furthermore, the practice of empirically restricting the data universe is typical in statistical disclosure control and in data analysis more broadly. Top-coding—setting a maximum limit on a continuous variable, usually after looking at the raw data—is one common example.

Before we proceed, we pause to address the astute reader who may wonder why we do not adopt a third, seemingly simpler approach to incorporating invariants into DP: Take an existing DP formulation and restrict its neighboring data sets to those that agree on the invariants. (For readers who are unfamiliar with the formulation of DP in terms of neighboring data sets, $\boldsymbol{x}$ and $\boldsymbol{x}'$ are called neighbors if $d_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}') = 1$; DP is then defined as the requirement that $D_{\mathrm{Pr}}[\mathsf{P}(T(\boldsymbol{x}, U) \in \cdot\,), \mathsf{P}(T(\boldsymbol{x}', U) \in \cdot\,)] \leq \varepsilon_{\mathcal{D}}$ holds for all neighbors $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}$.) The problem with this approach is that, after excluding neighboring data sets that have different values for the invariants, there may be no neighbors left, resulting in a completely vacuous formulation of DP. Indeed, we will see that this is what happens with data swapping. This is one reason to replace the concept of neighboring data sets with the more general notion of an input premetric, in addition to being the reason not to encode invariants through neighboring data sets.

2.3. **How the Multiverse Accommodates Invariants.** As we illustrated in Section 2.1, the multiverse $\mathscr{D}$ of a DP flavor $(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, D_{\mathrm{Pr}})$ is a collection of subsets of the domain $\mathcal{X}$, and these subsets are called the universes of the DP flavor. The idea behind the concept of a universe $\mathcal{D}$ is that any two data sets $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}$ should be 'mutually plausible,' in the sense that an attacker, upon observing the released statistics, should not be able to determine (with certainty and correctly) whether the true confidential data set is $\boldsymbol{x}$ or whether it is $\boldsymbol{x}'$. (Here, and in the next paragraph, we assume for simplicity that $d_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}')$ and $d_{\mathcal{X}}(\boldsymbol{x}', \boldsymbol{x})$ are both finite for every $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}$.)

In practice, it is often the case that the multiverse $\mathscr{D}$ is defined via a set-valued function that we call the *universe function* $\mathcal{D}(\cdot) : \mathcal{X} \to 2^{\mathcal{X}}$, that associates every potential data set $\boldsymbol{x} \in \mathcal{X}$ with a universe $\mathcal{D}(\boldsymbol{x}) \subset \mathcal{X}$. In this case, the multiverse $\mathscr{D} = \{\mathcal{D}(\boldsymbol{x})\}_{\boldsymbol{x} \in \mathcal{X}}$ is the image of this universe function $\mathcal{D}$. The resulting DP specification would then ensure that every $\boldsymbol{x}' \in \mathcal{D}(\boldsymbol{x})$ is plausible if $\boldsymbol{x}$ is also plausible—that is, an attacker would not be able to distinguish with certainty the true confidential data set to be $\boldsymbol{x}'$ and not $\boldsymbol{x}$, for any $\boldsymbol{x}' \in \mathcal{D}(\boldsymbol{x})$.

A set of invariants can be encoded by a universe function of the form:

$$\mathcal{D}_{\boldsymbol{c}}(\boldsymbol{x}) = \Big\{\boldsymbol{x}' \in \mathcal{X} : \boldsymbol{c}(\boldsymbol{x}') = \boldsymbol{c}(\boldsymbol{x})\Big\}, \tag{2.2}$$

for a given deterministic function $\boldsymbol{c} : \mathcal{X} \to \mathbb{R}^k$. Here the function $\boldsymbol{c}$ describes the features $\boldsymbol{c}(\boldsymbol{x})$ of the data set $\boldsymbol{x}$, which are taken to be invariant. For example, $\boldsymbol{c}(\boldsymbol{x})$ could be the state population totals calculated from the data set $\boldsymbol{x}$ of census responses. We call $\mathcal{D}_{\boldsymbol{c}}(\cdot)$ the *invariant-induced universe function* and its image $\{\mathcal{D}_{\boldsymbol{c}}(\boldsymbol{x})\}_{\boldsymbol{x} \in \mathcal{X}}$ the *invariant-induced multiverse* $\mathscr{D}_{\boldsymbol{c}}$.

By construction, the invariants take the same values across every data set in a universe $\mathcal{D}_{\boldsymbol{c}}(\boldsymbol{x})$. As such, invariants give rise to an equivalence relation $\sim$ over the domain $\mathcal{X}$, which is defined by

$\boldsymbol{x} \sim \boldsymbol{x}'$ if $\boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{c}(\boldsymbol{x}')$. The data universe function therefore induces a *partition* of $\mathcal{X}$ indexed by the image of the invariant function $\boldsymbol{c}$, splitting $\mathcal{X}$ into universes of mutually plausible data sets that share the same values for the invariants.

Example 1. Let the data set be an contingency table of $m \times n$ records taking non-negative integer values: $\mathcal{X} = \mathbb{N}^{m \times n}$. Suppose the function $\boldsymbol{c} : \mathbb{N}^{m \times n} \to \mathbb{N}^{m+n}$ tabulates the row- and column-margins:

$$\boldsymbol{c}(\boldsymbol{x}) = \left(\sum_{i=1}^{m} x_{i1}, \ldots, \sum_{i=1}^{m} x_{in}, \sum_{j=1}^{n} x_{1j}, \ldots, \sum_{j=1}^{n} x_{mj} \right). \tag{2.3}$$

The data curator may treat the row and column margins of the confidential data set as invariant (see Dobra & Fienberg, 2000, and references therein). This would be equivalent to employing the universe function $\mathcal{D}_{\boldsymbol{c}}(\cdot)$ as defined in Equation 2.2 using the function $\boldsymbol{c}$ from Equation 2.3, since $\mathcal{D}_{\boldsymbol{c}}(\cdot)$ ensures that only pairs of data sets $\boldsymbol{x}, \boldsymbol{x}'$ with the same row and column margins are subject to the DP condition (Equation 2.1).

More generally, the DP condition is required only for data sets that belong to the same universe $\mathcal{D} \in \mathscr{D}$. In the case of an invariant-induced multiverse $\mathscr{D}_{\boldsymbol{c}}$, this means the DP condition will only be applied to data sets $\boldsymbol{x}$ and $\boldsymbol{x}'$, which share the same values for the invariants. As we will prove in Section 2.4, this allows the invariants to be released as is, without contributing to the PLB.

In some applications, there are also inequality invariants (Abowd et al., 2022). As an example of such an invariant, the 2020 U.S. Decennial Census requires that the reported number of group quarters in any geographical unit is no larger than the number of persons in that unit. More generally, an inequality invariant is of the form $f(\boldsymbol{x}) \leq 0$ for some function $f : \mathcal{X} \to \mathbb{R}$. Such an invariant can be incorporated in the above framework by defining

$$c(\boldsymbol{x}) = \begin{cases} 1 & \text{if } f(\boldsymbol{x}) \leq 0, \\ 0 & \text{if } f(\boldsymbol{x}) > 0. \end{cases} \tag{2.4}$$

2.4. **Understanding the Impact of Invariants on Differential Privacy.** As we have repeatedly emphasized, invariants reduce the SDC protection provided by a DP specification because they restrict where the DP condition (Equation 2.1) must hold. This does not mean that invariants are antithetical to the fundamental idea of DP. To assert otherwise would impose an unduly restrictive interpretation of DP, one that would not only greatly reduce its applicability, but would also categorize the original formulation of DP as not satisfying DP. Indeed, this formulation (given in Dwork et al., 2006) uses the Hamming distance as its input premetric, as does any formulation of *bounded* DP more generally. Because the Hamming distance between two data sets is infinite whenever they have a different number of records, these formulations only require the DP condition to hold for data sets of the same size. Thus, having the data set size as an invariant has been a part of DP from its conception.

It should be clear then that the mere presence of invariants is not necessarily an issue; what matters is the extent that they impact actual SDC protection. To this end, this section provides some preliminary, theoretical results on invariants' effect. (A broader discussion on their impact is provided in Section 6.1.) Since releasing the invariants without modification was the aim of restricting the DP condition in the first place, we begin by demonstrating the necessity and sufficiency of the invariant-induced multiverse to achieving this aim. (All results in this section are proved in Appendix D.)

Proposition 1. *Fix a domain $\mathcal{X}$ and some invariants $\boldsymbol{c} : \mathcal{X} \to \mathbb{R}^k$. For any $d_{\mathcal{X}}$ and D_{Pr}, the mechanism $T(\boldsymbol{x}) = \boldsymbol{c}(\boldsymbol{x})$ that releases the invariants exactly satisfies $\varepsilon_{\mathcal{D}}$-DP$(\mathcal{X}, \mathscr{D}_{\boldsymbol{c}}, d_{\mathcal{X}}, D_{\mathrm{Pr}})$ with PLB $\varepsilon_{\mathcal{D}} = 0$ for all $\mathcal{D} \in \mathscr{D}_{\boldsymbol{c}}$.*

Now suppose $D_{\mathrm{Pr}}(\mathsf{P}, \mathsf{Q}) = \infty$ whenever the total variation distance between P and Q is one. (This assumption is satisfied by most common choices of D_{Pr}.) Let $\mathscr{D}$ be a multiverse such that there exists some universe $\mathcal{D}_0 \in \mathscr{D}$ and some $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}_0$ with $d_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}') < \infty$ and $\boldsymbol{c}(\boldsymbol{x}) \neq \boldsymbol{c}(\boldsymbol{x}')$. Then T does not satisfy $\varepsilon_{\mathcal{D}}$-DP$(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, D_{\mathrm{Pr}})$ whenever $\varepsilon_{\mathcal{D}_0} < \infty$.

The first part of Proposition 1 shows that the invariants can be released without modification 'for free'—that is, at no cost to the PLB. Hence, the invariant-induced multiverse is sufficient for achieving the aim of accommodating invariants. This result also holds if $\mathscr{D}$ is any multiverse with $\boldsymbol{c}$ constant within every universe $\mathcal{D} \in \mathscr{D}$ (i.e., if $\boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{c}(\boldsymbol{x}')$ for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}$ and all $\mathcal{D} \in \mathscr{D}$).

The second part of Proposition 1 demonstrates the necessity of the invariant-induced multiverse. It shows that a set of invariants cannot be published as is by a DP mechanism whenever $\mathscr{D}$ does not conform to these invariants. More precisely, if there are data sets $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}_0$, which take different values on the invariants (and satisfy $d_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}') < \infty$), then releasing the invariants exactly would require $\varepsilon_{\mathcal{D}_0} = \infty$.

The following result demonstrates that the invariant-induced multiverse does not allow for the release of additional information, above and beyond the information contained in the invariants, without incurring some loss of protection. This is important because while we want to allow the invariants to be released as is, we do not want this to imply that other information can also be released for free.

Proposition 2. *Suppose that a data release mechanism T varies within some universe $\mathcal{D}_0 \in \mathscr{D}$ in the sense that there exists $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}_0$ with $d_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}') < \infty$ but $\mathsf{P}(T(\boldsymbol{x}, U) \in \cdot\ ) \neq \mathsf{P}(T(\boldsymbol{x}', U) \in \cdot\ )$. When D_{Pr} is a metric, T satisfies $\varepsilon_{\mathcal{D}}$-DP$(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, D_{\mathrm{Pr}})$ only if $\varepsilon_{\mathcal{D}_0} > 0$.*

A mechanism T varies within a universe $\mathcal{D}_0 \in \mathscr{D}_{\boldsymbol{c}}$ whenever it includes information that is not logically equivalent to the invariants $\boldsymbol{c}$. In this case, Proposition 2 shows that releasing this information will require a nonzero PLB, even while the invariants can be released for free.

Beyond describing exactly what information can be released under an invariant-induced multiverse, the above two propositions more generally illustrate that the interpretation of the PLB cannot be isolated from the multiverse, and indeed this complicates the comparison of budgets across different DP flavors.

As a concrete example of how the meaning of the PLB $\varepsilon_{\mathcal{D}}$ changes with $\mathscr{D}$, consider evaluating the same mechanism T against two DP flavors $(\mathcal{X}, \mathscr{D}_{\boldsymbol{c}}, d_{\mathcal{X}}, D_{\mathrm{Pr}})$ and $(\mathcal{X}, \mathscr{D}_{\boldsymbol{c}'}, d_{\mathcal{X}}, D_{\mathrm{Pr}})$, which differ only on their invariants. Suppose the second set of invariants are nested within the first; that is, $\boldsymbol{c}$ is strictly more constraining than $\boldsymbol{c}'$. (For example, $\boldsymbol{c}$ are population counts at the block level and $\boldsymbol{c}'$ are counts at the county level.) Then Proposition 3 below proves that T's protection loss under $\mathscr{D}_{\boldsymbol{c}'}$ cannot be smaller and may be strictly larger than its loss under $\mathscr{D}_{\boldsymbol{c}}$.

By comparing T's two PLBs on their own, a reader might arrive at the seemingly paradoxical conclusion: we can increase the protection provided by T simply by increasing the number of invariants allowed by our DP flavor. Yet this ignores the critical fact that the PLB is only a 'within-system' evaluation of SDC. It is not an absolute, unitless measure of protection, but only a relative measure whose units are determined by the DP flavor. (See Bailie et al., 2026a, for an explanation of how the units of the PLB depend on the other three components of the DP flavor, not just $\mathscr{D}$.) It is dangerous therefore to think that the $\boldsymbol{c}$-release is actually afforded with less SDC protection

than the $\boldsymbol{c}'$-release because there is privacy leakage due to specifying additional invariants, which is not captured by the within-system evaluation $\varepsilon_{\mathcal{D}}$. Indeed in the extreme example where $\boldsymbol{c}$ is an injective function so that the universes $\mathcal{D}$ are singletons, the entire data set can be published, and hence there is no actual protection afforded by the DP specification $\varepsilon_{\mathcal{D}}$-DP$(\mathcal{X}, \mathscr{D}_{\boldsymbol{c}}, d_{\mathcal{X}}, D_{\mathrm{Pr}})$ even though in such cases $\varepsilon_{\mathcal{D}}$ can be set to zero. This point is crucial to understanding the comparative analysis between the PSA and the 2020 Census data releases as presented in Section 5.

Proposition 3. *Suppose that $\mathscr{D}$ is refinement of $\mathscr{D}'$—that is, for all $\mathcal{D} \in \mathscr{D}$, there exists some $\mathcal{D}' \in \mathscr{D}'$ such that $\mathcal{D} \subset \mathcal{D}'$. Define the protection loss of a data release mechanism T under the flavor $(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, D_{\mathrm{Pr}})$ as $L_{\mathcal{D}} = \inf\{\varepsilon_{\mathcal{D}} : T$ satisfies $\varepsilon_{\mathcal{D}}$-$DP(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, D_{\mathrm{Pr}})\}$. Similarly define $L'_{\mathcal{D}'} = \inf\{\varepsilon'_{\mathcal{D}'} : T$ satisfies $\varepsilon'_{\mathcal{D}'}$-$DP(\mathcal{X}, \mathscr{D}', d_{\mathcal{X}}, D_{\mathrm{Pr}})\}$. Then, the protection loss $L'_{\mathcal{D}'}$ under $(\mathcal{X}, \mathscr{D}', d_{\mathcal{X}}, D_{\mathrm{Pr}})$ is no smaller than the loss $L_{\mathcal{D}}$ under $(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, D_{\mathrm{Pr}})$:*

$$L'_{\mathcal{D}'} \geq L_{\mathcal{D}},$$

for any $\mathcal{D} \in \mathscr{D}$ and $\mathcal{D}' \in \mathscr{D}'$ with $\mathcal{D} \subset \mathcal{D}'$.

This proposition is an initial tool for comparing DP specifications, which vary on their invariants. The two sets of invariants in this result have to take a very particular form—one has to be nested inside the other. In this case, we can order the two sets of invariants by their impact on SDC protection. Adding invariants reduces the quality of a DP flavor, because it refines the multiverse from $\mathscr{D}'$ to $\mathscr{D}$, and in so doing shrinks the scope of protection and 'waters down' the DP flavor. Moreover, the new, lower quality DP flavor brings about an apparent 'saving' of protection loss—a change from $L'_{\mathcal{D}'}$ to $L_{\mathcal{D}}$. Section 5.3 dissects this phenomenon using a hypothetical swapping analysis of the 2020 Census. As a general matter, comparing the relative impact of non-nested sets of invariants is a much more difficult task because it requires weighting the disclosiveness of different invariants against each other. We suggest ways to approach this more difficult task in Section 6.1.

3. A Differentially Private Analysis of Data Swapping

3.1. **Data Swapping.** Given a data set $\boldsymbol{x}$, partition its set of variables $\boldsymbol{V}$ into two nonempty subsets: the *swapping variables* $\boldsymbol{V}_{\mathrm{Swap}}$ and the *holding variables* $\boldsymbol{V}_{\mathrm{Hold}}$. A data swapping method randomly selects some records of $\boldsymbol{x}$ and interchanges the values of their swapping variables $\boldsymbol{V}_{\mathrm{Swap}}$, while leaving their $\boldsymbol{V}_{\mathrm{Hold}}$ unchanged. This creates a new data set consisting of individual records whose $\boldsymbol{V}_{\mathrm{Hold}}$ values are as originally observed and whose $\boldsymbol{V}_{\mathrm{Swap}}$ values are possibly different. The exact procedure for selecting records and interchanging their $\boldsymbol{V}_{\mathrm{Swap}}$ values varies between different data swapping methods. (References to literature on various swapping methods can be found in Appendix B, which also summarizes the use of data swapping by national statistical offices across the world.)

Sometimes, swapping is restricted to records that share the same values on a subset of the holding variables $\boldsymbol{V}_{\mathrm{Hold}}$, called the *matching variables* $\boldsymbol{V}_{\mathrm{Match}}$. More exactly, whenever $\boldsymbol{V}_{\mathrm{Match}}$ is nonempty, records are partitioned into strata according to their $\boldsymbol{V}_{\mathrm{Match}}$ values and data swapping is repeated independently within each stratum. Also referred to as the *swap key* (Abowd & Hawes, 2023; McKenna, 2018), the matching variables often capture important demographic information that the data custodian would like to preserve. (Because swapping records with different $\boldsymbol{V}_{\mathrm{Match}}$ values is prohibited, this information is indeed preserved by data swapping.)

Example 2. This example is a simplification of the DAS for the 2010 U.S. Decennial Census. Represent the 2010 Census data as a list of household records, whose variables include all the household's characteristics, as well as the questionnaire responses from each individual associated with that household. The matching variables $\boldsymbol{V}_{\text{Match}}$ (i.e., swap key) include both the number of voting age persons and the total number of persons in the household. $\boldsymbol{V}_{\text{Match}}$ also includes a geographic variable V_g (see U.S. Census Bureau, 2021d), either the census tract, county, or state of the household. (To the best of our knowledge, the exact choice of V_g has never been made public by the USCB.) $\boldsymbol{V}_{\text{Swap}}$ are the geographic variables nested underneath V_g. For example if V_g is the county, then $\boldsymbol{V}_{\text{Swap}}$ is the block and tract of the household. All other variables belong to $\boldsymbol{V}_{\text{Hold}}$—in particular, the household and person characteristics. One can imagine the 2010 DAS as digging up pairs of houses of the same size in the same geographic area and swapping their locations but not changing the houses and their occupants. In the 2010 DAS, each household is assigned a risk score based on the USCB's assessment of how unique the household is within its neighborhood. These risk scores are used to compute each household's probability of being swapped. Every (non-imputed) household has a nonzero swap probability. Selected households are then swapped with one of their neighbors. (See Appendix C for a detailed description of the 2010 DAS and references for this information.)

3.2. **What Invariants Does Swapping Preserve?** Swapping is, very loosely, a synthetic data generation mechanism. Given a data set $\boldsymbol{x}$ as input, swapping produces a 'privacy enhanced' version $\boldsymbol{Z}$ of $\boldsymbol{x}$. Both $\boldsymbol{x}$ and $\boldsymbol{Z}$ contain the same variables as well as the same number of records. Hence, the invariants of swapping are determined by examining what swapping does, and does not, change in the data.

Consider the data set $\boldsymbol{x}$ as a matrix whose rows correspond to the records of $\boldsymbol{x}$ and whose columns correspond to the variables $\boldsymbol{V}$ of $\boldsymbol{x}$. Without loss of generality, the holding variables are ordered before the swapping variables so that $\boldsymbol{x}$ can be partitioned as $[\boldsymbol{x}_{\text{Hold}}, \boldsymbol{x}_{\text{Swap}}]$. A swapping algorithm randomly selects a permutation σ of the rows of $\boldsymbol{x}$ and interchanges the rows of the matrix $\boldsymbol{x}_{\text{Swap}}$ according to σ. This operation yields $\boldsymbol{x}^{\sigma}_{\text{Swap}}$, whose ith row is given by the $\sigma(i)$-th row of $\boldsymbol{x}_{\text{Swap}}$. This defines the swapped data set $\boldsymbol{Z}$ as the matrix $[\boldsymbol{x}_{\text{Hold}}, \boldsymbol{x}^{\sigma}_{\text{Swap}}]$, and the swapping mechanism releases as its output the fully saturated contingency table generated by $\boldsymbol{Z}$.

One can see that after swapping, any statistic generated by only the matrix $\boldsymbol{x}_{\text{Hold}}$ is invariant. Moreover, since $\boldsymbol{V}_{\text{Match}}$ is identical among swapped records, any statistic generated by only $\boldsymbol{x}_{\text{Match}}$ and $\boldsymbol{x}_{\text{Swap}}$ is also preserved by swapping. Only statistics that depend nontrivially on both variables $\boldsymbol{V}_{\text{Swap}}$ and $\boldsymbol{V}_{\text{Hold}} \setminus \boldsymbol{V}_{\text{Match}}$ can be altered by swapping.

Proposition 4. *Suppose that $\boldsymbol{V}_{\text{Hold}} \setminus \boldsymbol{V}_{\text{Match}}$ and $\boldsymbol{V}_{\text{Swap}}$ are nonempty. Then, without loss of generality, we may assume that each of $\boldsymbol{V}_{\text{Match}}, \boldsymbol{V}_{\text{Swap}}$, and $\boldsymbol{V}_{\text{Hold}} \setminus \boldsymbol{V}_{\text{Match}}$ are univariate. Denote a value of the matching variable $\boldsymbol{V}_{\text{Match}}$ by m. Similarly, let h and s be values of $\boldsymbol{V}_{\text{Hold}} \setminus \boldsymbol{V}_{\text{Match}}$ and $\boldsymbol{V}_{\text{Swap}}$ respectively.*

Disregarding the ordering of records, the data set $\boldsymbol{x}$ can be represented as a 3-dimensional contingency table $H(\boldsymbol{x}) = [n^{\boldsymbol{x}}_{mhs}]$ of counts in each combination of possible values for m, h, and s. (We will omit the superscript $^{\boldsymbol{x}}$ when it is clear from the context.) In general, interior cell counts n_{mhs} are not preserved under swapping and neither are the margins $n_{\cdot hs} = \sum_m n_{mhs}$. But swapping does keep $n_{m\cdot s} = \sum_h n_{mhs}$ and $n_{mh\cdot} = \sum_s n_{mhs}$ invariant.

In other words, there are two contingency tables that remain unchanged by swapping: 1) $\boldsymbol{V}_{\text{Match}} \times \boldsymbol{V}_{\text{Swap}}$: the cross-classification of the matching variables by the swapping variables; and 2) $\boldsymbol{V}_{\text{Hold}}$:

the cross-classification of all the holding variables; while the interior of the contingency table, $(\boldsymbol{V}_{\text{Hold}} \setminus \boldsymbol{V}_{\text{Match}}) \times \boldsymbol{V}_{\text{Swap}}$ is perturbed by swapping. (Here, as elsewhere in this work, "$\times$" and "$\setminus$" are respectively the Cartesian product and set difference operators.)

Proof. First we justify why we can assume that $\boldsymbol{V}_{\text{Match}}$, $\boldsymbol{V}_{\text{Swap}}$ and $\boldsymbol{V}_{\text{Hold}} \setminus \boldsymbol{V}_{\text{Match}}$ are univariate (i.e., that these variable sets are singletons). If $\boldsymbol{V}_{\text{Match}}$ is empty, replace it with a set consisting of a new variable taking the same value on every record. And if either of $\boldsymbol{V}_{\text{Match}}$, $\boldsymbol{V}_{\text{Swap}}$ or $\boldsymbol{V}_{\text{Hold}} \setminus \boldsymbol{V}_{\text{Match}}$ has more than one variable, then cross-classify these variables into a single variable. Neither of these two operations will change the behavior of a swapping method, so we may use them to ensure $\boldsymbol{V}_{\text{Match}}, \boldsymbol{V}_{\text{Swap}}$ and $\boldsymbol{V}_{\text{Hold}} \setminus \boldsymbol{V}_{\text{Match}}$ are univariate.

Since every permutation σ can be written as the composition of swaps (i.e., two-cycles), it suffices to show that all possible swaps preserve $n_{m\cdot s}$ and $n_{mh\cdot}$ but not necessarily $n_{\cdot hs}$. A swap pairs a record a in categories mhs with a record b in $mh's'$. It moves a to $mh's$ and b to mhs'. The matching category m is the same in a and b by construction. Unless $m = m'$ or $s = s'$, after the swap n_{mhs} and $n_{mh's'}$ decrease by one, and $n_{mh's}$ and $n_{mhs'}$ increase by one. Hence, $n_{m\cdot s}$ and $n_{mh\cdot}$ remain unchanged but $n_{\cdot hs}$ changes whenever $h \neq h'$ and $s \neq s'$. □

Example 2 (continued). In the 2010 U.S. Census DAS, the number of adults, children, and households in each block are invariant. (This is the $n_{m\cdot s}$ margin.) The counts of all the person and household characteristics inside each V_g are also invariant. (This is the $n_{mh\cdot}$ margin.) For example, if V_g is the county, then the aggregate characteristics at the county level remain unchanged by swapping, but these aggregates at the block and tract level are perturbed.

Definition 2. Under the setup of Proposition 4, define the *swapping invariants* $\boldsymbol{c}_{\text{Swap}}(\boldsymbol{x})$ for a given choice of $\boldsymbol{V}_{\text{Match}}, \boldsymbol{V}_{\text{Swap}}$, and $\boldsymbol{V}_{\text{Hold}}$ as the vector of all margins $n_{mh\cdot}$ and $n_{m\cdot s}$, for all possible values of m, h and s. For example, if $\boldsymbol{V}_{\text{Match}}$, $\boldsymbol{V}_{\text{Hold}} \setminus \boldsymbol{V}_{\text{Match}}$, and $\boldsymbol{V}_{\text{Swap}}$ take values in $\{1, \ldots, \mathcal{M}\}$, $\{1, \ldots, \mathcal{H}\}$, and $\{1, \ldots, \mathcal{S}\}$ respectively, then

$$\boldsymbol{c}_{\text{Swap}}(\boldsymbol{x}) = \begin{bmatrix} n_{11\cdot} & n_{12\cdot} & \cdots & n_{\mathcal{MH}\cdot} & n_{1\cdot 1} & n_{1\cdot 2} & \cdots & n_{\mathcal{M}\cdot\mathcal{S}} \end{bmatrix}^{\mathsf{T}}.$$

As the following example illustrates, we do not have complete flexibility in choosing the invariants of swapping.

Example 3. In the 2020 TDA, there are three invariants: 1) the number of people in each state; 2) the number of housing units in each block; and 3) the count of each type of occupied group quarters (e.g., residence halls, nursing facilities, prisons) in each block (U.S. Census Bureau, 2021e). We cannot design a swapping algorithm that preserves these—and only these—invariants. In other words, the 2020 U.S. Census invariants do not correspond to any swapping invariants $\boldsymbol{c}_{\text{Swap}}$, regardless of the choice of $\boldsymbol{V}_{\text{Match}}, \boldsymbol{V}_{\text{Swap}}$ and $\boldsymbol{V}_{\text{Hold}}$. Why? Swapping always preserves the one-dimensional marginals: $n_{m\cdot\cdot}, n_{\cdot h\cdot}$ and $n_{\cdot\cdot s}$; but the 2020 DAS does not. For example, the number of 25- to 34-year-old people in the United States is not invariant under the 2020 TDA, but it must necessarily be invariant under any swapping algorithm.

3.3. Permutation Swapping Satisfies Pure Differential Privacy Subject to Its Invariants.

In this section, we design a specific data swapping algorithm—called the *Permutation Swapping Algorithm* (PSA) to distinguish it from other data swapping methods—which satisfies the DP flavor $(\mathcal{X}, \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}, d^r_{\text{HamS}}, D_{\text{Mult}})$. Here $\mathcal{X}$ denotes any set of data sets that all have the same common set of variables, and d^r_{HamS} denotes the Hamming distance at the resolution r of the PSA's swapping

procedure. That is, $d^r_{\text{HamS}}(\boldsymbol{x}, \boldsymbol{x}')$ denotes the number of records that differ between data sets $\boldsymbol{x}$ and $\boldsymbol{x}'$ (ignoring the ordering of the records in $\boldsymbol{x}$ and $\boldsymbol{x}'$):

$$d^r_{\text{HamS}}(\boldsymbol{x}, \boldsymbol{x}') = \begin{cases} \frac{1}{2}|\boldsymbol{x} \ominus \boldsymbol{x}'| & \text{if } |\boldsymbol{x}| = |\boldsymbol{x}'|, \\ \infty & \text{otherwise,} \end{cases} \tag{3.1}$$

where $\ominus$ is the symmetric multiset difference and the resolution r is the type of record in consideration. For example, if the PSA swapped records that correspond to individual persons, then the input premetric of the PSA's DP flavor would be the Hamming distance d^p_{HamS} on person-records. Alternatively, the PSA could swap household-records, in which case its input premetric would be the Hamming distance d^{hh}_{HamS} at the resolution of households. (This distinction will become important when we compare the PSA with the TDA.) The output premetric of the PSA's DP flavor is the multiplicative distance, which corresponds to the notion of pure DP (Dwork et al., 2006) and is defined as:

$$D_{\text{Mult}}(\mathsf{P}, \mathsf{Q}) = \sup_{E \in \mathscr{F}} \left| \ln \frac{\mathsf{P}(E)}{\mathsf{Q}(E)} \right|, \tag{3.2}$$

for two probabilities P and Q on the measurable space $\mathcal{T}$ with σ-algebra $\mathscr{F}$.

While the PSA was not used in 2010, a specific instantiation of it does reflect the essential features of the 2010 DAS's data swapping algorithm (Section 3.5). However, certain aspects of the PSA were made with the specific goal of satisfying DP. For example, a swapping method cannot satisfy $(\mathcal{X}, \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}, d^r_{\text{HamS}}, D_{\text{Mult}})$ if the number of swaps it makes is fixed. (To be clear, based on the available public information, we do not believe the 2010 DAS fixes the number of swaps, although it does appear to control this number to some degree.) To see this, suppose that a possible output data set $\boldsymbol{z}$ differs from $\boldsymbol{x} \in \mathcal{D}_0$ by m swaps and from $\boldsymbol{x}' \in \mathcal{D}_0$ by $m+1$ swaps. If the swapping methods allows a maximum of m swaps, then $\boldsymbol{z}$ has nonzero probability given $\boldsymbol{x}$ as input but zero probability given $\boldsymbol{x}'$, thereby violating $\varepsilon_{\mathcal{D}}$-DP$(\mathcal{X}, \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}, d^r_{\text{HamS}}, D_{\text{Mult}})$ for any finite $\varepsilon_{\mathcal{D}_0}$. More generally, a necessary condition for a swapping method to satisfy $\varepsilon_{\mathcal{D}}$-DP$(\mathcal{X}, \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}, d^r_{\text{HamS}}, D_{\text{Mult}})$ for finite $\varepsilon_{\mathcal{D}_0}$ is that, given input $\boldsymbol{x} \in \mathcal{D}_0$, any data set $\boldsymbol{x}' \in \mathcal{D}_0$ has a nonzero probability of being outputted (up to reordering of the rows of $\boldsymbol{x}'$).

To ensure this condition, rather than swapping rows of $\boldsymbol{x}_{\text{Swap}}$ in the same matching category m, the PSA instead randomly permutes these rows, a type of data swapping method introduced in DePersio et al. (2012). Since we do not want to permute every row of $\boldsymbol{x}_{\text{Swap}}$, rows are randomly selected, independently with probability p, and only these selected rows are shuffled. Or, more accurately, after selecting rows of $\boldsymbol{x}_{\text{Swap}}$ with matching value m, the PSA samples uniformly at random a permutation $\sigma_m : \{1, \ldots, n_{m\cdot\cdot}\} \to \{1, \ldots, n_{m\cdot\cdot}\}$, which fixes nonselected rows (i.e., $\sigma_m(i) = i$ for all nonselected i), and deranges selected rows (i.e., $\sigma_m(i) \neq i$ for all selected i). This process is repeated for all values of m so that the final data set, after all permutations have been applied, is given by $\boldsymbol{Z} = [\boldsymbol{x}_{\text{Hold}}, \boldsymbol{x}^{\sigma}_{\text{Swap}}]$, where σ is defined by $\sigma(i) = \sigma_m(i)$ for record i with matching category m. In the case that only one record was selected, there are no possible σ_m and so records are reselected. Hence, the probability that a record with matching category m is swapped is

$$p \sum_{j=1}^{n_{m\cdot\cdot}-1} \binom{n_{m\cdot\cdot}-1}{j} p^j (1-p)^{n_{m\cdot\cdot}-1-j}.$$

When $n_{m\cdot\cdot} \gg 1$, the expected fraction of records that will have their swapping variables interchanged is approximately p. For this reason, we call p the swap rate.

Pseudocode for the PSA is provided in Algorithm 1. The output is a fully saturated contingency table $C(\boldsymbol{Z}) = [n^{\boldsymbol{Z}}_{mhs}]$ (i.e., a 3-way tensor) computed on the swapped data set $\boldsymbol{Z}$. When $\boldsymbol{V}_{\text{Match}}$, $\boldsymbol{V}_{\text{Hold}} \setminus \boldsymbol{V}_{\text{Match}}$, and $\boldsymbol{V}_{\text{Swap}}$ all take a finite number of values, $C(\boldsymbol{Z}) = [n^{\boldsymbol{Z}}_{mhs}]$ is a collection of $\mathcal{M}$ matrices $C_m(\boldsymbol{Z}) = [n^{\boldsymbol{Z}}_{mhs}]$, for $m = 1, \ldots, \mathcal{M}$, each of which has dimension $\mathcal{H} \times \mathcal{S}$. This contingency table $C(\boldsymbol{Z})$ fully determines $\boldsymbol{Z}$ up to reordering of the rows of $\boldsymbol{Z}$.

Theorem 1. *Suppose the domain $\mathcal{X}$ is such that every data set $\boldsymbol{x} \in \mathcal{X}$ shares the same common set $\boldsymbol{V}$ of variables. Partition $\boldsymbol{V}$ into swapping variables $\boldsymbol{V}_{\text{Swap}}$ and holding variables $\boldsymbol{V}_{\text{Hold}}$, and let $\boldsymbol{V}_{\text{Match}} \subset \boldsymbol{V}_{\text{Hold}}$ be the (possibly empty) set of matching variables. Let $\mathcal{B}$ be the set of matching strata m for which there exist at least two distinct records in stratum m. If $\mathcal{B}$ is not empty, define $b = \max_{m \in \mathcal{B}} n_{m\cdot\cdot}$; otherwise define $b = 0$.*

Suppose that the PSA (Algorithm 1) permutes records at resolution r. Then it satisfies $\varepsilon_{\mathcal{D}}$-DP($\mathcal{X}$, $\mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}, d^r_{\text{HamS}}, D_{\text{MULT}}$) with

$$\varepsilon_{\mathcal{D}} = \begin{cases} 0 & \text{if } b = 0, \\ \ln(b+1) - \ln o & \text{if } 0 < p \leq \frac{\sqrt{b+1}}{\sqrt{b+1}+1} \text{ and } b > 0, \\ \ln o & \text{if } \frac{\sqrt{b+1}}{\sqrt{b+1}+1} \leq p < 1 \text{ and } b > 0, \\ \infty & \text{if } p \in \{0, 1\} \text{ and } b > 0, \end{cases} \tag{3.3}$$

where $o = p/(1-p)$.

It is worth noting that the monotonic increase of $\varepsilon_{\mathcal{D}}$ with b may seem counterintuitive, until one realizes that the PLB quantified in Theorem 1 does not include the loss due to the invariants themselves. In other words, the more invariants the PSA imposes—which tends to lead to smaller b—the less information there is left for the PSA to protect, and hence it is easier to achieve smaller $\varepsilon_{\mathcal{D}}$. This phenomenon is not unique to the PSA, but reflects the fundamentally *relative* nature of DP.

A proof of Theorem 1 is presented in Appendix E. Here we give a broad sketch for the case $0 < p \leq 0.5$ and $b > 0$. Because $\sqrt{b+1}/(\sqrt{b+1}+1) > 0.5$, we need to show, for fixed data sets $\boldsymbol{x}, \boldsymbol{x}'$, and $\boldsymbol{z}$ in the same universe $\mathcal{D} \in \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}$, that the budget $\varepsilon_{\mathcal{D}} = \ln(b+1) - \ln o$ satisfies the inequality

$$\mathsf{P}[C([\boldsymbol{x}_{\text{Hold}}, \boldsymbol{x}^{\sigma}_{\text{Swap}}]) = C(\boldsymbol{z})] \leq \exp(k\varepsilon_{\mathcal{D}}) \mathsf{P}[C([\boldsymbol{x}'_{\text{Hold}}, \boldsymbol{x}'^{\sigma'}_{\text{Swap}}]) = C(\boldsymbol{z})], \tag{3.4}$$

where $k = d^r_{\text{HamS}}(\boldsymbol{x}, \boldsymbol{x}')$. The probabilities in Equation 3.4 are over the random sampling of the permutations σ and σ' in Algorithm 1. We can show that there exists a derangement ρ of k records such that $C(\boldsymbol{x}) = C([\boldsymbol{x}'_{\text{Hold}}, \boldsymbol{x}'^{\rho}_{\text{Swap}}])$ (Lemma 5). (A derangement is a permutation that does not fix any rows.) Moreover, there is a bijection between the possible σ and σ' given by $\sigma' = \sigma \circ \rho$. Hence, if k_σ is the number of records deranged by σ, we have

$$k_\sigma - k \leq k_{\sigma'} \leq k_\sigma + k. \tag{3.5}$$

For such pairs of possible σ and σ', the ratio $\mathsf{P}(\sigma)/\mathsf{P}(\sigma')$ can be bounded in terms of $o^{k_\sigma - k_{\sigma'}}$ and the ratio between the number of derangements of size $k_{\sigma'}$ and of size k_σ. For $o \leq 1$, this can in turn be bounded by $(b+1)^k o^{-k}$ using the inequality in Equation 3.5. Hence $\varepsilon_{\mathcal{D}} = \ln(b+1) - \ln o$ does indeed satisfy Equation 3.4.

In Appendix F, we prove that the PLB $\varepsilon_{\mathcal{D}}$ for the PSA given in Theorem 1 is tight in the weak sense that under some mild assumptions the difference between the right and left sides of the inequality in Equation 3.4 is arbitrarily close to zero for some choice of $\boldsymbol{x}$, $\boldsymbol{x}'$, and $\boldsymbol{z}$.

Algorithm 1: The Permutation Swapping Algorithm (PSA)

Input: A data set $\boldsymbol{x} \in \mathcal{X}$ whose set of variables $\boldsymbol{V}$ is partitioned into holding variables $\boldsymbol{V}_{\text{Hold}}$ and swapping variables $\boldsymbol{V}_{\text{Swap}}$; along with a set of matching variables $\boldsymbol{V}_{\text{Match}} \subset \boldsymbol{V}_{\text{Hold}}$ that define the matching strata.

1: Set $\boldsymbol{Z} \leftarrow \boldsymbol{x}$
2: **for all** matching strata m **do**
3: **if** $n_{m\cdot\cdot} = 0$ or $n_{m\cdot\cdot} = 1$ **then**
4: **continue**
5: **end if**
6: **for all** records i in stratum m **do**
7: Select i with probability p
8: **end for**
9: **if** 0 records selected **then**
10: **continue**
11: **else if** exactly 1 record selected **then**
12: Deselect all records
13: **go to** line 6
14: **end if**
15: Sample uniformly at random a permutation σ_m, which fixes the unselected records and deranges the selected records
16: /* *Permute the swapping variables according to* σ_m*:* */
17: $\boldsymbol{Z} \leftarrow [\boldsymbol{Z}_{\text{Hold}}, \boldsymbol{Z}^{\sigma_m}_{\text{Swap}}]$
18: Deselect all records
19: **end for**
20: **return** the fully saturated contingency table $C(\boldsymbol{Z})$

Remark 1. Since $n_{m\cdot\cdot}$ is an invariant, $n^{\boldsymbol{x}}_{m\cdot\cdot} = n^{\boldsymbol{x}'}_{m\cdot\cdot}$ for all $\boldsymbol{x}$ and $\boldsymbol{x}'$ in the same universe. Thus, b is a function of $\mathcal{D}$ and hence so is the PLB $\varepsilon_{\mathcal{D}}$ given in Equation 3.3. In the context of the PSA, we will use ε to denote the value of $\varepsilon_{\mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}(\boldsymbol{x}^*)}$ under the universe $\mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}(\boldsymbol{x}^*)$ corresponding to the realized confidential data $\boldsymbol{x}^*$. We will also report the PSA's protection loss budget in terms of this value ε and omit the values of $\varepsilon_{\mathcal{D}}$ for other universes $\mathcal{D}$. Even though it is a function of the realized data $\boldsymbol{x}^*$, the value of ε can still be publicly reported under the PSA's DP specification without additional protection loss (Bailie et al., 2026b).

3.4. **A Numerical Demonstration: The 1940 Census Full Count Data.** We demonstrate the PSA using the 1940 U.S. Decennial Census full count data, obtained from the IPUMS USA Ancestry Full Count Database (Ruggles et al., 2021). For the 1940 Census, the smallest geography level is county, hence swapping is performed among household units across counties within each state—that is, $\boldsymbol{V}_{\text{Swap}}$ is set to be each household's county indicator. The matching variables (or swap key) $\boldsymbol{V}_{\text{Match}}$ are the number of persons per household and the household's state. Our analysis is focused on the ownership status of household dwellings, an indicator variable taking value of either owned (including on loan) or rented. This is our $\boldsymbol{V}_{\text{Hold}} \setminus \boldsymbol{V}_{\text{Match}}$. The invariants $\boldsymbol{c}_{\text{Swap}}$ induced by this swapping scheme include 1) the total number of owned versus rented dwellings at each of the

Table 1. A comparison of two-way tabulations of dwelling ownership by county based on the 1940 Census full count for the state of Massachusetts (left) and one instantiation of the PSA at $p = 50\%$ (right).

County	Owned	Rented	Total	Owned (swapped)	Rented (swapped)	Total (swapped)
Barnstable	7,461	3,825	11,286	5,907	5,379	11,286
Berkshire	14,736	18,417	33,153	13,770	19,383	33,153
Bristol	33,747	63,931	97,678	35,537	62,141	97,678
Dukes	1,207	534	1,741	946	795	1,741
Essex	53,936	81,300	135,236	52,631	82,605	135,236
Franklin	7,433	6,442	13,875	6,337	7,538	13,875
Hampden	30,597	58,166	88,763	32,267	56,496	88,763
Hampshire	9,427	8,630	18,057	8,145	9,912	18,057
Middlesex	104,144	147,687	251,831	100,372	151,459	251,831
Nantucket	593	432	1,025	471	554	1,025
Norfolk	44,885	40,285	85,170	38,566	46,604	85,170
Plymouth	24,857	23,882	48,739	21,549	27,190	48,739
Suffolk	49,656	176,553	226,209	67,357	158,852	226,209
Worcester	53,126	78,535	131,661	51,950	79,711	131,661
Total	435,805	708,619	1,144,424	435,805	708,619	1,144,424

Note: Total dwellings per county, as well as total owned versus rented units per state, are invariant; the values of these statistics are not shown.

household sizes at the state level, and 2) the total number of dwellings at each of the household sizes at the county level. In our notation, these are the $n_{m\cdot s}$'s and the $n_{mh\cdot}$'s, respectively.

We restrict our illustration to the state of Massachusetts. Table 1 compares the two-way tabulations of dwelling ownership by county based on the original data and one instantiation of the swapping mechanism using a high swap rate of $p = 50\%$. The row margin of either table is the county-level total dwellings and is invariant due to $n_{\cdot h\cdot} = \sum_m n_{mh\cdot}$. The column margin is the total number of owned versus rented dwellings in Massachusetts and is invariant due to $n_{\cdot\cdot s} = \sum_m n_{m\cdot s}$.

Table 2 supplies the conversion between different swap rates to the protection loss ε of the PSA. Under the current swapping scheme, the largest stratum size delineated by $\boldsymbol{V}_{\text{Match}}$ is $b = 264,331$, consisting of all two-person households of Massachusetts. Therefore by Equation 3.3, we see that a low swap rate of 1% corresponds to $\varepsilon = 17.08$, whereas a high swap rate of 50% corresponds to $\varepsilon = 12.48$. It is worth noting that since the invariants $\boldsymbol{c}_{\text{Swap}}$ are fixed in this analysis, the different values of ε presented in this table can be directly interpreted as SDC guarantees of different quantified strengths. On the other hand, as we alluded to earlier, the protection losses corresponding to different invariants $\boldsymbol{c}_{\text{Swap}}$ are not directly comparable—see the discussion in Section 5.4.

We also examine the accuracy of the two-way tabulation as a function of swap rate. Figure 1 shows the mean absolute percentage error (MAPE) in the two-way tabulation induced by swapping at different swap rates from 1% to 50%. The variability across runs is small: each boxplot reflects 20 independent runs of the PSA.

Here, the MAPE of a swapped table from its true table is defined as the cell-wise average of the ratio between their absolute differences and the true table values. The MAPE in Figure 1 is with

Table 2. Conversion of (expected) swap rate p to protection loss ε.

p	0.01	0.05	0.10	0.50
ε	17.08	15.43	14.68	12.48

Note: Under this swapping scheme, the largest stratum size is $b = 264,331$, the number of all two-person households of Massachusetts.

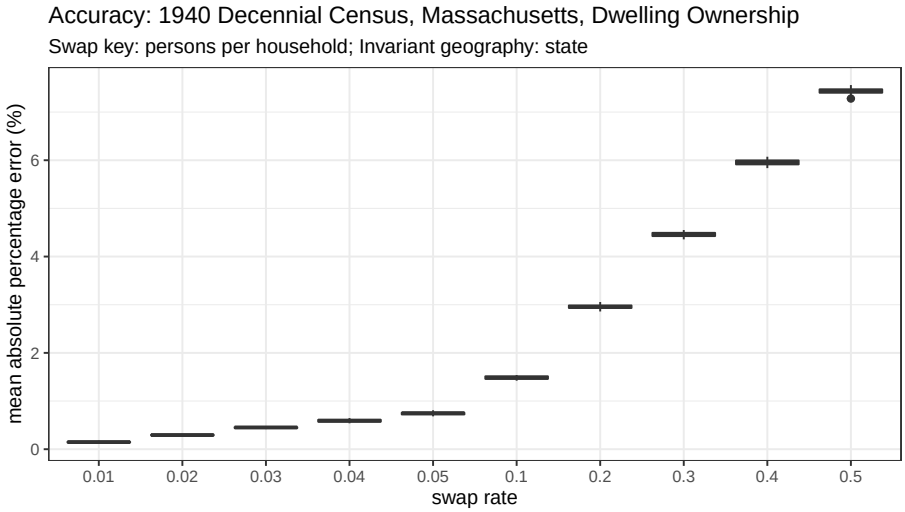

Figure 1. Mean absolute percentage error in the two-way tabulation of dwelling ownership by county induced by the Permutation Swapping Algorithm applied to the 1940 Census full count data of Massachusetts, at different swap rates from 1% to 50%. Each boxplot reflects 20 independent runs of the PSA at that swap rate.

respect to the contingency table of county by dwelling ownership in Massachusetts and is defined in the notation of Proposition 4 as

$$\frac{1}{\mathcal{HS}} \sum_{h,s} \frac{\left| n^{\boldsymbol{x}}_{\cdot hs} - n^{\boldsymbol{Z}}_{\cdot hs} \right|}{n^{\boldsymbol{x}}_{\cdot hs}},$$

where $\boldsymbol{x}$ is the true table, $\boldsymbol{Z}$ is the swapped table, h is the county indicator, and s the indicator of whether the house was rented or owned.

The accuracy assessment we demonstrate here is highly limited. The analysis above assesses only cell-wise departures of the swapped two-way marginal table from its confidential counterpart. It does not capture potential loss of data utility in terms of multivariate relational structures. It is well understood in the literature that swapping erodes the correlation between $\boldsymbol{V}_{\text{Swap}}$ and $\boldsymbol{V}_{\text{Hold}} \setminus \boldsymbol{V}_{\text{Match}}$ (see, e.g., Drechsler & Reiter, 2010; Mitra & Reiter, 2006; Slavković & Lee, 2010). For the current example, this means the countywide characteristics of household dwellings (other than their size) are not preserved, but other multivariate relationships are. While an in-depth investigation into the utility of swapping is out of the scope of this article, we return to the subject of data utility in (Bailie et al., 2026a) to discuss the implication this work may have on that line of inquiry.

3.5. **Estimating the Differential Privacy Specification of the 2010 DAS.** If we entertain the assumption that the 2010 DAS used the PSA, we could obtain a crude sketch of the SDC

guarantee afforded to the 2010 Census data. (We examine the validity of this assumption in detail in Appendix C.1.) As detailed in Example 2, the 2010 DAS swapped household-records. Therefore, the DP flavor for the 2010 DAS would be $(\mathcal{X}_{\text{CEF}}, \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}, d^{hh}_{\text{HamS}}, D_{\text{Mult}})$. Here the domain $\mathcal{X}_{\text{CEF}}$ is the set of all possible Census Edited Files (CEFs), where the term CEF refers to the data set that consists of the census data after editing and imputation and is the input into the USCB's DAS.

The 2010 DAS utilized a swap key that includes the household size as well as the household voting age population and some geography (either tract, county, or state) (Example 2). As we are unable to locate the 2010 Census data that allows for the precise calculation of b pertaining to this particular swapping scheme, the swap key we consider here is coarser: we set the matching variables $\boldsymbol{V}_{\text{Match}}$ to be the household's size and state (so we do not include the third matching variable of the 2010 DAS, the household count of voting age persons). This results in a value of $b = 3.65$ million, which—as we use a coarser swap key—is an upper bound for the actual b of the 2010 Census. Combining $b = 3.65$ million with a purported swap rate p between 2–4% (boyd & Sarathy, 2022) implies that the nominal budget ε is between 18.29 and 19.

The range 18.29–19 is an overestimate for the PLB; using the actual 2010 swap key would decrease the value of b and hence result in a smaller value of ε. However, we emphasize that the range 18.29–19 for the value of ε does not necessarily reflect the PLB of the 2010 DAS, but rather the PLB of the PSA when we choose its parameters to reflect what we know about the 2010 DAS. Moreover, as is always the case, this protection loss budget must be interpreted within the context of its DP flavor. Crucially, the DP flavor $(\mathcal{X}_{\text{CEF}}, \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}, d^{hh}_{\text{HamS}}, D_{\text{Mult}})$ for the above instantiation of the PSA includes the invariants $\boldsymbol{V}_{\text{Hold}}$ and $\boldsymbol{V}_{\text{Swap}} \times \boldsymbol{V}_{\text{Match}}$ (Proposition 4). Under the 2010 parameter choices, these invariants are the counts of households by number of occupants at the block level, and all cross-classifications of nongeographical variables at the state level. Hence, the measure of protection loss provided by the above values of ε are modulo any SDC leakage caused by the release of these invariants.

3.6. **What Does the Permutation Swapping Algorithm's Budget Look Like?** Figure 2 provides a visual illustration of Theorem 1, connecting the PLB ε to the swap rate p for a number of values for the largest stratum size b. Three observations are worth noting. First, for each b, there exists a smallest ε, call it $\varepsilon^{(b)}$, which lower bounds the PLB of the PSA (regardless of its swap rate $p \in [0, 1]$). The minimum budget $\varepsilon^{(b)} = \ln(b+1)/2$ is achieved by the swap rate $p^{(b)} = \sqrt{b+1}/(\sqrt{b+1}+1)$. For each b in Figure 2, this quantity $\varepsilon^{(b)}$ is marked by an outlined diamond. Importantly, the larger the b, the larger the minimum possible budget $\varepsilon^{(b)}$. For example, when $b = 10$, $\varepsilon^{(b)}$ is 1.20 (attained at $p^{(b)} = 77\%$); whereas when $b = 10^6$, $\varepsilon^{(b)}$ is 6.91 (at $p^{(b)} = 99.9\%$).

That some PLBs are not attainable for a fixed b follows from the fact that the ratio $\mathsf{P}(\boldsymbol{z} \mid \boldsymbol{x})/\mathsf{P}(\boldsymbol{z} \mid \boldsymbol{x}')$ of probabilities of a swapped data set $\boldsymbol{z}$ from two different input data sets $\boldsymbol{x}$ and $\boldsymbol{x}'$ depends not just on the swap rate p but also the ratio r of the number of derangements of size $d^{hh}_{\text{HamS}}(\boldsymbol{x}, \boldsymbol{z})$ to the number of derangements of size $d^{hh}_{\text{HamS}}(\boldsymbol{x}', \boldsymbol{z})$. (This is because the PSA selects $d^{hh}_{\text{HamS}}(\boldsymbol{x}, \boldsymbol{z})$ records and then samples derangements of size $d^{hh}_{\text{HamS}}(\boldsymbol{x}, \boldsymbol{z})$ uniformly at random.) This ratio r is upper bounded by $(b+1)^{d^{hh}_{\text{HamS}}(\boldsymbol{x}, \boldsymbol{x}')}$, which means ε must be at least $\ln(b+1) - \ln p + \ln(1-p)$.

Second, for every b and every budget $\varepsilon > \varepsilon^{(b)}$, two different swap rates can achieve that budget ε, with the higher rate often being very close to 100%. For example at $b = 10$, a swap rate of either 35.4% or 95.2% achieves the nominal budget of $\varepsilon = 3$. The mathematical reason behind this is that, for large p (i.e., $p > p^{(b)}$) the ratio r is dominated by the odds $o = p/(1-p)$, in which case $[\ln \mathsf{P}(\boldsymbol{z} \mid \boldsymbol{x}) - \ln \mathsf{P}(\boldsymbol{z} \mid \boldsymbol{x}')]/d^{hh}_{\text{HamS}}(\boldsymbol{x}, \boldsymbol{x}')$ is maximized when $d^{hh}_{\text{HamS}}(\boldsymbol{x}', \boldsymbol{z})$ is as small as possible.

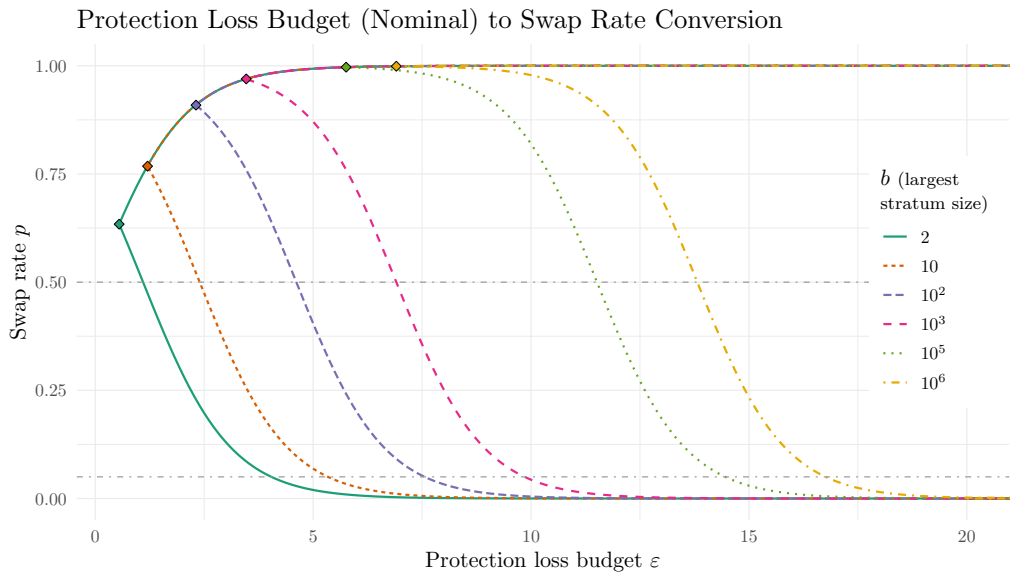

Figure 2. Conversion between the nominal protection loss budget ε and the swap rate p for the Permutation Swapping Algorithm. Color and line type encode different values of b, the size of the largest stratum delineated by $\boldsymbol{V}_{\text{Match}}$ (from 2 to 1 million). Outlined diamonds indicate the smallest ε attainable for each b. Grey dotted horizontal lines correspond to swap rates of 5% and 50% respectively. The protection loss budgets are nominal in that the statistical disclosure control guarantee they afford must be understood in the context of $\boldsymbol{c}_{\text{Swap}}$ (and hence the values of ε across different values of b are not immediately comparable).

This results in $\varepsilon = \ln o$, while, as explained in the previous paragraph, $\varepsilon = \ln(b+1) - \ln o$ for $p \leq p^{(b)}$. Since the former ε is monotone increasing in p and the latter monotone decreasing, there are two swap rates p corresponding to any $\varepsilon > \varepsilon^{(b)}$. This is akin to the behavior of the randomized response mechanism, where a large probability $p_{\text{RR}} > 0.5$ of flipping the binary confidential answer inadvertently preserves statistical information, thereby achieving the same budget $\varepsilon = |\ln o_{\text{RR}}|$ as $1 - p_{\text{RR}}$.

Third and most importantly, we emphasize that the budgets visualized in Figure 2 are *nominal* in the sense that the SDC guarantee they afford must be understood with respect to the full context as outlined by the PSA's DP specification. An aspect of this context is b, the size of the largest stratum of $\boldsymbol{V}_{\text{Match}}$, and as a result, the same value of ε across different b's should not be equated to be the same SDC guarantee. Indeed, the ordering of the b curves in the figure suggests a seemingly peculiar fact that, for a larger b, a higher p is needed to achieve the same ε. This apparent contradiction is explained by a point we have repeatedly made: for a fixed data set, a change in the value of b requires that the swapping invariants, and hence the PSA's SDC guarantee, also change.

4. A Differentially Private Analysis of the TopDown Algorithm

This section provides a DP specification for the TDA (Abowd et al., 2022; U.S. Census Bureau, 2023c). The TDA was used to produce the P.L. 94-171 Redistricting Summary File (PL) (U.S. Census Bureau, 2021a, 2021b) and the Demographic and Housing Characteristics File (DHC) (Cumings-Menon et al., 2025; U.S. Census Bureau, 2023e) for the 2020 U.S. Decennial Census. Four

other products—the Demographic Profile (U.S. Census Bureau, 2023b), the Privacy-Protected Microdata Files (PPMF) (U.S. Census Bureau, 2024c), the Redistricting and DHC Noisy Measurement Files (NMF) (U.S. Census Bureau, 2023j, 2023m), and 118th Congressional District Summary File (U.S. Census Bureau, 2023k)—were also derived during the production of the PL and the DHC. Hence the publication of these four additional data products do not contribute to additional privacy loss, and our DP specifications for the PL and the DHC automatically extend to cover the release of all six products.

We prove in Theorem 2 that the TDA satisfies zero-concentrated DP (zCDP) (Bun & Steinke, 2016) subject to its invariants. By this we mean that the TDA satisfies a DP specification whose output premetric is the normalized Rényi metric D_{NoR}—the output premetric corresponding to zCDP (see Appendix G)—and whose multiverse contains the TDA's invariants. More exactly, we show that the TDA satisfies the DP flavor $(\mathcal{X}_{\mathrm{CEF}}, \mathscr{D}_{\boldsymbol{c}_{\mathrm{TDA}}}, d^p_{\mathrm{HamS}}, D_{\mathrm{NoR}})$, where $\mathscr{D}_{\boldsymbol{c}_{\mathrm{TDA}}}$ is the multiverse induced by the TDA's invariants: the state population totals; the total number of housing units in each census block; and the count of each type of occupied group quarters in each block. By proving that the TDA cannot satisfy zCDP (with input premetric d^p_{HamS}) for any finite protection loss budget without conditioning on these invariants, we will also show that the TDA's DP flavor must have $\mathscr{D}_{\boldsymbol{c}_{\mathrm{TDA}}}$ (or a refinement of $\mathscr{D}_{\boldsymbol{c}_{\mathrm{TDA}}}$) as its multiverse.

The TDA, summarized in Algorithm 2, was run twice for the 2020 Census—once to produce the PL and then a second time for the DHC. The PLBs associated with releasing these files are given in Table 3. It is a two-step procedure: The first step (called the "measurement phase" in Abowd et al., 2022) produces the NMF $\boldsymbol{T}_p(\boldsymbol{x}_p)$ and $\boldsymbol{T}_{hh}(\boldsymbol{x}_{hh})$. Here $\boldsymbol{x}_p$ and $\boldsymbol{x}_{hh}$ denote the representations of the CEF at the person and household levels, respectively. The NMF are noisy versions of tabular summaries $\boldsymbol{Q}_p(\boldsymbol{x}_p)$, at the person level, and $\boldsymbol{Q}_{hh}(\boldsymbol{x}_{hh})$, at the household level, respectively. (In this section, we will include group quarters as households for the purposes of conciseness.) The tabular summaries $\boldsymbol{Q}_p(\boldsymbol{x}_p)$ and $\boldsymbol{Q}_{hh}(\boldsymbol{x}_{hh})$ are different for the PL and the DHC, but, roughly, they are the raw statistics that the USCB would like to include in each of these files. For example, when releasing the PL, $\boldsymbol{Q}_p(\boldsymbol{x}_p)$ and $\boldsymbol{Q}_{hh}(\boldsymbol{x}_{hh})$ are the statistics in this file, but aggregated directly from the census microdata without any noise. (However, to improve accuracy, the bureau adds some additional statistics to $\boldsymbol{Q}_p(\boldsymbol{x}_p)$ and $\boldsymbol{Q}_{hh}(\boldsymbol{x}_{hh})$, which do not appear in the PL.) Discrete Gaussian noise is added to $\boldsymbol{Q}_p(\boldsymbol{x}_p)$ and $\boldsymbol{Q}_{hh}(\boldsymbol{x}_{hh})$ to produce the NMFs $\boldsymbol{T}_p(\boldsymbol{x}_p)$ and $\boldsymbol{T}_{hh}(\boldsymbol{x}_{hh})$.

In the second step (called the "estimation phase" in Abowd et al., 2022), the PPMF $\boldsymbol{Z}_p$ and $\boldsymbol{Z}_{hh}$ are produced by solving a complex optimization problem. (The PPMF is also called the Microdata Detail File by Abowd et al., 2022.) The PPMF $\boldsymbol{Z}_p$ and $\boldsymbol{Z}_{hh}$ agree with the CEF $\boldsymbol{x}_p, \boldsymbol{x}_{hh}$ on the invariants $\boldsymbol{c}_{\mathrm{TDA}}$. In addition, the PPMF $\boldsymbol{Z}_p$ and $\boldsymbol{Z}_{hh}$ for the DHC are consistent with related statistics in the PL (U.S. Census Bureau, 2023h). To enforce this consistency, the PL $\boldsymbol{P}$ is passed as input into the TDA when producing the DHC and a constraint $\boldsymbol{H}(\boldsymbol{Z}_p, \boldsymbol{Z}_{hh}) = \boldsymbol{P}$ is added to the optimization problem. (The input $\boldsymbol{P}$ is not used by the TDA in producing the PL.)

The PL and the DHC are tabulations of the PPMF data sets $\boldsymbol{Z}_p$ and $\boldsymbol{Z}_{hh}$. In addition to the PL and the DHC, the USCB released the NMF $\boldsymbol{T}_p(\boldsymbol{x}_p)$ and $\boldsymbol{T}_{hh}(\boldsymbol{x}_{hh})$ produced for the PL and the DHC (U.S. Census Bureau, 2023d), and the PPMF $\boldsymbol{Z}_p$ and $\boldsymbol{Z}_{hh}$ produced for the DHC (U.S. Census Bureau, 2023g). The Demographic Profile and the 118th Congressional District Summary File are retabulations of the DHC (U.S. Census Bureau, 2023a, 2023k).

Table 3. The protection loss budgets of the mechanisms $\boldsymbol{T}_p$ (person) and $\boldsymbol{T}_{hh}$ (household) used in the first step of the TopDown Algorithm to produce the P.L. 94-171 Redistricting Summary File (PL) and Demographic and Housing Characteristics File (DHC).

		ρ^2	ε (with $\delta = 10^{-10}$)
PL	Household	0.07	2.70
	Person	2.56	17.90
DHC	Household	7.70	34.33
	Person	4.96	26.34
Total		15.29	52.83

Note: The value of ε for each row is computed using the conversion $\varepsilon = \rho^2 + 2\rho\sqrt{-\ln \delta}$ given in Bun & Steinke (2016) and adopted by the U.S. Census Bureau. (Hence the aggregate loss of 52.83 is not the sum of the individual ε's.) We follow the U.S. Census Bureau's choice of $\delta = 10^{-10}$.
Source: U.S. Census Bureau (2023f).

Algorithm 2: Overview of the TopDown Algorithm (Abowd et al., 2022), focusing on aspects salient to statistical disclosure control.

Input:

A CEF $\boldsymbol{x} \in \mathcal{X}_{\text{CEF}}$ with representations $\boldsymbol{x}_p$ and $\boldsymbol{x}_{hh}$ at the person and household levels respectively;
Person and household queries $\boldsymbol{Q}_p$ and $\boldsymbol{Q}_{hh}$;
Privacy noise scales Σ_p and Σ_{hh};
Constraints $\boldsymbol{c}^+_{\text{TDA}}$ (including invariants $\boldsymbol{c}_{\text{TDA}}$, edit constraints and structural zeroes);
(Optional) previously released statistics $\boldsymbol{P}$ along with an aggregation function $\boldsymbol{H}$, which specifies the relationship between $\boldsymbol{P}$ and the PPMF $\boldsymbol{Z}_p$ and $\boldsymbol{Z}_{hh}$.

Step 1: Noise Infusion
  Sample discrete Gaussian noise (Canonne et al., 2022):
    $\boldsymbol{W}_p \sim \mathcal{N}_{\mathbb{Z}}(\boldsymbol{0}, \Sigma_p)$
    $\boldsymbol{W}_{hh} \sim \mathcal{N}_{\mathbb{Z}}(\boldsymbol{0}, \Sigma_{hh})$
  Compute the NMF:
    $\boldsymbol{T}_p(\boldsymbol{x}_p) \leftarrow \boldsymbol{Q}_p(\boldsymbol{x}_p) + \boldsymbol{W}_p$
    $\boldsymbol{T}_{hh}(\boldsymbol{x}_{hh}) \leftarrow \boldsymbol{Q}_{hh}(\boldsymbol{x}_{hh}) + \boldsymbol{W}_{hh}$
Step 2: Post-Processing
  Compute the PPMF $\boldsymbol{Z}_p$ and $\boldsymbol{Z}_{hh}$ as a solution to the optimization problem:
    Minimize loss between $[\boldsymbol{T}_p(\boldsymbol{x}_p), \boldsymbol{T}_{hh}(\boldsymbol{x}_{hh})]$ and $[\boldsymbol{Q}_p(\boldsymbol{Z}_p), \boldsymbol{Q}_{hh}(\boldsymbol{Z}_{hh})]$
      subject to constraints $\boldsymbol{c}^+_{\text{TDA}}(\boldsymbol{Z}_p, \boldsymbol{Z}_{hh}) = \boldsymbol{c}^+_{\text{TDA}}(\boldsymbol{x}_p, \boldsymbol{x}_{hh})$ and $\boldsymbol{H}(\boldsymbol{Z}_p, \boldsymbol{Z}_{hh}) = \boldsymbol{P}$.

Output:

The PPMF $\boldsymbol{Z}_p$ and $\boldsymbol{Z}_{hh}$;
The NMF $\boldsymbol{T}_p(\boldsymbol{x}_p)$ and $\boldsymbol{T}_{hh}(\boldsymbol{x}_{hh})$ at the person and household levels.

Theorem 2. *The TDA satisfies the DP specification ρ-DP$(\mathcal{X}_{\text{CEF}}, \mathscr{D}_{\boldsymbol{c}_{\text{TDA}}}, d^p_{\text{HamS}}, D_{\text{NoR}})$ with PLB $\rho^2 = 2.63$ for the PL and $\rho^2 = 15.29$ for the DHC. (Note that these budgets do not vary with the universe $\mathcal{D} \in \mathscr{D}_{\boldsymbol{c}_{\text{TDA}}}$.)*

In the opposite direction, let $\boldsymbol{c}'$ be any proper subset of TDA's invariants. Then the TDA does not satisfy ρ-DP$(\mathcal{X}_{\text{CEF}}, \mathscr{D}_{\boldsymbol{c}'}, d^{p}_{\text{HamS}}, D_{\text{NoR}})$ with any finite budget ρ.

A proof of Theorem 2 is given in Appendix H.

Remark 2. Because the standard parametrization of zCDP's protection loss budget is equal to the square of our parametrization (Appendix G), throughout this article we report zCDP budgets in terms of ρ^2 to maintain consistency with the values reported in existing publications. We also drop the subscript $_{\mathcal{D}}$ when reporting 2020 Census budgets since these budgets do not vary with the universe $\mathcal{D}$.

5. What if the 2020 Census Used Swapping?

In this section we ask the counterfactual question: what if the PSA was applied to the 2020 Decennial Census? What would its protection guarantee look like? The answer is not a straightforward one for two reasons. First, the set of invariants implemented by the TDA for the 2020 Census does not conform to a valid invariant specification under any PSA regime (see Example 3). Second, the PSA may be deployed using different swapping schemes at varying swap rates, and we cannot know in the counterfactual world what swapping scheme and swap rate would have been chosen for the 2020 Census. Thus, our analysis proceeds in an exploratory fashion. We ask specifically: what would the disclosure risk for the 2020 Census look like under different choices for invariants, swapping schemes, and swap rates for the PSA, when these choices are made to mimic the design choices of the 2020 DAS, subject to reasonable and necessary departures?

To answer this question, we begin by spelling out the DP specifications of the 2020 Census and its components, and examine them alongside the DP specifications that can be achieved with the PSA under a range of hypothetical choices for its parameters. As the ensuing analysis will make clear, the distinct DP flavors of the 2020 DAS and the PSA hinder the extent to which their DP specifications may be directly compared against one another—most crucially due to the fact that the invariants of one of these flavors are neither strictly stronger nor weaker than the other.

5.1. An Overview of the 2020 Disclosure Avoidance System.

For the 'counterfactual PSA' to be at all relevant to the 2020 Decennial Census, we must first understand the DP specification of the 2020 DAS. This will allow us to ascertain the extent of compatibility between the PSA and the 2020 DAS, in order to facilitate an analysis of the DP specification of a would-be 2020 PSA deployment.

An overview of the 2020 DAS and its data products lays the foundation for our analysis of its DP specification. The USCB divides the 2020 Census data releases into three groups. Group 1 encompasses the two principal data products that we have already discussed, namely, the PL and the DHC. (The Demographic Profile is also included in Group 1 but as it is simply a subset of DHC's tabulations, we do not consider it as a standalone product.) As detailed in the previous section, both the PL and the DHC were protected using the TDA (Abowd et al., 2022; U.S. Census Bureau, 2023c). Group 2 encompasses the Detailed DHC-A (U.S. Census Bureau, 2023l) and Detailed DHC-B Files (U.S. Census Bureau, 2024a), which were produced by the SafeTab-P and SafeTab-H Algorithms respectively (Tumult Labs, 2022; U.S. Census Bureau, 2023n, 2024b). (Here '-P' and '-H' stand for 'persons' and 'households'.) Group 2 also includes the Supplemental DHC File (S-DHC) (U.S. Census Bureau, 2024d), protected using the PHSafe Algorithm (U.S. Census Bureau, 2024f).

Group 3 contains the additional products derived from the 2020 Census data, most notably the PPMF (U.S. Census Bureau, 2024c), the 118th Congressional District Summary File (U.S. Census Bureau, 2023k), and NMFs (U.S. Census Bureau, 2023j, 2023m). As explained in the previous section, because these data products are derived either from the Group 1 products or the privacy-protected intermediate outputs pertaining to those products, their production does not contribute to the overall 2020 PLB. As a result, we need not consider these Group 3 products in our analysis. Other Group 3 data releases, such as publications from researchers with access to census microdata, may be released in the future (Hawes, 2021b). Our analysis does not account for these releases, nor for products derived from 2020 Census data that are not listed here.

5.2. **Understanding the Differential Privacy Specifications of the 2020 Census.** The first five rows of Table 4 summarize the DP specifications of the 2020 DAS overall, as well as its constituent algorithms, including the TDA (Abowd et al., 2022; U.S. Census Bureau, 2023c), the SafeTab Algorithms (Tumult Labs, 2022; U.S. Census Bureau, 2023n, 2024b), and the PHSafe Algorithm (U.S. Census Bureau, 2024f). By way of contrast, the last row of Table 4 summarily presents the DP specification of the hypothetical application of the PSA to the 2020 Decennial Census, the details for which is expanded upon in Table 5 of Section 5.3.

The following remarks accompany the pertinent entries in Table 4:

Remark 3 (TDA's additional constraints). In addition to invariants, the TDA also enforces that the PPMFs $\boldsymbol{Z}_p$ and $\boldsymbol{Z}_h$ satisfy edit constraints and structural zeroes (Abowd et al., 2022). Edit constraints are rules that the USCB applies in their editing procedure to correct implausible or impossible census responses. One such rule is that a mother must be a certain number of years older than any of her children. If a census record does not satisfy an edit constraint, the USCB will modify it so that the edited record does. Structural zeroes are similar—they describe rules that cannot be broken by the census data because of constraints embedded in data collection. For example, if a block has an occupant, then one of the households in the block must have an occupant. As such, the CEF will always satisfy edit constraints and structural zeros. Because every possible $\boldsymbol{x} \in \mathcal{X}_{\text{CEF}}$ satisfies these rules by construction, these requirements need not be included as invariants.

Remark 4 (zCDP budget conversion). Because we reparametrize zCDP, we report its budgets in terms of ρ^2 (rather than ρ) to be consistent with other literature on the 2020 DAS (see Remark 2). Moreover, following the USCB, we use the formula from Bun and Steinke (2016), $\varepsilon = \rho^2 + 2\rho\sqrt{-\ln \delta}$, to convert from a zCDP budget ρ^2 to an approximate DP budget (ε, δ).

Remark 5 (Inequality invariants of PHSafe). The PHSafe Algorithm has inequality invariants (see Equation 2.4). Specifically, its invariant function $\boldsymbol{c}(\boldsymbol{x})$ is the vector of indicators for whether each census block has at least one housing unit or not.

Remark 6 (DP specification of the overall 2020 DAS). This DP specification covers all of the primary 2020 Census releases (U.S. Census Bureau, 2024e) discussed previously in this section, but not other data products that were derived from the 2020 CEF, such as the 2020 DAS accuracy metrics (U.S. Census Bureau, 2023i), the Population and Housing Unit Estimates (U.S. Census Bureau, 2023p), and the National Population Projections (U.S. Census Bureau, 2023o). We were unable to locate information on the DP specifications associated with these data products. Nevertheless, as with any data release, they necessarily increase the total PLB associated with the 2020 Census. They

Table 4. The differential privacy (DP) specifications of the 2020 disclosure avoidance system (DAS) and of the hypothetical application of the Permutation Swapping Algorithm (PSA) to the 2020 Decennial Census.

	D_{Pr}	$d_{\mathcal{X}}$ (Resolution)	Invariants[3]	Protection Loss Budget[4]
TopDown	D_{NoR}	d^{p}_{HamS} (person)	Population (state) Total housing units (block) Occupied group quarters by type (block)	PL & DHC: $\rho^2 = 15.29$ $(\varepsilon = 52.83, \delta = 10^{-10})$ See Table 3
SafeTab-P			Total housing units (block)	DDHC-A: $\rho^2 = 19.776$
SafeTab-H		d^{hh}_{HamS} (household)		DDHC-B: $\rho^2 = 17.79$
PHSafe			≥ 1 housing unit (block)[5]	S-DHC: $\rho^2 = 2.515$
Overall (to date) 2020 DAS[6]	D_{NoR}	d^{p}_{HamS} (person)	Same as TopDown	$\rho^2 = 55.371$ $(\varepsilon = 126.78, \delta = 10^{-10})$
Swapping (PSA)	D_{Mult}	d^{hh}_{HamS} (household)	Varies but much greater than TopDown[7]	ε between[8] 8.42–19.36

Note: The 2020 DAS consists of the TopDown Algorithm, which produced the P.L. 94-171 Redistricting Summary File (PL) and the Demographic and Housing Characteristics File (DHC); the SafeTab Algorithms, which produced the Detailed DHC-A File (DDHC-A) and Detailed DHC-B File (DDHC-B); and the PHSafe Algorithm, which produced the Supplemental DHC File (S-DHC). For each DP specification, the protection domain is the set $\mathcal{X}_{\text{CEF}}$ of all possible Census Edited Files and the multiverse is induced by the listed invariants. d^{p}_{HamS} and d^{hh}_{HamS} denote the Hamming distance at the resolution of person- and household-records, respectively (Equation 3.1); D_{NoR} the normalized Rényi metric (Appendix G), which is the output premetric underlying zCDP (Bun & Steinke, 2016); and D_{Mult} the multiplicative distance (Equation 3.2), which is pure DP's output premetric (Dwork et al., 2006). The numbered superscripts 3–8 in the table refer to the explanatory Remarks 3–8.

could also possibly weaken the 2020 Census's DP flavor by increasing the invariants, weakening the output premetric, or increasing the resolution of the input premetric. Moreover, the USCB may make additional releases in the future, such as the Surname File (U.S. Census Bureau, 2016) or research papers generated with access to census microdata (Hawes, 2021b). These releases would further weaken the DP specification for the 2020 Census. In comparison, under data swapping, the PLB and DP flavor covers all data releases (Section 5.3).

Remark 7 (PSA invariants). Depending on the swap key $\boldsymbol{V}_{\text{Match}}$ and the swapping variables $\boldsymbol{V}_{\text{Swap}}$, the invariants induced by the PSA are all (multivariate) household characteristics at either the state, county, or block group levels, and optionally the household size at the corresponding geography one level lower. See Section 5.3 for details.

Remark 8 (PSA budget). The exact PLB ε of the PSA depends on the swap rate p and the swap key $\boldsymbol{V}_{\text{Match}}$, with the combination of a higher swap rate and finer geography-household strata giving rise to the lower range and vice versa (Table 5).

Last but not least, we comment on the choice of protection units for both the 2020 DAS and the hypothetical PSA. The protection units (also known as the 'privacy' units) of a DP specification

are the basic entities that are protected under that DP specification. More exactly, a specification's PLB restricts how much a mechanism can change when a single protection unit's data changes. One might imagine that the protection units of a DP specification correspond to the resolution of its input premetric $d_{\mathcal{X}}$—and this is true, but only in simplistic examples. Here by 'the resolution of $d_{\mathcal{X}}$,' we mean the size of the change between $\boldsymbol{x}$ and $\boldsymbol{x}'$ when $d_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}') = 1$. For example, if $d_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}') = 1$ whenever $\boldsymbol{x}$ and $\boldsymbol{x}'$ differ on a person-record, then the resolution of $d_{\mathcal{X}}$ is a person. Common resolutions, in order from high to low, are: single transactions or interactions, persons, households, and businesses.

However, data preprocessing can create complications, so that the protection units of a specification are not always given by the resolution of $d_{\mathcal{X}}$. In the case of the Decennial Census, an individual contributor's data can be used for multiple records in the CEF $\boldsymbol{x} \in \mathcal{X}_{\text{CEF}}$ because the USCB's imputation procedure replaces missing records with copies of nonmissing records. As such, the protection units of $(\mathcal{X}_{\text{CEF}}, \mathscr{D}, d_{\mathcal{X}}, D_{\text{Pr}})$ do not correspond to the resolution of $d_{\mathcal{X}}$. Rather, the protection units of the 2020 DAS are 'post-imputation persons'—that is, those (fictional) entities with data that is exactly one record in the CEF. Similarly, the PSA's protection units are the 'post-imputation households' rather than actual households.

The distinction between actual and post-imputation protection units is not simply a matter of semantics. From the perspective of a data contributor, the resolution of $d_{\mathcal{X}}$ is not particularly informative in determining the SDC protection provided to them. Rather, a more accurate measure of a contributor's actual protection loss is given by the nominal PLB multiplied by the number of person-records the contributor contributed to. For example, if the 2020 imputation process duplicates a contributor's record once, then their substantive PLB is $\rho^2 \geq 221.48$ (or $\varepsilon \geq 364.31$ with $\delta = 10^{-10}$), rather than $\rho^2 \geq 55.37$. (Here the ρ^2 budgets are inflated by a factor of four since doubling ρ quadruples ρ^2. Also, we write $\geq$ rather than $=$ because the contributor's protection loss will increase due to data releases that we have not accounted for, as explained in Remark 6.) In general, the conversion from a DP flavor with post-imputation persons as units to a DP flavor with persons as units requires an inflation of the PLB by a factor equal to the maximum number of times a record can be duplicated (for a proof of this, see the section on group privacy in Bailie et al., 2026b). To avoid this complication, we have reported post-imputation budgets in Table 4, but we caveat this with the important observation that these budgets are relative to unusual protection units.

5.3. **The Protection Guarantee of the 2020 Census Under Swapping.** Table 5 shows the total nominal PLB ε that would be achieved by applying the PSA to the 2020 Decennial Census for a variety of possible parameter choices. For the purpose of illustration, we stipulate the swapping variable $\boldsymbol{V}_{\text{Swap}}$ to be the block, tract, or county membership of each household, and the matching variable $\boldsymbol{V}_{\text{Match}}$ to be the geography one level higher than $\boldsymbol{V}_{\text{Swap}}$, either alone or crossed with the household size variable. These choices justify what Table 4 claims, that the invariants of the PSA "varies" but are "much greater than" those of the TDA (see Remark 7).

Note that if the PSA were applied to the 2020 Decennial Census, the nominal ε reported in Table 5 would be the *total* PLB across all data products derived from the swapped data set $\boldsymbol{Z}$, including the PL, the DHC, the DDHC, and the S-DHC, for both persons and household product types. This is because swapping is performed on the full CEF, and hence produces a synthetic version of it from which all data products can be generated. Therefore, when comparing the PLBs in Table 5 with those reported for the 2020 DAS in Table 4, it should be understood that the

protection loss for the PSA covers all the 2020 data products. This characteristic of swapping leads to an additional desirable property that is not necessarily enjoyed by mechanisms based on output noise infusion (such as those used in the 2020 DAS): the *logical consistency* between, and within, multiple data products is automatically preserved under swapping without the need for any additional processing, because all these data products are produced from the same post-swapped microdata.

The $\boldsymbol{V}_{\text{Swap}}$ levels in Table 5 are ordered in increasing granularity of geography. Within each level of $\boldsymbol{V}_{\text{Swap}}$, the two $\boldsymbol{V}_{\text{Match}}$ levels are nested, in the sense that the swapping scheme represented in the latter row (i.e., crossed with household size) induces a logically stronger and more constrained set of invariants than the former one. These $\boldsymbol{V}_{\text{Match}} \times \boldsymbol{V}_{\text{Swap}}$–level combinations result in largest strata of varying sizes, as can be seen from b ranging from as large as 13.47 million (the total number of households in California) to as small as $4,549$ (the total number of two-person households in a Florida block group).

Table 5. The total nominal protection loss budget ε for the Permutation Swapping Algorithm applied to the 2020 Decennial Census for a variety of $\boldsymbol{V}_{\text{Match}}$, $\boldsymbol{V}_{\text{Swap}}$, and swap rate choices.

$\boldsymbol{V}_{\text{Match}}$	$\boldsymbol{V}_{\text{Swap}}$	b	Total ε $p = 5\%$	Total ε $p = 50\%$	Largest Stratum
State	county	13,475,623	19.36	16.42	California
State × household size	county	3,948,028	18.13	15.19	California, two-household
County	tract	3,420,628	17.99	15.05	Los Angeles County
County × household size	tract	939,185	16.70	13.75	Los Angeles County, two-household
Block group	block	6,204	11.68	8.73	a California block group
Block group × household size	block	4,549	11.37	8.42	a Florida block group, two-household

Note: The column b is the number of households in the largest stratum, obtained from the Demographic and Housing Characteristics File. (The California and Florida block groups identified in rows 5 and 6 have 2020 Census GEOIDs 060730187001 and 121199114024, respectively.)

This analysis highlights an important, yet perhaps counterintuitive, observation. When the swap rate p is fixed, including more invariants decreases the nominal PLB ε of the PSA. As Table 5 shows, when swaps are performed freely across counties in a state, even a high swap rate of 50% renders a nominal ε that is much larger than that pertaining to swaps among households of the same size within a block group at a low swap rate of 5% ($\varepsilon = 16.42$ and 11.37, respectively). If these nominal ε's are taken at face value, one may be tempted to conclude that swapping schemes with finer invariants should be preferred from an SDC standpoint. Furthermore, one may find it convenient to also recognize that finer invariants are desirable from a data utility standpoint, for the obvious reason that more exact statistics about the confidential data are made known. However, as we warned right after presenting Theorem 1, such a conclusion—that finer invariants should benefit both utility *and* privacy—would be dangerously mistaken, for it overlooks the privacy leakage, in an ordinary sense of the phrase, due to the invariants alone. Indeed, it is the loss of protection due to releasing more invariants that results in less information remaining that needs protection, thereby creating the illusion that we can achieve better DP protection with finer invariants. This illusion highlights the criticality of interpreting ε within the context of its DP flavor, and the necessity of treating the invariants as an integral part of any protection guarantee.

5.4. **Comparing the 2020 DAS With Swapping: The Need for a Disclosure Risk-Based Assessment.** With the aid of our system of DP specifications, we are able to articulate both the flavor and the intensity of the DP guarantees of the 2020 Census as well as the hypothetical application of swapping to it. A side-by-side examination of the building blocks of these DP specifications in Tables 4 and 5 elucidates the similarities and differences between the protection provided by the PSA and the 2020 DAS. But we caution against reading these tables as a *direct* comparison between the two SDC approaches. In our view, these tables illustrate precisely a *lack* of comparability between these approaches, due to their distinct DP flavors, which in turn render comparisons of their PLBs as effectively one of 'apples' versus 'oranges.'

To be sure, the DP specifications of the 2020 DAS and the PSA share similarities that facilitate their comparison on a conceptual level. Firstly, both the 2020 DAS and the PSA have the same protection domain: the set of all possible CEFs $\mathcal{X}_{\text{CEF}}$. This means that the PSA and the 2020 DAS protect the data $\boldsymbol{x} \in \mathcal{X}_{\text{CEF}}$ as it exists after collection, coding, editing, and imputation, rather than as it exists at other stages in its life cycle. As such, it is not the contributors' data (i.e., their 'raw' census responses) that are directly protected, but rather it is the edited and imputed data (i.e., the CEF) that receives the DP guarantee.

Secondly, because both the 2020 DAS and the PSA have invariants, each of their DP specifications partition the protection domain $\mathcal{X}_{\text{CEF}}$ into multiple universes. This operation constrains the scope of the 2020 DAS and the PSA's SDC protection to data sets that agree on their invariants. Therefore, for the same reasons that the 2020 DAS cannot satisfy the original specification of zCDP given in Bun and Steinke (2016), data swapping cannot satisfy the original pure DP specification of Dwork et al. (2006). In this sense, both the 2020 DAS and the PSA are DP only insofar as their invariants allow.

Thirdly, the input premetrics for the PSA and the 2020 DAS are both Hamming distances, although with differing resolutions—household-records for the PSA and person-records for the 2020 DAS. This means the protection units are post-imputation households and post-imputation persons, respectively. Since the input premetric is the yardstick for measuring change in the input data, using a lower resolution like household-records provides more protection than a higher resolution like person-records (all else being equal). That is, a household-level distance is a stronger notion than a person-level distance, since if the record of a single household changes part of its value, the multiple persons residing in a same household may all change their records.

Fourthly, the PSA's output premetric is also stronger than the 2020 DAS's. The PSA uses the multiplicative distance D_{Mult}—as used in pure DP (Dwork et al., 2006)—while the 2020 DAS uses the normalized Rényi metric D_{NoR}—as used in zCDP (Bun & Steinke, 2016). There exist probabilities P and Q with $D_{\text{NoR}}(\mathsf{P}, \mathsf{Q})$ arbitrarily small but $D_{\text{Mult}}(\mathsf{P}, \mathsf{Q}) = \infty$. As such, D_{Mult} ensures a greater level of SDC protection than D_{NoR} (again, assuming that all else is equal).

On the other hand, however, there are crucial differences between the DP specifications of the 2020 DAS and the PSA that ultimately render their comparison a wrought endeavor. The most damning obstacle to a meaningful comparison between the DP flavors of the 2020 DAS and the hypothetical PSA is that they do not, and cannot (on the hypothetical PSA's account) induce invariants in the data product that are equal, or even that are nested with respect to one another. As Proposition 4 makes clear, a swapping regime maintains invariant two marginal tables, that is, the cross-classifications of the matching variables by the swapping variables and that of all the holding variables, while perturbing the interior cells of the multiway contingency table. The 2020 DAS, on the other hand, used a list of invariants and other constraints on an as-needed basis.

They need not, by design, accord to some marginal tabulations of the underlying microdata. As a matter of fact, all of the SDC methods used in 2020 have invariants, but the PSA has many more invariants than any of these methods and, as such, places more restrictions on the scope of protection. On the flip side, while swapping almost always has stricter invariants for most variables, it does not necessarily have the TDA's group quarter invariants. Therefore, the 2020 DAS DP flavor is not strictly stronger than the PSA's flavor, nor visa versa—although the 2020 DAS places less restrictions on the scope of protection, these are not nested within the restrictions induced by the PSA's invariants.

Even with nested invariants—for example, in the case that the 2020 DAS's invariants were compatible with some swapping regime, the extent of comparison is still limited to a qualitative degree. We should all agree that when all else are held the same, a smaller set of invariants is strictly less disclosive than a larger superset of invariants (Section 2.4). But just how much less disclosive? We are still left to appeal to our intuition to reason with the privacy leakage induced by the additional invariants, with no quantitative guidance available.

As a consequence of the incommensurability between their DP flavors, the PLBs of the PSA and of the 2020 DAS are not directly comparable because a budget's 'unit of measurement' is determined by its DP flavor. That is to say, a PLB is a nominal measure of SDC protection, which is always relative to—and hence can only be understood within the context of—the four other building blocks. The DP flavors for the PSA and the 2020 DAS are different and so their budgets have different units of measurement.

An astute reader may point out that just like the USCB, we can convert the 2020 DAS zCDP budget $\rho^2 = 55.371$ to the approximate DP budget of $\varepsilon = 126.78$ with $\delta = 10^{-10}$, which is more comparable with a pure DP budget ε under the view that the latter has an $\delta = 0$. This calculation would appear to reveal that the PLB of the 2020 DAS is an order of magnitude larger than that of the PSA. We caution, again, that a budget conversion does not render the PSA and the 2020 DAS comparable because their DP flavors have different invariants and different input premetrics. While the 2020 DAS's budget would substantially increase under a household-level input premetric (Appendix H), removing even one invariant from the PSA's DP specification would result in an infinite PLB (Section 2.4.

It seems to us that the ultimate comparison between two competing SDC algorithms that sport different DP specifications should be phrased in terms of disclosure risk. The focus on disclosure risk is appropriate as it has been traditional in the SDC literature; see, for example, Kazan and Reiter (2025) for a demonstration of such assessment in the context of the 2020 Census. Articulating the DP specifications of SDC algorithms and recognizing the differences among competing options are essential to understanding their likely impact on disclosure risk. Yet, most crucially, disclosure risk stands to be the most promising, if not the only viable, yardstick for measuring the efficacy of SDC protection in the presence of invariants. Methodologically, we are encouraged by the observation that releasing a statistic under a large PLB is pragmatically equivalent to making that statistic an invariant. Hence, in principle it should be possible to effectively trade off invariants with large budgets, thus making the comparison between the budgets of SDC algorithms with distinct flavors, such as the PSA and the 2020 DAS, a more tractable one. Another possible equating principle is the vulnerability of a DP mechanism (and therefore by extension, its specification) to a reconstruction attack. For example, Abowd et al. (2025) show that swapping at a rate of 50% had about the same protection against reconstruction as the TDA (on the DHC). Hence, the increase from the TDA's invariants to the 2010 invariants might be roughly comparable to the increase in budget from

$\varepsilon \approx 8.42$–19.36 (the swapping budget, at the household level) to $\varepsilon \approx 52.83$ (the DHC budget, with $\delta = 10^{-10}$, at the person-level). Comparisons between the performance of swapping and the 2020 DAS against other privacy attacks could potentially provide additional heuristics on how invariants can be compared to large budgets (Ballesteros et al., 2025; Christ et al., 2022; Steed et al., 2025). Sections 6.1 and 6.2 offer some of our preliminary thoughts on the impact of invariants on disclosure risk, though we leave this question largely as a subject for future research.

6. Discussion

This article continues an existing line of research (Bailie & Chien, 2019; Chien & Sadeghi, 2024; Neunhoeffer et al., 2025; Rinott et al., 2018) examining traditional SDC methods—which are typically regarded as ad hoc and are motivated by intuitive notions of protection or specific attacker models —through the lens of DP. This body of literature shows that—even though they were designed without DP in mind—traditional SDC methods can still be fruitfully analyzed from the perspective of DP. By providing another example—data swapping—that can be studied theoretically via the lens of DP, we hope to inspire further formal analyses of other traditional SDC methods. This type of analysis improves our understanding of such methods by supplying mathematical descriptions of the level and substance (or, in our terminology, the intensity and flavor) of the methods' SDC. Such descriptions are important: they can provide assurance to data providers and custodians that their data is adequately protected; or, conversely, they can reveal inadequate SDC and spur additional protection.

However, it can be challenging to assess whether a given DP specification provides an adequate level of protection. To do so, we must understand how choices for each of the five building block can affect SDC—both individually and in conjunction with choices for the other building blocks. This requires answering a range of difficult sociotechnical questions. For instance, taking the other four building blocks as fixed, what PLB (if any) is sufficient for adequate SDC—adequate for whom, and who should decide this adequacy? Also, what is the practical impact of the protection units being post-imputation persons, and how should, if at all, this impact be mitigated? Furthermore—and most relevant for this article—what is the effect of invariants on SDC?

While we know that increasing the invariants strictly weakens the DP specification (Section 2.4), it is more difficult to determine how they affect an attacker's ability to make disclosures. Ashmead et al. (2019) have investigated the effect of the 2020 DAS invariants, but there is a need for future work that studies the effect of invariants at the scale of those induced by data swapping. In addition to building technical understanding—and parallel to studies that survey preferences on appropriate settings for the PLB, protection domain and input and output premetrics—it could be beneficial to gauge public opinion on the acceptability of specific invariants.

Nevertheless, by providing DP specifications for both the PSA and the TDA, we demonstrate the feasibility of mathematically comparing and contrasting on fair grounds traditional SDC methods with DP-based mechanisms. With these two algorithms as prime examples, the paper points to the possibility of similar comparative analyses between other SDC methods, both those that were explicitly inspired by DP and those that were designed and motivated from non-DP perspectives. By explicating the five building blocks for the PSA and the TDA, we hope to promote nuanced assessments of DP deployments that go beyond discussion of the PLB.

6.1. **Understanding the Impact of Invariants on Disclosure Risk.**

A major criticism of the swapping method implemented in the 2010 Census is that it induces too many invariants. One

salient consequence of a plurality of invariants is that it severely constrains the permissible values for the confidential data. Indeed, the larger the number of invariants, the more data sets an attacker can rule out as impossible, and, consequently, the higher the risk of disclosure. As Abowd and Hawes (2023) discuss, the invariants in the 2010 Census elevate disclosure risk because not only are they numerous but they also include information at a very fine granularity, for example, the total and voting age populations at the block level.

The USCB's reconstruction and reidentification attack against their 2010 Census (Abowd et al., 2025) provides some understanding of the impact of its invariants on disclosure risk, although, as we will see, there were also other confounding factors at play. As its name suggests, a reconstruction and reidentification attack consists of two steps: a reconstruction attack, which creates a plausible version of the confidential microdata; followed by a reidentification attack, which links this 'reconstructed' data set to an external source in order to attach names or other personally identifying information to (some or all of) the reconstructed records. Because this second step assigns identities to the linked microdata records, an attacker can use the resulting data to infer some information about some specific census respondents—such as their race or ethnicity—although they cannot be certain that this inference is correct due to both the potential for linkage errors and the uncertainty introduced by swapping.

A reconstruction attack works by collating many publicly available, aggregate statistics about the confidential (unknown) microdata. Then it constructs a data set that agrees with these statistics—or, in the case where the published statistics are noisy functions of the confidential microdata, it *infers* a data set based on a statistical model of the published data.[1] This reconstructed data set is a plausible guess for the confidential microdata, since, in the case where the statistics are deterministic functions of the confidential microdata, it generates identical statistics to the ones generated by the microdata. When the statistics are noisy, the situation is more complex: the reconstructed data set does not necessarily reproduce the published statistics exactly due to their stochasticity, but nevertheless it still could plausibly be the microdata that generated these statistics. In any case, the larger the number of published statistics and the more accurate they are, the more heavily they constrain the possible configurations of the reconstructed data set, and hence the more likely it is that this reconstruction agrees with the true confidential microdata.

For the reconstruction attack on the 2010 Census, the underlying microdata the USCB targeted was not the CEF$\boldsymbol{x}_{\text{CEF}}$, but the post-swapped data—that is, the resulting records after swapping had been applied to $\boldsymbol{x}_{\text{CEF}}$. This avoids the complication of designing a reconstruction attack that accounts for the noise introduced by swapping. Yet, because of the low swap rate and the large number of invariants in the 2010 DAS, there is a high degree of alignment between the reconstructed data and $\boldsymbol{x}_{\text{CEF}}$. As a result, linking the reconstructed data to an external data set containing personally identifiable information allows for the possibility of learning potentially sensitive data: the race and ethnicity of census respondents. Indeed, the USCB found that an attacker could predict the race and ethnicity of about 3.4 million vulnerable individuals (i.e., about 1% of the U.S. population) with 95% accuracy (Abowd et al., 2025), although an actual attacker would not be able to verify their level of accuracy or identify which individuals are vulnerable since they do not

[1]This model has the unknown microdata as its parameters, the released statistics as its data, and the SDC method that generated the published statistics from the confidential microdata as its data-generating process. For example, the reconstructed data set could be the maximum likelihood estimate under this model, or it could be a draw from a Bayesian posterior that is compatible with this model.

have access to the confidential data like the USCB does. As discussed in Section 1.1, it is partly from these observations that the bureau concluded an urgent need to revamp their swapping-based SDC. However, while the bureau's later experiments further suggest that the rate of swapping must be significantly increased to achieve what it deemed as an acceptable level of protection (Abowd & Hawes, 2023), the question remains open: in what ways does imposing a specific set of invariants impact the disclosure risk of the resulting data product?

To understand the relationship between swapping's invariants and disclosure risk, two caveats are worth noting at the outset. First, the degree of vulnerability to a reconstruction attack is a measure of *absolute disclosure risk* (Duncan & Lambert, 1986; Reiter, 2005), defined as the degree of certainty with which an attacker can make inferences about confidential information from the published data. Unless strong assumptions about the attacker's prior knowledge are made, DP does not directly translate into any quantifiable degree of control over the absolute disclosure risk; see, for example, Dwork and Naor (2010), Hotz and Salvo (2020), Kifer and Machanavajjhala (2011), and McClure and Reiter (2012). Similarly, absolute measures of reconstruction attack success (such as percentages of successfully reconstructed records, as reported by the USCB) are not controlled by DP (Francis, 2022; Kenny et al., 2021). Second, invariants are not unique to swapping, nor should they be viewed as a static byproduct. The final choice of invariants used in the 2020 TDA was arrived at by the USCB through an iterative process. For example, block-level populations were once considered as an invariant, but were ultimately not included (Abowd et al., 2022; Ashmead et al., 2019; Kifer, 2019).

Notwithstanding these caveats, it is worthwhile to inquire, to the extent possible, about the impact of invariants on disclosure risk through the lens of DP. Such an inquiry can be challenging within the standard DP paradigm, because the impact of invariants cannot be captured by the PLB. By contrast, our system of DP specifications is more dexterous because invariants can be explicitly included through the multiverse $\mathscr{D}$. An analysis of the impact of invariants on disclosure risk therefore amounts to a five-dimensional comparison between alternative DP specifications that differ on $\mathscr{D}$ and potentially on other building blocks as well. A comprehensive description of the five-way dynamics remains open for future research, although investigations with a restricted scope (e.g., only varying two or three dimensions at a time, rather than all five) can already be informative. For example, it can be shown that reconstruction attacks can be increasingly successful if applied to DP-protected data when more invariants are imposed on them (Protivash et al., 2024). Analysis in Section 5.3 also indicates that the granularity of swapping's invariants has a large numerical impact on the PLB ε (through the largest stratum size b), while the swap rate has comparatively little influence. This suggests that a reduction of invariants may have a larger impact on SDC protection compared to an increase in the swap rate—at least when protection is measured by a DP flavor.

Finally, crude comparisons can be made between swapping and the TDA based on their vulnerability to reconstruction attacks. From experiments in Abowd et al. (2025), swapping with a high swap rate of 50% is roughly comparable in its protection against reconstruction attacks as the TDA (using the 2020 production settings). This suggests that the reduction in invariants from the 2010 data swapping method's invariants (which are numerous) to the TDA's invariants (which are few) is approximately equivalent to an increase in PLB from $\varepsilon = 15.19$ at the household level to $\varepsilon = 52.83$ (with $\delta = 10^{-10}$) at the person level.

6.2. **Mitigating the Impact of Invariants on Disclosure Risk.** Because some small number of invariants is frequently mandated, while a large number may have an adverse impact on disclosure

risk, statistical agencies need methodologies that allow for the specification of invariants in a flexible and precise manner. To this end, swapping—as instantiated either in 2010 or in our work—does not suffice because its invariants are largely hardwired into its mechanics. However, several extensions to data swapping enable more customization in the choice of invariants, thereby allowing for a better balance between SDC protection and accuracy targets. (On this topic, it is also worth noting that the TDA allows for a range of invariant choices.)

One such extension is *probabilistic unit matching*, which was considered by the USCB as part of its comparative analysis between data swapping and the TDA. Instead of using the swapping key to form hard strata within which swapping is confined, this method permits units across different strata to be swapped with a small probability, which could be inversely proportional to some distance metric on the strata. For example, consider using county as the swapping variable, with state and household size as the matching variables. Suppose that, for some $\alpha > 0$, a household chosen for swapping would have a $(1 - \alpha)\%$ chance of being swapped with another household of the same size, but an $\alpha\%$ chance of being swapped with a differently sized household. Doing so retains the countywide household counts as invariant, but the countywide total populations would no longer be invariant.

Two other approaches to remove some of swapping's invariants are *pre-* and *post-swap perturbation*. As their names suggest, the former infuses noise into the confidential records prior to applying swapping (Hawes & Rodríguez, 2021, p. 23), whereas the latter perturbs an intermediate data product after applying swapping. Notably, data swapping followed by tabular perturbation is a common SDC strategy; for example, it is the approach taken by the Office of National Statistics (ONS) for the protection of the 2021 UK Census (Office for National Statistics, 2023). In this case, the cell key method (CKM) (Fraser & Wooton, 2005; Marley & Leaver, 2011; Thompson et al., 2013) is employed to perturb the cells of contingency tables after targeted swapping has been applied to the underlying microdata. In addition to its use at ONS, applying swapping and then CKM perturbation is also recommended by Eurostat's Centre of Excellence on Statistical Disclosure Control for EU censuses (Glessing & Schulte Nordholt, 2017).

That the CKM has already been analyzed through the lens of DP (Bailie & Chien, 2019; Chien & Sadeghi, 2024; Rinott et al., 2018) suggests the possibility that its use as a post-swap perturbation method may deliver a formal guarantee of protection that is stronger than that provided by swapping or CKM alone. Nevertheless, we leave to future work a full investigation of the protections supplied by probabilistic matching and by swapping combined with pre- or post-swap perturbation. Note that compared to standard swapping algorithms such as the PSA, all three of the above procedures introduce strictly more auxiliary randomness into the data product. It would therefore be reasonable to expect the resulting algorithms to enjoy DP guarantees while supplying fewer and more flexible choices of invariants. One salient question for this line of research is to determine the DP specification for the 'chaining' (i.e., sequential composition) of two mechanisms (e.g., swapping followed by tabular perturbation), when these mechanisms satisfy different DP specifications.

Disclosure Statement. JB gratefully acknowledges partial financial support from the Australian-American Fulbright Commission and the Kinghorn Foundation; RG and XLM acknowledge partial financial support from the NSF; and XLM acknowledges partial financial support from Harvard University's Office of the Vice Provost for Research.

Acknowledgments. We thank Cory McCartan for his assistance with the 2010 U.S. Census data, as well as for many stimulating discussions; and Xiaodong Yang, Nathan Cheng, and Souhardya Sengupta for their help in proving Lemma 5. We are grateful to the participants of the National Bureau of Economic Research's conference *Data Privacy Protection and the Conduct of Applied Research: Methods, Approaches and Their Consequences* (May 4-5, 2023); the 36th New England Statistical Symposium's invited session *A Private Refreshment on Statistical Principles and Senses* (June 6, 2023); the 2023 Joint Statistical Meetings session *Methodological Approaches to Privacy Concerns Across Multiple Domains* (August 10, 2023); the CA Census retreat at the Boston University Center for Computing and Data Science (September 26, 2023); and the Statistics Canada Methodology Seminar (October 31, 2023) for their thoughtful comments and questions. We appreciate enormously the detailed feedback provided by Daniel Kifer, John Abowd, Philip Leclerc, Ryan Cummings, Rolando Rodriguez, Robert Ashmead, Sallie Keller, and Michael Hawes at the USCB, which greatly improved and corrected all three parts of this trio of papers. All remaining errors are purely our own.

References

Abowd, J., Ashmead, R., Cumings-Menon, R., Garfinkel, S., Heineck, M., Heiss, C., Johns, R., Kifer, D., Leclerc, P., Machanavajjhala, A., Moran, B., Sexton, W., Spence, M., & Zhuravlev, P. (2022). The 2020 Census disclosure avoidance system TopDown Algorithm. *Harvard Data Science Review*, (Special Issue 2). https://doi.org/10.1162/99608f92.529e3cb9

Abowd, J. M. (2017). How will statistical agencies operate when all data are private? *Journal of Privacy and Confidentiality*, *7*(3). https://doi.org/10.29012/jpc.v7i3.404

Abowd, J. M. (2018). The U.S. Census Bureau adopts differential privacy. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '18* (pp. 2867–2867). ACM Press. https://doi.org/10.1145/3219819.3226070

Abowd, J. M. (2021). Declaration of John M. Abowd [Exhibit A, Document 10-1, Case 1:21-cv-01361-ABJ of the United States District Court for the District of Columbia]

Abowd, J. M., Adams, T., Ashmead, R., Darais, D., Dey, S., Garfinkel, S., Goldschlag, N., Hawes, M. B., Kifer, D., Leclerc, P., Lew, E., Moore, S., Rodríguez, R. A., Tadros, R. N., & Vilhuber, L. (2025). A simulated reconstruction and reidentification attack on the 2010 U.S. Census. *Harvard Data Science Review*, *7*(3). https://doi.org/10.1162/99608f92.4a1ebf70

Abowd, J. M., & Hawes, M. B. (2023). Confidentiality protection in the 2020 US Census of Population and Housing. *Annual Review of Statistics and Its Application*, *10*(1), 119–144. https://doi.org/10.1146/annurev-statistics-010422-034226

Ashmead, R., Kifer, D., Leclerc, P., Machanavajjhala, A., & Sexton, W. (2019). *Effective privacy after adjusting for invariants with applications to the 2020 Census* (tech. rep.). GitHub. https://github.com/uscensusbureau/census2020-das-e2e/blob/master/doc/20190711_0941_Effective_Privacy_after_Adjusting_for_Constraints__With_applications_to_the_2020_Census.pdf

Australian Bureau of Statistics. (2021). Treating microdata. https://www.abs.gov.au/about/data-services/data-confidentiality-guide/treating-microdata

Bailie, J. (2020). Big data, differential privacy and national statistical organisations. *Statistical Journal of the IAOS*, *36*(4), 1067–1074. https://doi.org/10.3233/SJI-200685

Bailie, J., & Chien, C.-H. (2019). ABS perturbation methodology through the lens of differential privacy. In *Work Session on Statistical Data Confidentiality* (p. 13). UN Economic Commission for Europe. https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2019/mtg1/SDC2019_S2_ABS_Bailie_D.pdf

Bailie, J., & Drechsler, J. (2024). Whose data is it anyway? Towards a formal treatment of differential privacy for surveys. In V. J. Hotz, R. Gong, & I. M. Schmutte (Eds.), *Data privacy protection and the conduct of applied research: Methods, approaches and their consequences* (p. 33). National Bureau of Economic Research. https://conference.nber.org/conf_papers/f194306.pdf

Bailie, J., Gong, R., & Meng, X.-L. (2026a). *Differential privacy meets invariant statistics: Some conundrums in quantifying trade-offs*. ArXiv. https://doi.org/10.48550/arXiv.2504.15246

Bailie, J., Gong, R., & Meng, X.-L. (2026b). *Privacy differentials in differential privacy* [Article in preparation]. Department of Statistics, Harvard University.

Ballesteros, M., Dwork, C., King, G., Olson, C., & Raghavan, M. (2025). Evaluating the impacts of swapping on the US Decennial Census. In *CSLAW '25: Proceedings of the 2025 Symposium on Computer Science and Law* (pp. 64–76). Association for Computing Machinery. https://doi.org/10.1145/3709025.3712210

Benthall, S., & Cummings, R. (2024). *Integrating differential privacy and contextual integrity*. ArXiv. https://doi.org/10.48550/arXiv.2401.15774

boyd, d., & Sarathy, J. (2022). Differential perspectives: Epistemic disconnects surrounding the U.S. Census Bureau's use of differential privacy. *Harvard Data Science Review*, (Special Issue 2). https://doi.org/10.1162/99608f92.66882f0e

Bun, M., & Steinke, T. (2016). Concentrated differential privacy: Simplifications, extensions, and lower bounds. In M. Hirt & A. Smith (Eds.), *Theory of cryptography* (pp. 635–658). Springer. https://doi.org/10.1007/978-3-662-53641-4_24

Canonne, C., Kamath, G., & Steinke, T. (2022). The discrete Gaussian for differential privacy. *Journal of Privacy and Confidentiality*, *12*(1). https://doi.org/10.29012/jpc.784

Chien, C.-H., & Sadeghi, P. (2024). On the connection between the ABS perturbation methodology and differential privacy. *Journal of Privacy and Confidentiality*, *14*(2). https://doi.org/10.29012/jpc.859

Cho, Y. H., & Awan, J. (2024). *Formal privacy guarantees with invariant statistics*. ArXiv. https://doi.org/10.48550/arXiv.2410.17468

Christ, M., Radway, S., & Bellovin, S. M. (2022). Differential privacy and swapping: Examining de-identification's impact on minority representation and privacy preservation in the U.S. Census. In *2022 IEEE Symposium on Security and Privacy* (pp. 457–472). Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/SP46214.2022.9833668

Cohen, A. (2022). Attacks on deidentification's defenses. In *Proceedings of the 31st USENIX Security Symposium (USENIX Security 22)* (pp. 1469–1486). USENIX. https://www.usenix.org/conference/usenixsecurity22/presentation/cohen

Cumings-Menon, R., Ashmead, R., Kifer, D., Leclerc, P., Spence, M., Zhuravlev, P., & Abowd, J. M. (2025). Disclosure avoidance for the 2020 Census Demographic and Housing Characteristics File. *Harvard Data Science Review*, *7*(3). https://doi.org/10.1162/99608f92.f1065159

Dalenius, T., & Reiss, S. P. (1978). Data-swapping: A technique for disclosure control (extended abstract). In *Proceedings of the ASA Section on Survey Research Methods* (pp. 191–194,

Vol. 6). American Statistical Association. http://www.asasrms.org/Proceedings/papers/1978_038.pdf

Dalenius, T., & Reiss, S. P. (1982). Data-swapping: A technique for disclosure control. *Journal of Statistical Planning and Inference*, *6*(1), 73–85. https://doi.org/10.1016/0378-3758(82)90058-1

de Vries, M., Golmajer, M., Tent, R., Giessing, S., & de Wolf, P.-P. (2023). An overview of used methods to protect the European Census 2021 tables. In *UNECE Conference of European Statisticians Expert Meeting on Statistical Data Confidentiality* (p. 16). UN Economic Commission for Europe. https://unece.org/statistics/documents/2023/08/working-documents/overview-used-methods-protect-european-census-2021

DePersio, M., Lemons, M., Ramanayake, K. A., Tsay, J., & Zayatz, L. (2012). n-cycle swapping for the American Community Survey. In J. Domingo-Ferrer & I. Tinnirello (Eds.), *PSD 2012: Privacy in statistical databases* (pp. 143–164, Vol. 7556). Springer. https://doi.org/10.1007/978-3-642-33627-0_12

Desfontaines, D., Mohammadi, E., Krahmer, E., & Basin, D. (2020). *Differential privacy with partial knowledge*. ArXiv. https://doi.org/10.48550/arXiv.1905.00650

Desfontaines, D., & Pejó, B. (2020). SoK: Differential privacies. In *Proceedings on Privacy Enhancing Technologies* (pp. 288–313, Vol. 2020). https://doi.org/10.2478/popets-2020-0028

Dharangutte, P., Gao, J., Gong, R., & Yu, F.-Y. (20237). Integer subspace differential privacy. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence* (pp. 7349–7357, Vol. 37). AAAI Press. https://doi.org/10.1609/aaai.v37i6.25895

Dobra, A., & Fienberg, S. E. (2000). Bounds for cell entries in contingency tables given marginal totals and decomposable graphs. *Proceedings of the National Academy of Sciences*, *97*(22), 11885–11892. https://doi.org/10.1073/pnas.97.22.11885

Drechsler, J., & Reiter, J. P. (2010). Sampling with synthesis: A new approach for releasing public use census microdata. *Journal of the American Statistical Association*, *105*(492), 1347–1357. https://doi.org/10.1198/jasa.2010.ap09480

Duncan, G. T., & Lambert, D. (1986). Disclosure-limited data dissemination. *Journal of the American Statistical Association*, *81*(393), 10–18. https://doi.org/10.1080/01621459.1986.10478229

Dwork, C., Kohli, N., & Mulligan, D. (2019). Differential privacy in practice: Expose your epsilons! *Journal of Privacy and Confidentiality*, *9*(2). https://doi.org/10.29012/jpc.689

Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In S. Halevi & T. Rabin (Eds.), *Proceedings of the Third Theory of Cryptography Conference, TCC 2006* (pp. 265–284, Vol. 3876). Springer. https://doi.org/10.1007/11681878_14

Dwork, C., & Naor, M. (2010). On the difficulties of disclosure prevention in statistical databases or the case for differential privacy. *Journal of Privacy and Confidentiality*, *2*(1). https://doi.org/10.29012/jpc.v2i1.585

Dwork, C., Smith, A., Steinke, T., & Ullman, J. (2017). Exposed! A survey of attacks on private data. *Annual Review of Statistics and Its Application*, *4*(1), 61–84. https://doi.org/10.1146/annurev-statistics-060116-054123

Fienberg, S., & McIntyre, J. (2004). Data swapping: Variations on a theme by Dalenius and Reiss. In J. Domingo-Ferrer & V. Torra (Eds.), *Privacy in Statistical Databases: PSD 2004 Proceedings*. Springer. https://doi.org/10.1007/978-3-540-25955-8_2

Francis, P. (2022). *A note on the misinterpretation of the US Census re-identification attack*. ArXiv. https://doi.org/10.48550/arXiv.2202.04872

Fraser, B., & Wooton, J. (2005). A proposed method for confidentialising tabular output to protect against differencing. In *Joint UNECE/Eurostat work session on statistical data confidentiality* (p. 6). UN Economic Commission for Europe. https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2005/wp.35.e.pdf

Gao, J., Gong, R., & Yu, F.-Y. (2022). Subspace differential privacy. *Proceedings of the AAAI Conference on Artificial Intelligence*, *36*(4), 3986–3995. https://doi.org/10.1609/aaai.v36i4.20315

Garfinkel, S. L. (2019). *Formal privacy: Making an impact at large organizations – deploying differential privacy for the 2020 Census of Population and Housing* [Slides]. U.S. Census Bureau. https://www.census.gov/content/dam/Census/newsroom/press-kits/2019/jsm/presentation-deploying-differential-privacy-for-the-2020-census-of-pop-and-housing.pdf

Glessing, S., & Schulte Nordholt, E. (2017). *Recommendations for best practices to protect the Census 2021 hypercubes*. Centre of Excellence on Statistical Disclosure Control, Eurostat. https://wayback.archive-it.org/12090/20181005014459/https://ec.europa.eu/eurostat/cros/system/files/recommendations_for_the_protection_of_hypercubes.pdf

Gong, R., & Meng, X.-L. (2020). Congenial differential privacy under mandated disclosure. In *Proceedings of the 2020 ACM-IMS on Foundations of Data Science Conference* (pp. 59–70). Association for Computing Machinery. https://doi.org/10.1145/3412815.3416892

Hawala, S. (2008). Producing partially synthetic data to avoid disclosure. In *JSM section on government statistics* (pp. 1345–1350). American Statistical Association. http://www.asasrms.org/Proceedings/y2008/Files/301018.pdf

Hawes, M. (2021a). Understanding the 2020 Census Disclosure Avoidance System [Slides]. https://www2.census.gov/about/training-workshops/2021/2021-08-10-das-presentation.pdf

Hawes, M. (2021b). The 2020 Census Disclosure Avoidance System [Slides]. https://planning.maryland.gov/MSDC/Documents/affiliate_meeting/2021/Census2021_MHawes.pdf

Hawes, M., & Rodríguez, R. (2021). Determining the privacy-loss budget: Research into alternatives to differential privacy [Presentation to the Census Scientific Advisory Committee]. https://www2.census.gov/about/partners/cac/sac/meetings/2021-05/presentation-research-on-alternatives-to-differential-privacy.pdf

Hawes, M., Rodriguez, R., & Goldschlag, N. (2021). The Census Bureau's simulated reconstruction-abetted re-identification attack on the 2010 Census [Transcript of presentation at NWX-U.S. Dept Of Commerce]. https://www2.census.gov/about/training-workshops/2021/2021-05-07-das-transcript.pdf

He, X., Machanavajjhala, A., & Ding, B. (2014). Blowfish privacy: Tuning privacy-utility trade-offs using policies. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data – SIGMOD '14* (pp. 1447–1458). Association for Computing Machinery. https://doi.org/10.1145/2588555.2588581

Hotz, V. J., & Salvo, J. (2020). *Assessing the use of differential privacy for the 2020 Census: Summary of what we learned from the CNSTAT workshop*. National Academies Committee on

National Statistics. https://www.amstat.org/asa/files/pdfs/POL-CNSTAT_CensusDP_WorkshopLessonsLearnedSummary.pdf

Ito, S., & Hoshino, N. (2014). Data swapping as a more efficient tool to create anonymized Census microdata in Japan [Paper presentation]. In *Privacy in Statistical Databases 2014 (PSD 2014)* (p. 14). UNESCO Chair in Data Privacy. https://www.nstac.go.jp/sys/files/static/services/society_paper/26_04_01_Paper.pdf

Kairouz, P., Liu, Z., & Steinke, T. (2021). The distributed discrete Gaussian mechanism for federated learning with secure aggregation. In M. Meila & T. Zhang (Eds.), *Proceedings of the 38th International Conference on Machine Learning* (pp. 5201–5212). PMLR. https://proceedings.mlr.press/v139/kairouz21a.html

Kazan, Z., & Reiter, J. P. (2025). Assessing statistical disclosure risk for differentially private, hierarchical count data, with application to the 2020 U. S. Decennial Census. *Statistica Sinica*, *35*, 629–649. https://doi.org/10.5705/ss.202022.0187

Kenny, C. T., Kuriwaki, S., McCartan, C., Rosenman, E. T., Simko, T., & Imai, K. (2021). The use of differential privacy for census data and its impact on redistricting: The case of the 2020 U.S. Census. *Science Advances*, *7*(41), eabk3283. https://doi.org/10.1126/sciadv.abk3283

Keyes, O., & Flaxman, A. D. (2022). How Census data put trans children at risk. *Scientific American*. https://www.scientificamerican.com/article/how-census-data-put-trans-children-at-risk/

Kifer, D. (2019). Design principles of the TopDown algorithm [Presentation]. JASON, La Jolla, CA, United States.

Kifer, D., Abowd, J. M., Ashmead, R., Cumings-Menon, R., Leclerc, P., Machanavajjhala, A., Sexton, W., & Zhuravlev, P. (2022). *Bayesian and frequentist semantics for common variations of differential privacy: Applications to the 2020 Census*. ArXiv. https://doi.org/10.48550/arXiv.2209.03310

Kifer, D., & Machanavajjhala, A. (2011). No free lunch in data privacy. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data* (pp. 193–204). ACM Press. https://doi.org/10.1145/1989323.1989345

Kifer, D., & Machanavajjhala, A. (2014). Pufferfish: A framework for mathematical privacy definitions. *ACM Transactions on Database Systems*, *39*(1), 1–36. https://doi.org/10.1145/2514689

Kim, N. (2015). The effect of data swapping on analyses of American Community Survey data. *Journal of Privacy and Confidentiality*, *7*(1). https://doi.org/10.29012/jpc.v7i1.644

Lauger, A., Wisniewski, B., & McKenna, L. (2014). *Disclosure avoidance techniques at the U.S. Census Bureau: Current practices and research* (Research Report Series – Disclosure Avoidance No. #2014-02). Center for Disclosure Avoidance Research, U.S. Census Bureau. https://www.census.gov/content/dam/Census/library/working-papers/2014/adrm/cdar2014-02-discl-avoid-techniques.pdf

Lemons, M., Dajani, A., You, J., & Jordan, J. (2015). Measuring the degree of difference in perturbed data. In *Proceedings of the 2015 Joint Statistical Meetings*. American Statistical Association. https://ww2.amstat.org/meetings/jsm/2015/onlineprogram/AbstractDetails.cfm?abstractid=316960

Machanavajjhala, A., Kifer, D., Gehrke, J., & Venkitasubramaniam, M. (2007). ℓ-Diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data*, *1*(1). https://doi.org/10.1145/1217299.1217302

Marley, J. K., & Leaver, V. L. (2011). A method for confidentialising user-defined tables: Statistical properties and a risk-utility analysis. In *Proceedings of the 58th World Statistical Congress* (pp. 1072–1081). International Statistical Institute. https://2011.isiproceedings.org/papers/450007.pdf

McClure, D., & Reiter, J. P. (2012). Differential privacy and statistical disclosure risk measures: An investigation with binary synthetic data. *Transactions on Data Privacy*, *5*(3), 535–552. https://doi.org/10.5555/2423656.2423658

McKenna, L. (2018). *Disclosure avoidance techniques used for the 1970 through 2010 Decennial Censuses of Population and Housing* (working paper). The Research and Methodology Directorate, U.S. Census Bureau. https://www.census.gov/content/dam/Census/library/working-papers/2018/adrm/Disclosure%20Avoidance%20for%20the%201970-2010%20Censuses.pdf

McKenna, L., & Haubach, M. (2019). *Legacy techniques and current research in disclosure avoidance at the U.S. Census Bureau* (working paper). Research and Methodology Directorate, U.S. Census Bureau. https://www.census.gov/content/dam/Census/library/working-papers/2019/adrm/5%20Legacy%20Techniques(tagged)%20CED-DA%20Report%20Series.pdf

Mitra, R., & Reiter, J. P. (2006). Adjusting survey weights when altering identifying design variables via synthetic data. In *Privacy in Statistical Databases: PSD 2006 Proceedings* (pp. 177–188). Springer. https://doi.org/10.1007/11930242_16

Neunhoeffer, M., Seeman, J., & Drechsler, J. (2025). On the formal privacy guarantees of synthetic data (generated without formal privacy guarantees). *Harvard Data Science Review*. https://doi.org/10.1162/99608f92.1af82b35

Nissenbaum, H. F. (2010). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford Law Books

Office for National Statistics. (2023). Protecting personal data in Census 2021 results. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/protectingpersonaldataincensus2021results

Population Reference Bureau & U.S. Census Bureau's 2020 Census Data Products and Dissemination Team. (2023). *Disclosure avoidance and the 2020 Census: How the TopDown algorithm works* (2020 Census Briefs No. C2020BR-04). https://www2.census.gov/library/publications/decennial/2020/census-briefs/c2020br-04.pdf

Protivash, P., Durrell, J., Kifer, D., Ding, Z., & Zhang, D. (2024). Reconstruction attacks on aggressive relaxations of differential privacy. *Journal of Privacy and Confidentiality*, *14*(3). https://doi.org/10.29012/jpc.871

Radway, S., & Christ, M. (2023). *The impact of de-identification on single-year-of-age counts in the U.S. Census*. ArXiv. https://doi.org/10.48550/arXiv.2308.12876

Reiter, J. P. (2005). Estimating risks of identification disclosure in microdata. *Journal of the American Statistical Association*, *100*(472), 1103–1112. https://doi.org/10.1198/016214505000000619

Rinott, Y., O'Keefe, C. M., Shlomo, N., & Skinner, C. (2018). Confidentiality and differential privacy in the dissemination of frequency tables. *Statistical Science*, *33*(3), 358–385. https://doi.org/10.1214/17-STS641

Robertson Ishii, T., & Atkins, P. (2023). Essential vs. accidental properties. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Spring 2023 Edition). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2023/entries/essential-accidental/

Ruggles, S., Fitch, C. A., Goeken, R., Hacker, J. D., Nelson, M. A., Roberts, E., Schouweiler, M., & Sobek, M. (2021). IPUMS ancestry full count data: Version 3.0 [data set]. https://doi.org/10.18128/D014.V3.0

Seeman, J., Reimherr, M., & Slavkovic, A. (2022). *Formal privacy for partially private data*. ArXiv. https://doi.org/10.48550/arXiv.2204.01102

Seeman, J., & Susser, D. (2024). Between privacy and utility: On differential privacy in theory and practice. *ACM Journal on Responsible Computing*, *1*(1), 3:1–18. https://doi.org/10.1145/3626494

Shlomo, N., Tudor, C., & Groom, P. (2010). Data swapping for protecting Census tables. In J. Domingo-Ferrer & E. Magkos (Eds.), *Privacy in statistical databases* (pp. 41–51). Springer. https://doi.org/10.1007/978-3-642-15838-4_4

Slavković, A., & Seeman, J. (2023). Statistical data privacy: A song of privacy and utility. *Annual Review of Statistics and Its Application*, *10*(1), 189–218. https://doi.org/10.1146/annurev-statistics-033121-112921

Slavković, A. B., & Lee, J. (2010). Synthetic two-way contingency tables that preserve conditional frequencies. *Statistical Methodology*, *7*(3), 225–239. https://doi.org/10.1016/j.stamet.2009.11.002

Spicer, K. (2020). *Statistical disclosure control (SDC) for 2021 UK Census* (tech. rep. No. EAP125). UK Statistics Authority. https://uksa.statisticsauthority.gov.uk/wp-content/uploads/2020/07/EAP125-Statistical-Disclosure-Control-SDC-for-2021-UK-Census.docx

Steed, R., Qing, D., & Wu, S. (2025). Quantifying privacy risks of public statistics to residents of subsidized housing. *Harvard Data Science Review*, (Special Issue 6). https://doi.org/10.1162/99608f92.39d8bfa4

Steel, P., & Zayatz, L. (2003). *The effects of the disclosure limitation procedure on Census 2000 tabular data products (abridged)* (tech. rep. No. Census 2000 Evaluation C.1). Statistical Research Division, U.S. Census Bureau. https://www2.census.gov/programs-surveys/decennial/2000/program-management/5-review/txe-program/C_1.pdf

Sweeney, L. (2000). *Simple demographics often identify people uniquely* (Data Privacy Working Paper No. 3). Carnegie Mellon University. http://dataprivacylab.org/projects/identifiability/

Sweeney, L. (2002). *k*-Anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *10*(5), 557–570. https://doi.org/10.1142/S0218488502001648

Thompson, G., Broadfoot, S., & Elazar, D. (2013). Methodology for the automatic confidentialisation of statistical outputs from remote servers at the Australian Bureau of Statistics [Paper presentation]. In *Joint UNECE/Eurostat work session on statistical data confidentiality* (p. 38). UN Economic Commission for Europe. https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2013/Topic_1_ABS.pdf

Tumult Labs. (2022). *SafeTab: DP algorithms for 2020 Census Detailed DHC Race & Ethnicity*. https://www2.census.gov/about/partners/cac/sac/meetings/2022-03/dhc-attachment-1-safetab-dp-algorithms.pdf

UK Statistics Authority. (2021). *Transparency of SDC methods and parameters (post-meeting EAP paper for SDC for Census August 2021)* (tech. rep. No. EAP168). https://uksa.statisticsauthority.gov.uk/wp-content/uploads/2022/02/EAP168-Statistical-Disclosure-Control-for-Census.pdf

U.S. Census Bureau. (2010). *2010 Sample Census Form*. https://www.census.gov/history/pdf/2010questionnaire.pdf

U.S. Census Bureau. (2012). *2010 Census Summary File 1—Technical documentation* (tech. rep. No. SF1/10-4 (RV)). https://www2.census.gov/programs-surveys/decennial/2010/technical-documentation/complete-tech-docs/summary-file/sf1.pdf

U.S. Census Bureau. (2016). *Decennial Census Surname Files (2010, 2000)*. https://www.census.gov/data/developers/data-sets/surnames.html

U.S. Census Bureau. (2021a). *2020 Census National Redistricting Data Summary File*. https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/summary-file/2020Census_PL94_171Redistricting_NationalTechDoc.pdf

U.S. Census Bureau. (2021b). *2020 Census State Redistricting Data Summary File*. https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/summary-file/2020Census_PL94_171Redistricting_StatesTechDoc_English.pdf

U.S. Census Bureau. (2021c). *Comparing differential privacy with older disclosure avoidance methods* (Factsheet No. D-FS-GP-EN-0509). https://www.census.gov/content/dam/Census/library/factsheets/2021/comparing-differential-privacy-with-older-disclosure-avoidance-methods.pdf

U.S. Census Bureau. (2021d). *Guidance for geography users: Hierarchy diagrams*. https://www.census.gov/programs-surveys/geography/guidance/hierarchy.html

U.S. Census Bureau. (2021e). *Disclosure avoidance for the 2020 Census: An introduction* (tech. rep.). U.S. Government Publishing Office. https://www2.census.gov/library/publications/decennial/2020/2020-census-disclosure-avoidance-handbook.pdf

U.S. Census Bureau. (2022). *2010 Demonstration Data Product – Demographic and Housing Characteristics technical document* (tech. rep.). https://www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/2010-demonstration-data-products/02-Demographic_and_Housing_Characteristics/2022-03-16_Summary_File/2022-03-16_Technical%20Document/2022-03-16_Technical%20Document.pdf

U.S. Census Bureau. (2023a). *2020 Census Demographic and Housing Characteristics File (DHC) technical documentation* (tech. rep.). https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/demographic-and-housing-characteristics-file-and-demographic-profile/2020census-demographic-and-housing-characteristics-file-and-demographic-profile-techdoc.pdf

U.S. Census Bureau. (2023b). *2020 Census Demographic Profile*. https://www.census.gov/data/tables/2023/dec/2020-census-demographic-profile.html

U.S. Census Bureau. (2023c). *Disclosure avoidance and the 2020 Census: How the TopDown Algorithm works*. https://www2.census.gov/library/publications/decennial/2020/census-briefs/c2020br-04.pdf

U.S. Census Bureau. (2023d). *Release dates set for next 2020 Census data products; New reader-friendly disclosure avoidance briefs*. https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/newsletters/release-dates-and-da-briefs.html

U.S. Census Bureau. (2023e). *2020 Census Demographic and Housing Characteristics File (DHC)*. https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/demographic-and-housing-characteristics-file-and-demographic-

profile/2020census-demographic-and-housing-characteristics-file-and-demographic-profile-techdoc.pdf

U.S. Census Bureau. (2023f). *2023-04-03 privacy-loss budget allocations*. https://www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/2010-demonstration-data-products/04-Demonstration_Data_Products_Suite/2023-04-03/2023-04-03_Privacy-Loss_Budget_Allocations.pdf

U.S. Census Bureau. (2023g). *About 2020 Census data products*. https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/release/about-2020-data-products.html

U.S. Census Bureau. (2023h). *Factsheet on disclosure avoidance for the 2010 Demonstration Data Products Suite – Redistricting and Demographic and Housing Characteristics File – production settings (2023-04-03)*. https://www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/2010-demonstration-data-products/04-Demonstration_Data_Products_Suite/2023-04-03/2023-04-03_Factsheet.pdf

U.S. Census Bureau. (2023i). *2020 Census Disclosure Avoidance System detailed summary metrics*. https://www2.census.gov/programs-surveys/decennial/2020/data/demographic-and-housing-characteristics-file/2020-Census-Disclosure-Avoidance-System-Detailed-Summary-Metrics.xlsx

U.S. Census Bureau. (2023j). *2020 Census Redistricting Data (P.L. 94-171) Noisy Measurement File README File*. https://www2.census.gov/programs-surveys/decennial/2020/data/01-Redistricting_File--PL_94-171/00-2020-Redistricting-Noisy-Measurement-File/2020%20Redistricting%20NMF%202023-06-15%20README.html

U.S. Census Bureau. (2023k). *2020 Census 118th Congressional District Summary File (CD118)*. https://www.census.gov/data/tables/2023/dec/2020-census-CD118.html

U.S. Census Bureau. (2023l). *2020 Census Detailed Demographic and Housing Characteristics File A (Detailed DHC-A) technical documentation*. https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/detailed-demographic-and-housing-characteristics-file-a/2020census-detailed-dhc-a-techdoc.pdf

U.S. Census Bureau. (2023m). *2020 Census Demographic and Housing Characteristics (DHC) Demonstration Noisy Measurement File (2023-10-23) README file*. https://www2.census.gov/programs-surveys/decennial/2020/data/demographic-and-housing-characteristics-file/00-2020-DHC-Noisy-Measurement-File/2020_DHC_NMF_README.html

U.S. Census Bureau. (2023n). *Disclosure avoidance methods for the Detailed Demographic and Housing Characteristics File A (Detailed DHC-A): How SafeTab-P works*. https://www2.census.gov/library/publications/decennial/2020/census-briefs/c2020br-05.pdf

U.S. Census Bureau. (2023o). *Methodology, assumptions and inputs for the 2023 National Population Projections* (tech. rep.). https://www2.census.gov/programs-surveys/popproj/technical-documentation/methodology/methodstatement23.pdf

U.S. Census Bureau. (2023p). *Methodology for the United States Population Estimates: Vintage 2023* (tech. rep.). https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/2020-2023/methods-statement-v2023.pdf

U.S. Census Bureau. (2024a). *2020 Census Detailed Demographic and Housing Characteristics File B (Detailed DHC-B) technical documentation*. https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/detailed-demographic-and-housing-characteristics-file-b/2020census-detailed-dhc-b-techdoc.pdf

U.S. Census Bureau. (2024b). *Disclosure avoidance and the 2020 Census: How the SafeTab-H works.* https://www2.census.gov/library/publications/decennial/2020/census-briefs/c2020br-11.pdf

U.S. Census Bureau. (2024c). *2020 Census Privacy-Protected Microdata File (PPMF).* https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/privacy-protected-microdata-file/2020census-privacy-protected-microdata-file.pdf

U.S. Census Bureau. (2024d). *2020 Supplemental Demographic and Housing Characteristics File (S-DHC) technical documentation.* https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/supplemental-demographic-and-housing-characteristics-file/2020census-supplemental-dhc-techdoc.pdf

U.S. Census Bureau. (2024e). *Census Bureau releases final 2020 Census data product.* https://www.census.gov/newsroom/press-releases/2024/final-2020-census-data-product-s-dhc.html

U.S. Census Bureau. (2024f). *Disclosure avoidance and the Supplemental Demographic and Housing Characteristics File (S-DHC): How PHSafe works.* https://www2.census.gov/library/publications/decennial/2020/census-briefs/c2020br-12.pdf

Zayatz, L., Lucero, J., Massell, P., & Ramanayake, A. (2010). Disclosure avoidance for Census 2010 and American Community Survey five-year tabular data products. In *JSM Section on Survey Research Methods* (pp. 2279–2288). American Statistical Association. http://www.asasrms.org/Proceedings/y2010/Files/307156_57962.pdf

Zayatz, L. (2003). Disclosure limitation for Census 2000 tabular data. In *Joint ECE/Eurostat workshop on statistical data confidentiality* (p. 9). United Nations Economic Commission for Europe. https://unece.org/fileadmin/DAM/stats/documents/ece/ces/2003/04/confidentiality/wp.15.e.pdf

Zayatz, L. (2007). Disclosure avoidance practices and research at the U.S. Census Bureau: An update. *Journal of Official Statistics*, *23*(2), 253–265. https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/disclosure-avoidance-practices-and-research-at-the-u.s.-census-bureau-an-update.pdf

Appendix A. Other Related Work

In this appendix, we briefly review some related work that the main body of this article does not discuss in sufficient depth. Firstly, there is existing literature examining differential privacy (DP) under invariants. One branch of this literature develops DP mechanisms that report invariants without noise. In addition to the United States Census Bureau's work on the TopDown Algorithm (TDA) (Abowd et al., 2022), other papers in this branch include Gong and Meng (2020), Gao et al. (2022), and Dharangutte et al. (2023). Because invariants can be viewed from an attacker's perspective as background knowledge, work addressing how to incorporating this knowledge into DP (Desfontaines et al., 2020; He et al., 2014; Kifer & Machanavajjhala, 2011, 2014) is also relevant. In particular, Seeman et al. (2022) applies the Pufferfish privacy framework to construct a DP formulation that can handle invariants, although with the additional complication that the data must be modeled. Although not specifically addressing invariant-respecting DP, Protivash et al. (2024) demonstrate that related DP formulations—which, like invariants, also restrict the data universes considered by DP—may not provide sufficient SDC. More recent work on invariants can be found in Cho and Awan (2024) and Bailie (2020).

Many of these works, including Cho and Awan (2024), condition on the specific, realized value of the invariants, which we showed in Section 2 is not a valid way to incorporate invariants into DP. Further, those works above that incorporate invariants into DP only do so for specific flavors of DP, while in this work, we use a system that can integrate invariants into any DP flavor. This unified way to handle invariants, which is missing in other work, is needed for our comparisons of the Permutation Swapping Algorithm's and the TDA's DP specifications, as these specifications vary on other dimensions, not just on their invariants.

There is also related literature studying SDC for the U.S. Decennial Census. Ashmead et al. (2019) and Kifer et al. (2022) describe DP semantics for the 2020 Census, with the former focusing on the impact of invariants. Abowd et al. (2025) examines the 2010 DAS, using a reconstruction attack to demonstrate that aggregation did not provide SDC, as has traditionally been assumed. Christ et al. (2022) compares data swapping with standard DP-based mechanisms. Finally, the paper that first proposed data swapping (Dalenius & Reiss, 1982) includes arguments for the SDC provided by data swapping, which were reviewed by Fienberg and McIntyre (2004).

An extensive body of literature inspired and contributed to the development of the system of DP specifications outlined in Section 2.2. A review of this literature is given in Bailie et al. (2026b).

Appendix B. Background on Data Swapping

Invented by Dalenius and Reiss (1978, 1982) and further expanded upon by Fienberg and McIntyre (2004), data swapping (also called record swapping, particularly in Europe) refers to a family of statistical disclosure control (SDC) methods that select some subset of records and permute the values these records take for a subset of variables. These methods differ on which variables are swapped, how records are selected to be swapped, and how the interchanging of the values of the swapping variables between the selected records is conducted. (See DePersio et al., 2012; Fienberg and McIntyre, 2004; Kim, 2015; Shlomo et al., 2010, for examples of different data swapping methods.) Traditionally, claims of SDC protection provided by swapping methods have been based on the intuition that a successful disclosure requires linking inferred information about a *sensitive variable* to an individual entity using some *quasi-identifying variables*. By sensitive variable, we mean a variable that is plausibly of interest to an attacker—for example, a person's race or a household's income. Learning the value of a sensitive variable for an individual record may not be problematic on its own since the attacker does not know to whom the record belongs. Thus, an attacker has two goals: 1) to infer the value of a sensitive variable for an individual record and 2) to determine, using quasi-identifying variables, the individual entity associated with that record. Since the sensitive variable and the quasi-identifiers must belong to the same record, the attacker needs to infer them jointly. The idea behind data swapping is to hinder such joint inference by randomly permuting the records' quasi-identifiers while keeping the sensitive variables fixed (or visa versa). In this way, there are multiple plausible values for the original data set that are compatible with the swapped data set—thereby adding uncertainty to the relationship between any record's sensitive variables and its quasi-identifiers.

It is important to emphasize that the above discussion is only an intuitive justification for data swapping. A major motivation for this article is to supplement such intuitive arguments with mathematical SDC guarantees. Some such guarantees are provided by the Permutation Swapping Algorithm's differential privacy specification. In fact, Theorem 1 can be interpreted as a formalization of the above intuitive argument because it provides a bound on how plausible the true confidential

data set is compared to other compatible data sets. This bound ensures a degree of uncertainty in the relationship between $\boldsymbol{V}_{\text{Swap}}$ and $\boldsymbol{V}_{\text{Hold}} \setminus \boldsymbol{V}_{\text{Swap}}$. Taking $\boldsymbol{V}_{\text{Swap}}$ to be the quasi-identifiers and $\boldsymbol{V}_{\text{Hold}} \setminus \boldsymbol{V}_{\text{Swap}}$ the sensitive variables (or visa versa), this recovers the above argument. However, theoretically any set of variables can function as quasi-identifiers, depending on the attacker's auxiliary knowledge and the context of the data collection (see, e.g., Cohen, 2022; Machanavajjhala et al., 2007; Sweeney, 2000, 2002). As such, arguments that rely on knowing what variables are quasi-identifiers may have limited utility outside the scope of context-specific SDC analyses.

Data swapping is widely utilized—typically in combination with other SDC methods—by statistical offices across the globe. As we remark in the main body of this article, it has been used and studied extensively by the United States Census Bureau (Lauger et al., 2014; Lemons et al., 2015; McKenna & Haubach, 2019; Steel & Zayatz, 2003; Zayatz et al., 2010; Zayatz, 2007). Further, the Office of National Statistics (ONS) has employed it for the 2001, 2011, and 2021 UK Censuses (Office for National Statistics, 2023; Shlomo et al., 2010; Spicer, 2020). It is also one of the two protection methods recommended by Eurostat's Centre of Excellence on Statistical Disclosure Control (Glessing & Schulte Nordholt, 2017) and was used (or is intended to be used) for protecting census data by 15 of 30 European Union states surveyed by de Vries et al. (2023). The Australian Bureau of Statistics uses it as one of their primary SDC methods for releasing microdata (Australian Bureau of Statistics, 2021), and it has been explored as a method for protecting the Japanese Population Census (Ito & Hoshino, 2014).

While we largely focus on the swapping procedure used in the 2010 U.S. Decennial Census, much of this article applies to other statistical agencies, especially when their swapping mechanisms are similar to the Disclosure Avoidance System (DAS) used in 2010. In particularly, the ONS's Targeted Record Swapping (UK Statistics Authority, 2021) closely aligns with the 2010 DAS and hence the current article is also relevant for the 2021 UK Census.

Appendix C. The 2010 U.S. Census Disclosure Avoidance System

This appendix collates information about the 2010 Disclosure Avoidance System (DAS) that have been made public by the United States Census Bureau (USCB). Most of this information also applies to the 2000 DAS—as it was very similar to the 2010 DAS—but likely not to the 1990 DAS, which used a significantly different data swapping procedure (McKenna, 2018).

The main references are McKenna (2018), McKenna and Haubach (2019) and Abowd (2021), with additional information spread across various other USCB publications (Garfinkel, 2019; Hawala, 2008; Hawes, 2021a; Hawes et al., 2021; Lauger et al., 2014; Lemons et al., 2015; Steel & Zayatz, 2003; U.S. Census Bureau, 2021c, 2022; Zayatz et al., 2010; Zayatz, 2003, 2007). However, the publicly available documentation on the 2010 DAS is deliberately incomplete as some implementation details have been deemed confidential by the USCB due to concerns that they may allow the privacy protections of the 2010 DAS to be undermined. We are not the only researchers external to the USCB who have attempted to reproduce the 2010 DAS (Christ et al., 2022; Keyes & Flaxman, 2022; Kim, 2015; Radway & Christ, 2023); however, we believe this documentation is the most comprehensive of those that are currently publicly available.

The primary protection method of the 2010 DAS was data swapping. Special tabulations had additional rules-based protections (see Appendix A of McKenna, 2018 for these rules). Synthetic data methods were used to protect the confidentiality of group quarters (GQs) since swapping was infeasible for GQs due to their sparsity and the consequent lack of matching records (Hawala,

2008). These synthetic data methods involved replacing some GQ data with predicted values from a generalized linear model (McKenna, 2018, Section 6.5).

The data swapping procedure for the 2000 and 2010 DAS had three main steps:

Step 1: A random set S of household records was selected.
Step 2: Each record in S was paired with a similar, nearby household.
Step 3: The location of each household in S was swapped with the location of its pair.

We will describe each of these steps in detail below. The data swapping procedure was applied only to households (i.e., 'occupied housing units', U.S. Census Bureau, 2012) and not to unoccupied housing units or group quarters. At the end of step 3, the DAS swapping procedure outputs a data set —called the post-swapped data set—which differs from its input (the Census Edited File [CEF]) only on the locations of the selected households and their pairs. All publications from the 2010 Census were derived from this post-swapped data set (Lauger et al., 2014; Zayatz et al., 2010). (The post-swapped data set is called the Hundred-percent Detailed File by the USCB.)

Step 1: Household records were randomly selected into the set S with a probability that, for 2010, depended on (possibly among other factors):

(A) The size of the household's block (larger blocks decreased the probability of selection)
(B) Whether the household contained individuals of a race category not found elsewhere in its block (unique race categories increased the probability of selection)
(C) The imputation rate within the household's block (higher imputation rates decreased the probability of selection)
(D) Whether the household was unique within their geographical area on some set of variables (such households were always included in S). (It is not clear what geographical area was used, but we speculate that it may have been either the household's block group, tract, or county.)
(E) Whether the household record was imputed and what proportion of the record was imputed (Abowd, 2021; McKenna, 2018; McKenna & Haubach, 2019).

Note that (at least in 2000) the selection of records into S was not mutually independent. This was because the number of records in S was capped so that the proportion of swapped records (i.e., the swap rate) was controlled at prespecified thresholds at the state level (Steel & Zayatz, 2003). (The swap rates for each state were approximately equal (Steel & Zayatz, 2003).) There may have also been other dependencies between households' selection into S.

Exactly how a household's probability of selection was calculated is not public information. However, the USCB has confirmed that in 2010, the marginal selection probability (unconditional on other selections) was zero for totally imputed households, and was nonzero for all other households (Abowd, 2021; Hawes et al., 2021; McKenna, 2018).[2]

[2]However, this appears to be contradicted by another statement from the USCB: "there was a threshold value for not swapping in blocks with a high imputation rate" (McKenna & Haubach, 2019). Assuming that this imputation rate threshold was under 100%, there would be not-totally-imputed household records with zero selection probability.

There is a possible explanation of this contradiction. All not-totally-imputed households may have had the possibility of being swapped (in Step 2) even though some of them had zero probability of being selected into S. Yet this would require that, for all the not-totally-imputed households h with zero selection probability, there was a household h' with nonzero selection probability that matched h on the five criteria in Step 2 and furthermore that h and h' could possibly be matched given the DAS's prioritization of certain

Step 2: For each household record in S, the DAS swapping procedure found a household that

- had the same number of adults (over 18 years of age);
- had the same number of minors (under 18 years of age);[3]
- had the same tenure status;[4]
- was located in the same state; but
- was located in a different block (Abowd, 2021; Garfinkel, 2019; U.S. Census Bureau, 2021c).[5]

A household that satisfies these requirements is called a matching household. Each record in S was paired with a matching household. In 2000 (and hence plausibly in 2010 as well), the swapping procedure prioritized pairings where

(1) the matching record was also in S; or
(2) both records were geographically close (e.g., they were in the same tract or county); or
(3) the matching record had a high "disclosure risk" (Steel & Zayatz, 2003).

It is possible that there were other criteria for deciding the pairing when there were multiple matching households. It is unclear how these criteria were ranked in their importance. (For example, how did the swapping procedure decide between I. a pair where both records were in S but were in different counties; and II. a pair where both records were in the same block group but only one record was in S?) However, it is likely that criteria 1. was considered the most important since it minimizes the number of swaps (Steel & Zayatz, 2003).

Step 3: Steps 1 and 2 produced pairs of household records. These pairs consisted of one record from S along with its matching record found in Step 2 (which may also be in S). In Step 3, all pairs had their locations swapped. More exactly, for each household in S, the value of its block, block group, tract, and county were swapped with the corresponding values of its paired record. (Note that a pair of records might have had the same block group, tract, or county, in which case these values did not change. Garfinkel, 2019 states that the paired households were always in the same state, so this location variable was never swapped.)

C.1. **Comparing the 2010 DAS With the PSA.** In this section, we compare the 2010 data swapping procedure with the PSA. The PSA is a general algorithm (in the sense that its parameters—such as the swapping and matching variables—are not set but must be chosen). Thus, for the purposes of this comparison, we will consider the PSA using the implementation choices that attempt to mirror the 2010 DAS, as given in Section 3.5.

There are a number of key similarities between the data swapping procedure in the 2010 DAS and the PSA from Section 3.5:

---

matches over other matches (e.g., 1.–3. in Step 2). It seems infeasible to guarantee such requirements for all possible CEFs.

[3]As a consequence, the paired housing units also matched on the occupancy status (occupied versus unoccupied), and total number of persons.

[4]The 2010 Census classified households' tenure as either owner-occupied (owned outright), owner-occupied (with a loan or mortgage), renter occupied, or occupied without payment of rent. It is unclear if the swapping procedure matched households on these categories, or only on the broader categories of A. owner-occupied vs B. renter-occupied (including without payment of rent) (U.S. Census Bureau, 2010, 2012).

[5]The public documentation from the USCB is contradictory on whether there were additional requirements beyond the five listed here (Abowd, 2021; Hawes et al., 2021; U.S. Census Bureau, 2022).

(1) The *swapping units* (i.e., the records that are swapped) are household-records for both the 2010 DAS and the PSA.
(2) *Swap rates:* The swap rate is defined as the fraction of records that were swapped. For the 2010 DAS, the swap rate is the fraction of records that were selected into S or were paired with a record in S. This rate was explicitly controlled by the USCB at the state level and all states had approximately the same swap rate (Zayatz, 2003). Although the USCB has not released the value of the 2010 DAS's swap rate, at the national level it is purported to be between 2-4% (boyd & Sarathy, 2022).

    In comparison, the PSA controls the expected swap rate (where the expectation is over the randomness in the PSA). An implementer of the PSA cannot precisely fix the swap rate—but only the expected swap rate (via the parameter p). However, when the number of records n is large, the swap rate is typically very close to p, since its variance is approximately $p(1-p)/n \approx 0$.

    Hence, one may set the PSA's parameter p so that the swap rates for the PSA and the 2010 DAS are similar at the state and national levels.
(3) The *matching variables* of the 2010 DAS include the household's state, the number of adult occupants, the number of child occupants, and the household's tenure status. There may be other matching variables (which have not been disclosed by the USCB), but Abowd (2021) implicitly suggests that this is not the case. The PSA could be implemented with exactly the same matching variables. However, the matching variables of the PSA implementation in Section 3.5 are the household's state and counts of adults and children—the household's tenancy status was not included. By excluding a matching variable, the PSA from Section 3.5 has fewer invariants and its protection loss budget (PLB) ε is a conservative estimate, compared to a PSA implementation that mirrored exactly the 2010 DAS matching variables. (The reasoning here mirrors the discussion in Section 3.5 regarding the exclusion of the household count of voting age persons from the swap key.)
(4) The *swapping variables* of the 2010 DAS and the PSA from Section 3.5 are the same: the households' county, tract, block group, and block are swapped by the 2010 DAS and the PSA. (As we will discuss later in this section, the 2010 DAS sometimes used the households' county, tract, or block group as matching variables in an adaptive matching procedure. For our purposes, they can still be considered as swapping variables; matching variables can always be swapped since swapping them does not change the data.)

There are a number of significant differences between the PSA and the 2010 swapping procedure:

(1) The 2010 DAS *swapped* pairs of records, whereas the PSA *permutes* multiple records. While any permutation is equal to a sequence of multiple pairwise swaps, the 2010 DAS does not allow for such arbitrary swaps. However, permutation swapping (under the name n-Cycle swapping) was actively being investigated by the USCB (DePersio et al., 2012; Lauger et al., 2014) before this work was supplanted by their shift toward differential privacy (DP) (McKenna & Haubach, 2019). The USCB found that permutation swapping provided both better data utility and better data protection than the pairwise swapping used in 1990-2010 Censuses; their second finding is corroborated by our DP analysis of permutation swapping.
(2) *Swap probabilities:* The probability of a given household being swapped was not constant in the 2010 DAS. In fact, swapping was highly targeted to households that were "vulnerable to re-identification" (Hawes et al., 2021). Moreover, the probability of a household being

swapped was dependent on whether other households were selected for swapping (for example, because the absolute statewide swap rates were controlled). In comparison, in the PSA, the probability of a household being swapped is constant and independent of other households.

(3) *Adaptive matching:* The 2010 DAS paired households according to a complicated matching procedure. For example, they prioritized matching households that shared the same county or tract. (More details on their matching procedure is given in Step 2 of the 2010 DAS description above.) In essence, this means that sometimes the household's county or tract were included as matching variables, and sometimes they were not; and whether they were included was a function of the household as well as its matching households. The matching procedure for the PSA is much simpler by comparison: the matching variables are static and the choice of how to swap the selected matching households is made uniformly at random.

(4) *Nonvacuous swaps:* A swap is vacuous if it does not change the data set, except (possibly) by reordering the records. A pairwise swap is not vacuous if and only if the paired records have different values for both their swapping variables $\boldsymbol{V}_{\text{Swap}}$ and their holding variables $\boldsymbol{V}_{\text{Hold}}$. It is unclear whether the 2010 DAS prohibited vacuous swaps but we suspect so. On the other hand, vacuous swaps are allowed by the PSA.

C.2. **Modifying the PSA to Further Align With the 2010 DAS.** We discuss some possible extensions to the PSA in Section 6.2. Those extensions aimed to reduce the PSA's invariants without foregoing its DP guarantee. In this section, we propose four additional extensions to the PSA that also provide a DP guarantee while being more faithful to the 2010 DAS. These extensions address the differences between the 2010 DAS and the PSA identified in the previous section. We show that these differences can be bridged without losing the guarantee of DP—at the cost of greatly complicating the calculation of the PLB. We do not attempt these calculations; we only argue why the PLBs for these extensions remain bounded away from infinity.

First, we address one aspect of the 2010 DAS that cannot be incorporated into a DP swapping mechanism. The 2010 DAS used disjoint, pairwise swapping (Lauger et al., 2014). This is not a transitive action—its orbit space does not equal the universe induced by the swapping invariants—and hence it cannot satisfy differential privacy. (Roughly speaking, a necessary condition for a mechanism T to be pure DP is that $\mathsf{P}\big(T(\boldsymbol{x}, U) \in \cdot\ \big)$ and $\mathsf{P}\big(T(\boldsymbol{x}', U) \in \cdot\ \big)$ have common support for all $\boldsymbol{x}$ and $\boldsymbol{x}'$ in the same data universe with $d_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}') < \infty$. When the statistical disclosure control method is a random group action on $\boldsymbol{x}$, as is the case for permutation swapping, this condition is equivalent to the group action being transitive.)

Variable Swapping Probabilities: The PSA uses the same swapping probability p for all records. However, we can modify the PSA to use a different swapping probability p_i for each record i. As long as these probabilities are uniformly bounded away from zero and one (so that p_i^{-1} and $(1-p_i)^{-1}$ are bounded away from infinity), this modification will satisfy the same DP flavor as the original PSA (with a finite budget). The proof would follow the same strategy as in Appendix E; only the final computations would change, as one would need to optimize over $o_i = p_i/(1-p_i)$ for all i.

Nonuniform Permutations: The PSA samples derangements of the selected records uniformly at random, whereas the 2010 DAS prioritizes certain swaps over others. We can mirror this aspect of the 2010 DAS by sampling from a nonuniform distribution over the derangements. This would allow

for some derangements to be selected with higher probability than other derangements. The advantage here is that some derangements, which result in poor data utility (such as when geographically distant records are swapped), can be undersampled; while other, more desirable derangements can be oversampled. This would mimic the adaptive matching of the 2010 DAS. By reasoning that is analogous to the previous extension, this extension will also retain the PSA's DP flavor, provided that $\mathsf{P}_{\boldsymbol{x}}(\sigma = g)/\mathsf{P}_{\boldsymbol{x}'}(\sigma = g')$ is uniformly bounded by $\exp[O(|k_g - k_{g'}|)]$, for all derangements g and g' (of k_g and $k_{g'}$ records, respectively).

Prohibiting Imputed Records From Being Swapped: The 2010 DAS never swaps records that have been completely imputed. (The rationale is that imputed records do not require privacy protection.) We can modify the PSA so that $p_i = 0$ for all records i that are imputed. Suppose that the records that are imputed are constant. If the PSA satisfies $\varepsilon_{\mathcal{D}}$-DP$(\mathcal{X}_{\text{CEF}}, \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}, d^{hh}_{\text{HamS}}, D_{\text{Mult}})$, then this modification would satisfy $\varepsilon'_{\mathcal{D}}$-DP$(\mathcal{X}_{\text{CEF}}, \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}, d'_{\mathcal{X}}, D_{\text{Mult}})$, where

$$d'_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}') = \begin{cases} d^{hh}_{\text{HamS}}(\boldsymbol{x}, \boldsymbol{x}') & \text{if } \boldsymbol{x}, \boldsymbol{x}' \text{ do not differ on any imputed record,} \\ \infty & \text{otherwise'} \end{cases}$$

and $\varepsilon'_{\mathcal{D}} \leq \varepsilon_{\mathcal{D}}$ since the maximum stratum size b is reduced when the imputed records are removed.

Prohibiting Vacuous Swaps: A swap (or more generally a permutation) is vacuous if it does not change the data set, except perhaps by reordering the records.

We assume that the 2010 DAS does not allow vacuous swaps. We can similarly prohibit the PSA from allowing vacuous swaps. Instead of sampling derangements uniformly at random, we would put zero probability on vacuous derangements. Under the action of nonvacuous derangements, the orbit space is still the entire data universe. Hence, this modification will still satisfy the same DP flavor as the PSA, however, the calculation of the PLB $\varepsilon_{\mathcal{D}}$ will be difficult as one must optimize $\mathsf{P}_{\boldsymbol{x}}(\sigma = g)/\mathsf{P}_{\boldsymbol{x}'}(\sigma = g')$ over all permutations g and g' that are nonvacuous with respect to $\boldsymbol{x}$ and $\boldsymbol{x}'$, over all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}$.

Appendix D. Appendix to Section 2

D.1. An Additional Result on the Impact of Invariants.

Proposition 5. *Suppose that $\mathscr{D}'$ is refinement of $\mathscr{D}$. Then, for all PLBs $\varepsilon'_{\mathcal{D}'} : \mathscr{D}' \to [0, \infty]$, any mechanism that satisfies $\varepsilon_{\mathcal{D}}$-DP$(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, D_{\text{Pr}})$ also satisfies $\varepsilon'_{\mathcal{D}'}$-DP$(\mathcal{X}, \mathscr{D}', d_{\mathcal{X}}, D_{\text{Pr}})$, whenever $\varepsilon_{\mathcal{D}} \leq \inf\{\varepsilon'_{\mathcal{D}'} : \mathcal{D}' \in \mathscr{D}' \text{ with } \mathcal{D}' \subset \mathcal{D}\}$. Furthermore, for all PLBs $\varepsilon_{\mathcal{D}} : \mathscr{D} \to [0, \infty]$, any mechanism that satisfies $\varepsilon_{\mathcal{D}}$-DP$(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, D_{\text{Pr}})$ also satisfies $\varepsilon'_{\mathcal{D}'}$-DP$(\mathcal{X}, \mathscr{D}', d_{\mathcal{X}}, D_{\text{Pr}})$, whenever $\varepsilon'_{\mathcal{D}'} \geq \inf\{\varepsilon_{\mathcal{D}} : \mathcal{D} \in \mathscr{D} \text{ with } \mathcal{D}' \subset \mathcal{D}\}$.*

This result shows how protection loss budgets (PLBs) under one flavor of differential privacy (DP) can be transferred into PLBs under another flavor, when the two flavors are related by the fact that one's multiverse is refinement of the other (such as is the case when one has more invariants than the other). It will be used in the proofs of Proposition 3 and Theorem 2. Its second part is a stronger and more general result than Proposition 2.7 of Gao et al. (2022), which was restricted to nested linear subspaces and did not consider shrinking of the PLB.

D.2. Proofs for Section 2's Results.

Proof of Proposition 1. T is constant within any universe $\mathcal{D} \in \mathscr{D}_{\boldsymbol{c}}$. Therefore,

$$D_{\text{Pr}}[\mathsf{P}(T(\boldsymbol{x}, U) \in \cdot\ ), \mathsf{P}(T(\boldsymbol{x}', U) \in \cdot\ )] = 0,$$

for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}$. This proves the first half of the proposition.

To prove the second half, observe that $\mathsf{P}(T(\boldsymbol{x}, U) \in \cdot\,)$ and $\mathsf{P}(T(\boldsymbol{x}', U) \in \cdot\,)$ have disjoint support (the former's support is concentrated on $\boldsymbol{c}(\boldsymbol{x})$, the latter's on $\boldsymbol{c}(\boldsymbol{x}')$). Hence their total variation distance is one. This implies

$$D_{\mathrm{Pr}}[\mathsf{P}(T(\boldsymbol{x}, U) \in \cdot\,), \mathsf{P}(T(\boldsymbol{x}', U) \in \cdot\,)] = \infty,$$

Because $d_{\mathcal{X}}(\boldsymbol{x}_1, \boldsymbol{x}_2) < \infty$, the DP condition (Equation 2.1) can therefore only be satisfied if $\varepsilon_{\mathcal{D}_0} = \infty$. □

Proof of Proposition 2. This proposition relies on the metric axiom $D_{\mathrm{Pr}}(\mathsf{P}, \mathsf{Q}) > 0$ if $\mathsf{P} \neq \mathsf{Q}$. This implies

$$D_{\mathrm{Pr}}[\mathsf{P}(T(\boldsymbol{x}, U) \in \cdot\,), \mathsf{P}(T(\boldsymbol{x}', U) \in \cdot\,)] > 0.$$

Because $d_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}') < \infty$, the DP condition can therefore only be satisfied if $\varepsilon_{\mathcal{D}_0} > 0$. (Note that on the right hand side of Equation 2.1, we define $0 \times \infty$ to be ∞.) □

Lemma 1. *Given two DP specifications $\varepsilon_{\mathcal{D}}$-DP$(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, D_{\mathrm{Pr}})$ and $\varepsilon'_{\mathcal{D}'}$-DP$(\mathcal{X}, \mathscr{D}', d_{\mathcal{X}}, D_{\mathrm{Pr}})$, suppose that, for all $\mathcal{D}' \in \mathscr{D}'$ and all $\delta > 0$, there exists $\mathcal{D} \in \mathscr{D}$ such that $\mathcal{D}' \subset \mathcal{D}$ and $\varepsilon_{\mathcal{D}} \leq \varepsilon'_{\mathcal{D}'} + \delta$. Then*

$$\mathcal{M}(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, D_{\mathrm{Pr}}, \varepsilon_{\mathcal{D}}) \subset \mathcal{M}(\mathcal{X}, \mathscr{D}', d_{\mathcal{X}}, D_{\mathrm{Pr}}, \varepsilon'_{\mathcal{D}'}).$$

In the above lemma, $\mathcal{M}(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, D_{\mathrm{Pr}}, \varepsilon_{\mathcal{D}})$ denotes the set of mechanisms satisfying the DP specification $\varepsilon_{\mathcal{D}}$-DP$(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, D_{\mathrm{Pr}})$.

Proof. Suppose that $T \in \mathcal{M}(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, D_{\mathrm{Pr}}, \varepsilon_{\mathcal{D}})$. Let $\mathcal{D}' \in \mathscr{D}'$ and $\delta > 0$. Suppose $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}'$. By assumption, there exists $\mathcal{D} \in \mathscr{D}$ such that $\mathcal{D}' \subset \mathcal{D}$ and

$$D_{\mathrm{Pr}}[\mathsf{P}(T(\boldsymbol{x}, U) \in \cdot\,), \mathsf{P}(T(\boldsymbol{x}', U) \in \cdot\,)] \leq \varepsilon_{\mathcal{D}} d_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}') \leq (\varepsilon'_{\mathcal{D}'} + \delta) d_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}').$$

Since $D_{\mathrm{Pr}}[\mathsf{P}(T(\boldsymbol{x}, U) \in \cdot\,), \mathsf{P}(T(\boldsymbol{x}', U) \in \cdot\,)] \leq (\varepsilon'_{\mathcal{D}'} + \delta) d_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}')$ holds for all $\delta > 0$, it follows that $D_{\mathrm{Pr}}[\mathsf{P}(T(\boldsymbol{x}, U) \in \cdot\,), \mathsf{P}(T(\boldsymbol{x}', U) \in \cdot\,)] \leq \varepsilon'_{\mathcal{D}'} d_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}')$. This proves $T \in \mathcal{M}(\mathcal{X}, \mathscr{D}', d_{\mathcal{X}}, D_{\mathrm{Pr}}, \varepsilon'_{\mathcal{D}'})$. □

Proof of Proposition 5. Because $\mathscr{D}'$ is a refinement of $\mathscr{D}$, the assumption of Lemma 1 holds for both choices of the budgets $\varepsilon_{\mathcal{D}}$ and $\varepsilon'_{\mathcal{D}'}$. The results then follow by this lemma. □

Proof of Proposition 3. Suppose for contradiction that the result does not hold. Then there exists some $\mathcal{D}'_0 \in \mathscr{D}'$ and some $\mathcal{D}_0 \in \mathscr{D}$ such that $L'_{\mathcal{D}'_0} < L_{\mathcal{D}_0}$ and $\mathcal{D}_0 \subset \mathcal{D}'_0$.

Given a budget $\varepsilon'^{(i)}_{\mathcal{D}'} : \mathscr{D}' \to [0, \infty]$, define the budget $\varepsilon^{(i)}_{\mathcal{D}} : \mathscr{D} \to [0, \infty]$ as follows: For $\mathcal{D}_0$, define $\varepsilon^{(i)}_{\mathcal{D}_0} = \varepsilon'^{(i)}_{\mathcal{D}'_0}$. For all other $\mathcal{D}_1 \in \mathscr{D}$, fix a $\mathcal{D}'_1 \in \mathscr{D}'$ (which does not depend on i) with $\mathcal{D}' \supset \mathcal{D}$ and define $\varepsilon^{(i)}_{\mathcal{D}_1} = \varepsilon'^{(i)}_{\mathcal{D}'_1}$.

Now, if T satisfies $\varepsilon'^{(i)}_{\mathcal{D}'}$-DP$(\mathcal{X}, \mathscr{D}', d_{\mathcal{X}}, D_{\mathrm{Pr}})$, then it also satisfies $\varepsilon^{(i)}_{\mathcal{D}}$-DP$(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, D_{\mathrm{Pr}})$ by Proposition 5. Yet we can construct a sequence of budgets $\varepsilon'^{(1)}_{\mathcal{D}'}, \varepsilon'^{(2)}_{\mathcal{D}'}, \ldots$ such that

(1) $\varepsilon'^{(i)}_{\mathcal{D}'_0}$ converges to $L'_{\mathcal{D}'_0}$ as $i \to \infty$; and
(2) T satisfies $\varepsilon'^{(i)}_{\mathcal{D}'}$-DP$(\mathcal{X}, \mathscr{D}', d_{\mathcal{X}}, D_{\mathrm{Pr}})$ for all i.

Because $L'_{\mathcal{D}'_0} < L_{\mathcal{D}_0}$, this implies there exists some N such that T satisfies $\varepsilon^{(N)}_{\mathcal{D}}$-DP$(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, D_{\mathrm{Pr}})$ with $\varepsilon^{(N)}_{\mathcal{D}_0} < L_{\mathcal{D}_0}$. This contradicts the definition of $L_{\mathcal{D}_0}$. □

Appendix E. Proof of Theorem 1

In this appendix, we prove that Algorithm 1 (the Permutation Swapping Algorithm [PSA]) satisfies $\varepsilon_{\mathcal{D}}$-DP$(\mathcal{X}, \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}, d^r_{\text{HamS}}, D_{\text{Mult}})$ for the value of $\varepsilon_{\mathcal{D}}$ given in Theorem 1. Assume throughout this appendix the conditions of Theorem 1: that all $\boldsymbol{x} \in \mathcal{X}$ share a common set of variables $\boldsymbol{V}$, which is partitioned into subsets $\boldsymbol{V}_{\text{Swap}}$ and $\boldsymbol{V}_{\text{Hold}}$; and that the PSA swaps records at the same resolution as d^r_{HamS}.

By Proposition 4, we may also assume that there is exactly one matching variable, one non-matching holding variable, and one swapping variable. For ease of exposition, we assume that each of these variables can take on a finite number of values, which we denote by $m = 1, \ldots, \mathcal{M}$ and $h = 1, \ldots, \mathcal{H}$ and $s = 1, \ldots, \mathcal{S}$, respectively, although the proof immediately generalizes beyond this assumption. Recall that $n^{\boldsymbol{x}}_{mhs}$ is the count of records in $\boldsymbol{x}$, which take the value (m, h, s). Replacing a category m, h, s with $\cdot$ denotes a marginal count—for example, $n^{\boldsymbol{x}}_{m\cdot s} = \sum_{h=1}^{\mathcal{H}} n^{\boldsymbol{x}}_{mhs}$. We will drop $\boldsymbol{x}$ in the superscript and $\cdot$ in the subscript when this does not cause ambiguity.

Write $(M_i^{\boldsymbol{x}}, H_i^{\boldsymbol{x}}, S_i^{\boldsymbol{x}})$ for the i-th record in $\boldsymbol{x}$, so that we can write $\boldsymbol{x}$ as the vector $[(M_i, H_i, S_i)]_{i=1}^n$, where $n = n^{\boldsymbol{x}}_{\cdots} = |\boldsymbol{x}|$ is the number of records in $\boldsymbol{x}$. With this notation,

$$n^{\boldsymbol{x}}_{mhs} = \sum_{i=1}^{n} 1_{M_i^{\boldsymbol{x}}=m} 1_{H_i^{\boldsymbol{x}}=h} 1_{S_i^{\boldsymbol{x}}=s}.$$

Let $\ell_1^r(\boldsymbol{x}, \boldsymbol{x}')$ be the ℓ_1-distance on the interior cells of the fully saturated contingency table

$$\ell_1^r(\boldsymbol{x}, \boldsymbol{x}') := \sum_{m,h,s} \left| n^{\boldsymbol{x}}_{mhs} - n^{\boldsymbol{x}'}_{mhs} \right|. \tag{E.1}$$

Lemma 2. *$\ell_1^r(\boldsymbol{x}, \boldsymbol{x}') = 2d^r_{\text{HamS}}(\boldsymbol{x}, \boldsymbol{x}')$ if $|\boldsymbol{x}| = |\boldsymbol{x}'|$.*

Lemma 3. *D_{Mult} is a metric on the space of a.e. equal random variables (over the same probability space $\mathcal{T}$).*

Proof. It is easy to see that D_{Mult} is symmetric and $D_{\text{Mult}}(X, Y) = 0$ if and only if $X = Y$ a.e. All that remains is to verify the triangle inequality. Let $\{E_n\} \subset \mathcal{F}$ such that

$$\left| \ln \frac{\mathsf{P}(X \in E_n)}{\mathsf{P}(Z \in E_n)} \right| \to D_{\text{Mult}}(X, Z),$$

as $n \to \infty$. Then

$$\begin{aligned} \left| \ln \frac{\mathsf{P}(X \in E_n)}{\mathsf{P}(Z \in E_n)} \right| &\leq |\ln[\mathsf{P}(X \in E_n)] - \ln[\mathsf{P}(Y \in E_n)]| + |\ln[\mathsf{P}(Y \in E_n)] - \ln[\mathsf{P}(Z \in E_n)]| \\ &\leq D_{\text{Mult}}(X, Y) + D_{\text{Mult}}(Y, Z). \end{aligned}$$

□

Recall that σ_m is the random perturbation sampled by the PSA, which deranges the selected records in matching stratum m. Let σ be the permutation defined by $\sigma(i) = \sigma_{M_i}(i)$. Since σ_m fixes i whenever $M_i \neq m$, it is the case that $\sigma = \sigma_{\mathcal{M}} \circ \cdots \circ \sigma_1$. (Note that σ is a random function of the input data set $\boldsymbol{x}$, although we leave this dependence implicit.) For a permutation g, write $g(\boldsymbol{x})$ as shorthand for the data set in which the the values of the swapping variables have been permuted according to g. That is, if $\boldsymbol{x} = [(M_i, H_i, S_i)]_{i=1}^n$ then $g(\boldsymbol{x}) = [(M_i, H_i, S_{g(i)})]_{i=1}^n$. Given an input data set $\boldsymbol{x}$, the swapped data set $\sigma(\boldsymbol{x})$ generated by the PSA is denoted by $\boldsymbol{Z}$.

Let $\mathsf{P}_{\boldsymbol{x}}$ denote the probability induced by the randomness in the PSA (i.e., the randomness in selecting records and in sampling the permutation σ), taking the input data set $\boldsymbol{x}$ as fixed. Recall that the output of the PSA is the fully saturated contingency table $C(\boldsymbol{Z}) = [n^{\boldsymbol{Z}}_{jkl}]$.

Lemma 4. *If $\boldsymbol{x}$ and $\boldsymbol{x}'$ differ only by reordering of rows (i.e., $d^r_{\mathrm{HamS}}(\boldsymbol{x}, \boldsymbol{x}') = 0$), then*

$$D_{\mathrm{MULT}}[\mathsf{P}_{\boldsymbol{x}}(C(\boldsymbol{Z}) \in \cdot\ ), \mathsf{P}_{\boldsymbol{x}'}(C(\boldsymbol{Z}) \in \cdot\ )] = 0.$$

Proof. The contingency table $[n^{\boldsymbol{Z}}_{mhs}]$ is invariant to reordering of rows of $\boldsymbol{Z}$. Thus $\mathsf{P}_{\boldsymbol{x}}(C(\boldsymbol{Z}) \in \cdot\ ) = \mathsf{P}_{\boldsymbol{x}'}(C(\boldsymbol{Z}) \in \cdot\ )$. □

Lemma 5. *Fix some data universe $\mathcal{D} \in \mathscr{D}_{\boldsymbol{c}_{\mathrm{Swap}}}$ and some $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}$ with $d^r_{\mathrm{HamS}}(\boldsymbol{x}, \boldsymbol{x}') = \Delta$. Then there exists a permutation ρ that fixes exactly $n - \Delta$ records such that $C(\rho(\boldsymbol{x})) = C(\boldsymbol{x}')$.*

Proof. We have that $\Delta < \infty$ since the invariants $\boldsymbol{c}_{\mathrm{Swap}}$ imply that all data sets in $\mathcal{D}$ have the same number of records. Hence the symmetric difference $\boldsymbol{x} \ominus \boldsymbol{x}'$ contain 2Δ records, with Δ records from $\boldsymbol{x}$ and Δ records from $\boldsymbol{x}'$. Denote the records in $\boldsymbol{x} \ominus \boldsymbol{x}'$ that come from $\boldsymbol{x}$ by $\boldsymbol{x}_0$ and the records from $\boldsymbol{x}'$ by $\boldsymbol{x}'_0$, so that $\boldsymbol{x} \ominus \boldsymbol{x}'$ is the disjoint union of $\boldsymbol{x}_0$ and $\boldsymbol{x}'_0$.

Without loss of generality, we may assume that there is a single matching category ($\mathcal{M} = 1$). (If there is more than one matching category, apply the following argument to each category separately.) Then the data set $\boldsymbol{x}$ (disregarding the order of the records) can be represented as the matrix $C(\boldsymbol{x}) = [n^{\boldsymbol{x}}_{hs}]$.

We will need the following result ($*$) whose proof is straightforward: For any $\boldsymbol{x}'', \boldsymbol{x}''' \in \mathcal{X}$, the matrix $C(\boldsymbol{x}'') - C(\boldsymbol{x}''') = [n^{\boldsymbol{x}''}_{hs} - n^{\boldsymbol{x}'''}_{hs}]$ has zero row- and column-sums if and only if $\boldsymbol{x}'' \in \mathcal{D}_{\boldsymbol{c}_{\mathrm{Swap}}}(\boldsymbol{x}''')$. Moreover, $\boldsymbol{x}'' \in \mathcal{D}_{\boldsymbol{c}_{\mathrm{Swap}}}(\boldsymbol{x}''')$ implies $\boldsymbol{x}''_0 \in \mathcal{D}_{\boldsymbol{c}_{\mathrm{Swap}}}(\boldsymbol{x}'''_0)$ and $C(\boldsymbol{x}'') - C(\boldsymbol{x}''') = C(\boldsymbol{x}''_0) - C(\boldsymbol{x}'''_0)$.

By the above result ($*$), the marginal counts of $\boldsymbol{x}_0$ and $\boldsymbol{x}'_0$ agree: $n^{\boldsymbol{x}_0}_h = n^{\boldsymbol{x}'_0}_h$ and $n^{\boldsymbol{x}_0}_s = n^{\boldsymbol{x}'_0}_s$ for all h and s. But the interior cells disagree: if $n^{\boldsymbol{x}_0}_{hs} > 0$ then $n^{\boldsymbol{x}'_0}_{hs} = 0$ (and visa versa, swapping $\boldsymbol{x}_0$ and $\boldsymbol{x}'_0$). Further $C(\boldsymbol{x}_0) - C(\boldsymbol{x}'_0)$ has positive entries that sum to Δ and negative entries that sum to $-\Delta$, and zero row- and column-sums.

By construction of $\boldsymbol{x}_0$ and $\boldsymbol{x}'_0$, if we can permute $\boldsymbol{x}_0$ to produce $\boldsymbol{x}'_0$, then we can use the same permutation to produce $\boldsymbol{x}'$ from $\boldsymbol{x}$ (up to reordering of records). Critically, permutations of $\boldsymbol{x}_0$ can only derange Δ records (since there are only Δ records in $\boldsymbol{x}_0$) and indeed must derange Δ records to produce $\boldsymbol{x}'_0$ (since there are no records in common between $\boldsymbol{x}_0$ and $\boldsymbol{x}'_0$). Therefore we have reduced the problem: we need to find a permutation ρ (regardless of the number of records it fixes) such that $C(\rho(\boldsymbol{x}_0)) = C(\boldsymbol{x}'_0)$.

We construct this permutation ρ by induction on $\Delta = d^r_{\mathrm{HamS}}(\boldsymbol{x}, \boldsymbol{x}') = d^r_{\mathrm{HamS}}(\boldsymbol{x}_0, \boldsymbol{x}'_0)$. There are two base cases: The case $\Delta = 1$ is vacuous since $d^r_{\mathrm{HamS}}(\boldsymbol{x}, \boldsymbol{x}') = 1$ implies that $\boldsymbol{x}, \boldsymbol{x}'$ are not in the same data universe. Why? If $\ell^r_1(\boldsymbol{x}, \boldsymbol{x}') = 2$ then $C(\boldsymbol{x}) - C(\boldsymbol{x}')$ only has one or two nonzero cells. But this implies $C(\boldsymbol{x}) - C(\boldsymbol{x}')$ has a row or column with nonzero sum.

For $\Delta = 2$, the result ($*$) implies that the 2×2 top-left submatrix of $\boldsymbol{A} = C(\boldsymbol{x}_0) - C(\boldsymbol{x}'_0)$ looks like

$$\boldsymbol{A}_{1:2,1:2} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix},$$

(up to reordering of rows and columns). Therefore (up to reordering of records), $\boldsymbol{x}_0$ and $\boldsymbol{x}'_0$ differ by a single swap: if h, h', s, s' are indices such that $A_{hs} = A_{h's'} = 1$ then define ρ to be the swap of the records (k, l) and (k', l') in $\boldsymbol{x}_0$. We have $C(\rho(\boldsymbol{x}_0)) = C(\boldsymbol{x}'_0)$ as desired.

This completes the base cases. Now we will prove the induction step. By ($*$), we can always reorder the rows and columns of $\boldsymbol{A} = C(\boldsymbol{x}_0) - C(\boldsymbol{x}'_0)$ such that the 2×2 top-left submatrix looks like

$$\boldsymbol{A}_{1:2,1:2} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

with $A_{11}, A_{22} > 0$ and $A_{21} < 0$. Define $\boldsymbol{x}_1$ by swapping the records $(1,1)$ and $(2,2)$ in $\boldsymbol{x}_0$. Then the top-left submatrix of $\boldsymbol{A}' = C(\boldsymbol{x}_1) - C(\boldsymbol{x}_0')$ looks like

$$\boldsymbol{A}'_{1:2,1:2} = \begin{bmatrix} A_{11} - 1 & A_{12} + 1 \\ A_{21} + 1 & A_{22} - 1 \end{bmatrix},$$

and the rest of $\boldsymbol{A}'$ is the same as $\boldsymbol{A}$. If $A_{12} < 0$, then $\ell_1^r(\boldsymbol{x}_1, \boldsymbol{x}_0') = \ell_1(\boldsymbol{A}') = \ell_1(\boldsymbol{A}) - 4$. If $A_{12} \geq 0$ then $\ell_1(\boldsymbol{A}') = \ell_1(\boldsymbol{A}) - 2$. In both cases, we can use the induction hypothesis to give us a permutation ρ_1 of $\boldsymbol{x}_1$, which produces $\boldsymbol{x}_0'$ (up to reordering of records). Define the permutation ρ as the composition of ρ_1 with the swap of $(1,1)$ and $(2,2)$. Then $C(\rho(\boldsymbol{x}_0)) = C(\boldsymbol{x}_0')$ as desired. □

Proof of Theorem 1. Fix $\boldsymbol{x}$ and $\boldsymbol{x}'$ in the same data universe $\mathcal{D} \in \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}$. Let $\Delta = d^r_{\text{HamS}}(\boldsymbol{x}, \boldsymbol{x}')$. We need to prove that $D_{\text{Mult}}[\mathsf{P}_{\boldsymbol{x}}(C(\boldsymbol{Z})), \mathsf{P}_{\boldsymbol{x}'}(C(\boldsymbol{Z}))] \leq \Delta\varepsilon_{\mathcal{D}}$ or equivalently

$$\mathsf{P}_{\boldsymbol{x}}[C(\sigma(\boldsymbol{x})) = C(\boldsymbol{z})] \leq \exp(\Delta\varepsilon_{\mathcal{D}})\mathsf{P}_{\boldsymbol{x}'}[C(\sigma(\boldsymbol{x}')) = C(\boldsymbol{z})],$$

for all possible swapped data sets $\boldsymbol{z}$, where the probability is over the random permutation σ sampled by the PSA. Since the output $C(\boldsymbol{Z})$ does not depend on the ordering of the records in the input $\boldsymbol{x}$, we may without loss of generality reorder the records in $\boldsymbol{x}'$. Hence, there exists a permutation ρ, which fixes exactly $n - \Delta$ records such that $\rho(\boldsymbol{x}') = \boldsymbol{x}$ by Lemma 5.

Since

$$\mathsf{P}_{\boldsymbol{x}}[C(\sigma(\boldsymbol{x})) = C(\boldsymbol{z})] = \sum_{\boldsymbol{z}'} \mathsf{P}_{\boldsymbol{x}}[\sigma(\boldsymbol{x}) = \boldsymbol{z}'],$$

where the sum is over data sets $\boldsymbol{z}'$ with $d^r_{\text{HamS}}(\boldsymbol{z}, \boldsymbol{z}') = 0$, it suffices to show

$$\mathsf{P}_{\boldsymbol{x}}[\sigma(\boldsymbol{x}) = \boldsymbol{z}] \leq \exp(\Delta\varepsilon_{\mathcal{D}})\mathsf{P}_{\boldsymbol{x}'}[\sigma(\boldsymbol{x}') = \boldsymbol{z}], \tag{E.2}$$

for all possible swapped data sets $\boldsymbol{z}$.

Recall

$$b = \max\{0, n_{m\cdot\cdot} \mid \text{there are two records with different values in matching stratum } m\}.$$

If $b = 0$, then $\boldsymbol{x}$ and $\boldsymbol{x}'$ only differ by reordering of rows and hence $\varepsilon_{\mathcal{D}} = 0$ satisfies the differential privacy (DP) condition (Equation E.2) by Lemma 4. Having taken care of the case $b = 0$, from herein we may assume $b \geq 2$. (The case $b = 1$ is not possible.)

If $p \in \{0, 1\}$, then $\varepsilon_{\mathcal{D}} = \infty$ and the DP condition holds vacuously.

All that remains is to prove Equation E.2 holds in the case where $0 < p < 1$. Since $\boldsymbol{x}$ and $\boldsymbol{x}'$ themselves differ by the permutation ρ, we can permute $\boldsymbol{x}$ to produce $\boldsymbol{z}$ if and only if we can permute $\boldsymbol{x}'$ to produce $\boldsymbol{z}$. Thus, either $\mathsf{P}_{\boldsymbol{x}}(\sigma(\boldsymbol{x}) = \boldsymbol{z})$ and $\mathsf{P}_{\boldsymbol{x}'}(\sigma(\boldsymbol{x}') = \boldsymbol{z})$ are both zero, or they are both nonzero. We need only focus on the case where both probabilities are nonzero.

Recall that any permutation σ selected with nonzero probability by the PSA can be decomposed as $\sigma = \sigma_{\mathcal{M}} \circ \ldots \circ \sigma_1$, where σ_m will leave any unit i with matching category $M_i \neq m$ fixed. Write $\boldsymbol{x}_m$ for the records of $\boldsymbol{x}$ with $M_i = m$. Because we perform random selection and permutation independently for each stratum m,

$$\frac{\mathsf{P}_{\boldsymbol{x}}(\sigma(\boldsymbol{x}) = \boldsymbol{z})}{\mathsf{P}_{\boldsymbol{x}'}(\sigma(\boldsymbol{x}') = \boldsymbol{z})} = \frac{\prod_{m=1}^{\mathcal{M}} \mathsf{P}_{\boldsymbol{x}}(\sigma_m(\boldsymbol{x}_m) = \boldsymbol{z}_m)}{\prod_{m=1}^{\mathcal{M}} \mathsf{P}_{\boldsymbol{x}'}(\sigma_m(\boldsymbol{x}'_m) = \boldsymbol{z}_m)}.$$

Thus, to prove Equation E.2 it suffices to show

$$\frac{\mathsf{P}_{\boldsymbol{x}}(\sigma_m(\boldsymbol{x}_m) = \boldsymbol{z}_m)}{\mathsf{P}_{\boldsymbol{x}'}(\sigma_m(\boldsymbol{x}'_m) = \boldsymbol{z}_m)} \leq \exp(\Delta_m\varepsilon_{\mathcal{D}}), \tag{E.3}$$

for all m where $\Delta_m = d^r_{\text{HamS}}(\boldsymbol{x}_m, \boldsymbol{x}'_m)$.

Fix some m. For notation simplicity, whenever it is not essential to indicate the role of m, we will drop the subscript m from herein (until the end of the proof when we need to optimize over m). (This is the same as assuming $\boldsymbol{V}_{\text{Match}}$ is empty.)

Let $G_{\boldsymbol{x}\to\boldsymbol{z}} = \{\text{permutation } g : g(\boldsymbol{x}) = \boldsymbol{z}\}$. We use the notation g instead of σ to emphasize that g is not random, while the permutation σ chosen by Algorithm 1 is random. There is a bijection between $G_{\boldsymbol{x}\to\boldsymbol{z}}$ and $G_{\boldsymbol{x}'\to\boldsymbol{z}}$ given by $g \mapsto g \circ \rho$. Since

$$\mathsf{P}_{\boldsymbol{x}}(\sigma(\boldsymbol{x}) = \boldsymbol{z}) = \sum_{g\in G_{\boldsymbol{x}\to\boldsymbol{z}}} \mathsf{P}_{\boldsymbol{x}}(\sigma = g),$$

we will prove (E.3) by showing

$$\mathsf{P}_{\boldsymbol{x}}(\sigma = g) \le \exp(\Delta\varepsilon_{\mathcal{D}})\mathsf{P}_{\boldsymbol{x}'}(\sigma = g \circ \rho),$$

for all $g \in G_{\boldsymbol{x}\to\boldsymbol{z}}$. (Note that this may not obtain the best possible bound for specific $\boldsymbol{x}$ and $\boldsymbol{x}'$, but it is mathematically easier to bound $\mathsf{P}_{\boldsymbol{x}}(\sigma = g)/\mathsf{P}_{\boldsymbol{x}'}(\sigma = g \circ \rho)$ than bound the desired ratio

$$\frac{\sum_{g\in G_{\boldsymbol{x}\to\boldsymbol{z}}} \mathsf{P}_{\boldsymbol{x}}(\sigma = g)}{\sum_{g\in G_{\boldsymbol{x}\to\boldsymbol{z}}} \mathsf{P}_{\boldsymbol{x}'}(\sigma = g \circ \rho)}$$

directly. Yet in the case where $G_{\boldsymbol{x}\to\boldsymbol{z}}$ and $G_{\boldsymbol{x}'\to\boldsymbol{z}}$ are singletons, this approach gives tight bounds.)

Let k_g be the number of records (in category m) that were deranged (i.e., not fixed) by g and let $d(k)$ denote the k-th derangement number (i.e., the number of derangements of size k):

$$\begin{aligned} d(k) &= k!\sum_{j=0}^{k}\frac{(-1)^j}{j!} \\ &= kd(k-1) + (-1)^k \qquad \text{for } k \ge 0. \end{aligned} \tag{E.4}$$

Fix $g \in G_{\boldsymbol{x}\to\boldsymbol{z}}$ and $g' = g \circ \rho$. We now compute $\mathsf{P}_{\boldsymbol{x}}(\sigma = g)$. The permutation g is sampled in Algorithm 1 via a two-step procedure. Firstly records are independently selected for derangement with probability p. Suppose that g deranges records $\{i_1, \ldots, i_{k_g}\}$. Since we disallow the possibility of selecting only one record,

$$\mathsf{P}_{\boldsymbol{x}}(\text{the selected records are } \{i_1, \ldots, i_{k_g}\}) = \frac{p^{k_g}(1-p)^{n-k_g}}{1 - \mathsf{P}_{\boldsymbol{x}}(\text{exactly 1 record selected})}.$$

Secondly we sample uniformly from the set of all derangements of k_g records. Hence, we sample g with probability $[d(k_g)]^{-1}$and therefore,

$$\mathsf{P}_{\boldsymbol{x}}(\sigma = g) = \frac{p^{k_g}(1-p)^{n-k_g}}{[1 - \mathsf{P}_{\boldsymbol{x}}(\text{exactly 1 record selected})]d(k_g)}.$$

This gives

$$\frac{\mathsf{P}_{\boldsymbol{x}}(\sigma = g)}{\mathsf{P}_{\boldsymbol{x}'}(\sigma = g')} = o^{\delta}\frac{d(k_g - \delta)}{d(k_g)}, \tag{E.5}$$

where $o = p/(1-p)$ and $\delta = k_g - k_{g'}$.

Our aim is now to bound the right hand side of (E.5) by $\exp(\Delta\varepsilon_{\mathcal{D}})$. Since g' and g differ only by the permutation ρ (which fixes $n - \Delta$ records), we must have $k_g - \Delta \le k_{g'} \le k_g + \Delta$. Therefore,

there are at most $2\Delta + 1$ possible cases:

$$\begin{aligned}\delta \in S &= \big\{\delta \in \mathbb{Z} \mid -\Delta \le \delta \le \Delta \text{ and } \big(k_g - \delta = 0 \text{ or } 2 \le k_g - \delta \le n\big)\big\} \\ &= \{\delta \in \mathbb{Z} \mid \max(-\Delta, k_g - n) \le \delta \le \min(\Delta, k_g) \text{ and } \delta \ne k_g - 1\}.\end{aligned}$$

Suppose $0 < p \le 0.5$. Since $d(k)$ is nondecreasing (except at $k = 1$, which is not realizable by g or g') and $(1-p)/p \ge 1$, the right hand side of (E.5) is maximized when $k_{g'_m} = n_m$ and $k_{g_m} = n_m - \Delta_m$ (i.e., $\delta = -\Delta_m$), in which case

$$\begin{aligned}\frac{\mathsf{P}_{\boldsymbol{x}}(\sigma = g)}{\mathsf{P}_{\boldsymbol{x}'}(\sigma = g')} &= o^{-\Delta} \prod_{m=1}^{\mathcal{M}} \frac{d(n_m)}{d(n_m - \Delta_m)} \\ &\le o^{-\Delta} \prod_{m=1}^{\mathcal{M}} (n_m + 1)^{\Delta_m} \\ &\le o^{-\Delta}(b+1)^{\Delta} \\ &= \exp(\Delta\varepsilon_{\mathcal{D}}),\end{aligned} \tag{E.6}$$

for $\varepsilon_{\mathcal{D}} = \ln(b+1) - \ln o$. (The second line uses Lemma 6, which is given below this proof.)

Now suppose $0.5 < p < 1$. In the case of $\delta_m = \Delta_m$, the ratio (E.5) is maximized at o^{Δ_m} when $k_{g_m} = \Delta_m = 2$. Moreover, o^{Δ_m} also dominates $o^{\delta_m} \frac{d(k_{g_m} - \delta_m)}{d(k_{g_m})}$ for all $0 \le \delta_m \le \Delta_m$ and all possible k_{g_m}. Thus,

$$\begin{aligned}\frac{\mathsf{P}_{\boldsymbol{x}}(\sigma = g)}{\mathsf{P}_{\boldsymbol{x}'}(\sigma = g')} &\le \prod_{m=1}^{\mathcal{M}} \max\left\{o^{\Delta_m}, o^{\delta_m} \frac{d(k_{g_m} - \delta_m)}{d(k_{g_m})} : \delta_m \in S_m \text{ and } \delta_m < 0\right\} \\ &\le \prod_{m=1}^{\mathcal{M}} \max\big\{o^{\Delta_m}, o^{\delta_m}(k_{g_m} - \delta_m + 1)^{-\delta_m} : \delta_m \in S_m \text{ and } \delta_m < 0\big\} \\ &\le \prod_{m=1}^{\mathcal{M}} \max\big\{o^{\Delta_m}, o^{-\delta_m}(n_m + 1)^{\delta_m} : 0 < \delta_m \le \Delta_m\big\} \\ &\le \max\big\{o^{\Delta}, o^{-\delta}(b+1)^{\delta} : 0 < \delta \le \Delta\big\}.\end{aligned}$$

If $o^{-1}(b+1) \ge 1$ then $o^{-\delta}(b+1)^{\delta}$ is maximised at $\delta = \Delta$. Otherwise $o^{-\delta}(b+1)^{\delta} < 1 < o^{\Delta}$. Hence

$$\frac{\mathsf{P}_{\boldsymbol{x}}(\sigma = g)}{\mathsf{P}_{\boldsymbol{x}'}(\sigma = g')} \le \exp(\Delta\varepsilon_{\mathcal{D}}), \tag{E.7}$$

for $\varepsilon_{\mathcal{D}} = \max\big\{\ln o, \ln(b+1) - \ln o\big\}$. Combining Equations E.6 and E.7, we have

$$\varepsilon_{\mathcal{D}} = \begin{cases}\ln(b+1) - \ln o & \text{if } 0 < p \le 0.5 \text{ and } b > 0, \\ \max\big\{\ln o, \ln(b+1) - \ln o\big\} & \text{if } 0.5 < p < 1 \text{ and } b > 0.\end{cases} \tag{E.8}$$

When $b > 0$, we have $b \ge 2$ and hence also $\max\big\{\ln o, \ln(b+1) - \ln o\big\} = \ln(b+1) - \ln o$ for $0.5 < p \le \sqrt{b+1}/(\sqrt{b+1} + 1)$. Thus, Equation E.8 simplifies to

$$\varepsilon_{\mathcal{D}} = \begin{cases}\ln(b+1) - \ln o & \text{if } 0 < p \le \frac{\sqrt{b+1}}{\sqrt{b+1}+1} \text{ and } b > 0, \\ \ln o & \text{if } \frac{\sqrt{b+1}}{\sqrt{b+1}+1} < p < 1 \text{ and } b > 0.\end{cases}$$

as required. □

Lemma 6. *For any $k \in \mathbb{N}$ and any $a \in \mathbb{N}$ satisfying $0 \le a \le k$ and $a \neq k-1$,*

$$\frac{d(k)}{d(k-a)} \le (k+1)^a,$$

where $d(k)$ is the number of derangements of k elements (see Equation E.4).

Proof. We use induction on k. The base cases $k = 0, 1, 2$ are straightforward to verify since $d(0) = d(2) = 1$ and $d(1) = 0$. For the induction step, we can assume $k \ge 3$ so that $d(k-1) \ge 1$ and hence

$$\begin{aligned}\frac{d(k)}{d(k-a)} &= \frac{d(k)}{d(k-1)}\frac{d(k-1)}{d(k-a)} \\ &\le \frac{d(k)}{d(k-1)}k^{a-1}\end{aligned}$$

by the induction hypothesis. The result then follows by the identity given in Equation E.4:

$$\begin{aligned}\frac{d(k)}{d(k-1)} &= \frac{kd(k-1)+(-1)^k}{d(k-1)} \\ &\le k+1.\end{aligned}$$

□

Appendix F. Optimality of Theorem 1

Throughout this appendix we make the following assumptions. Following Proposition 4, we may assume there is a single matching variable, a single nonmatching holding variable, and a single swapping variable. Let $\mathscr{M}$, $\mathscr{H}$, and $\mathscr{S}$ be the domains for the matching variable, the nonmatching holding variable, and the swapping variable, respectively. Define $\mathcal{X}_\times = \bigcup_{k=1}^{\infty}(\mathscr{M}\times\mathscr{H}\times\mathscr{S})^k$. (Note $\mathcal{X}_{\text{CEF}} \subset \mathcal{X}_\times$, but we cannot assume the reverse inclusion.)

Recall that $b = \max\{0, n_{m\cdot} \mid$ there are two records with different values in matching stratum $m \in \mathscr{M}\}$; that $o = p/(1-p)$; and that $d(k)$ denotes the k-th derangement number (see Equation E.4).

Theorem 3. *Assume that $|\mathscr{H}|, |\mathscr{S}| \ge 2$ (so that $\mathcal{D} \in \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}$ are not all singletons and swapping is not completely vacuous).*

Suppose that the PSA satisfies $\varepsilon_{\mathcal{D}}$-DP$(\mathcal{X}_\times, \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}, d^r_{\text{HamS}}, D_{\text{Mult}})$. Then:

- *(A) If $p \in \{0, 1\}$, then there exists a universe $\mathcal{D}_0 \in \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}$ such that $\varepsilon_{\mathcal{D}_0} = \infty$.*
- *(B) If $0 < p < 1$, then there exists a universe $\mathcal{D}_0 \in \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}$ such that $\varepsilon_{\mathcal{D}_0} \ge \ln o$.*
- *(C) If $0 < p < 1$, then there exists a universe $\mathcal{D}_0 \in \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}$ such that $\varepsilon_{\mathcal{D}_0} \ge 0.5\ln[d(b)/d(b-2)] - \ln(o)$.*

The above values of $\varepsilon_{\mathcal{D}_0}$ describe lower bounds on the protection loss budget (PLB) of the Permutation Swapping Algorithm (PSA); any differential privacy (DP) specification for the PSA must have a PLB at least equal to these values. Comparing these lower bounds to the PLB $\varepsilon^{(1)}_{\mathcal{D}}$ given in Theorem 1 shows that $\varepsilon^{(1)}_{\mathcal{D}}$ is optimal in the weak sense that there exists universes $\mathcal{D}_0$ for which $\varepsilon^{(1)}_{\mathcal{D}_0}$ is arbitrarily close to the best possible budget $\varepsilon^{(\text{inf})}_{\mathcal{D}_0}$.

Theorem 4. *Assume $|\mathscr{H}|, |\mathscr{S}| \ge 4$. For each $\mathcal{D}_0 \in \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}$, define*

$$\varepsilon^{(\text{inf})}_{\mathcal{D}_0} = \inf\{\varepsilon_{\mathcal{D}_0} \mid \text{the PSA satisfies } \varepsilon_{\mathcal{D}}\text{-DP}(\mathcal{X}_\times, \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}, d^r_{\text{HamS}}, D_{\text{Mult}})\}.$$

(That is, $\varepsilon^{(\text{inf})}_{\mathcal{D}}$ is the pointwise infimum over all PLBs $\varepsilon_{\mathcal{D}}$ satisfied by the PSA.) Then $\varepsilon^{(\text{inf})}_{\mathcal{D}}$ is the smallest budget under which the PSA satisfies the DP flavor $(\mathcal{X}_\times, \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}, d^r_{\text{HamS}}, D_{\text{Mult}})$.

Let $\varepsilon^{(1)}_{\mathcal{D}}$ be the PLB given in Theorem 1. There exists $\mathcal{D}_0 \in \mathscr{D}_{\boldsymbol{c}_{\mathrm{Swap}}}$ such that

$$\varepsilon^{(1)}_{\mathcal{D}_0} - \varepsilon^{(\inf)}_{\mathcal{D}_0} \leq \begin{cases} f(b) & \text{if } 0 < p < \frac{\sqrt{b+1}}{\sqrt{b+1}+1} \text{ and } b > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where

$$f(b) = \frac{1}{2} \ln \left[\frac{(b+1)^2}{b(b-1)} \frac{1 + \frac{e}{2(b-2)!}}{1 - \frac{e}{2b!}} \right],$$

is a positive, monotonically decreasing function for $b \geq 2$ that converges to zero, and satisfies, for example, $f(b) \leq 0.148$ for all $b \geq 10$.

We emphasize that this is a weak form of optimality. A budget $\varepsilon_{\mathcal{D}}$ can be tight at the level of the output (in the sense that $\frac{\mathsf{P}_{\boldsymbol{x}}(C(\sigma)(\boldsymbol{x})=\boldsymbol{z})}{\mathsf{P}_{\boldsymbol{x}'}(C(\sigma)(\boldsymbol{x}')=\boldsymbol{z})} = \exp[\varepsilon_{\mathcal{D}} d^r_{\mathrm{HamS}}(\boldsymbol{x}, \boldsymbol{x}')]$ for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}$, all $\boldsymbol{z}$ and all $\mathcal{D}$); or at the level of the data (in the sense that $D_{\mathrm{Mult}}(\mathsf{P}_{\boldsymbol{x}}, \mathsf{P}_{\boldsymbol{x}'}) = \varepsilon_{\mathcal{D}} d^r_{\mathrm{HamS}}(\boldsymbol{x}, \boldsymbol{x}')$ for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}$ and all $\mathcal{D}$); or at the level of the universe (in the sense that $D_{\mathrm{Mult}}(\mathsf{P}_{\boldsymbol{x}}, \mathsf{P}_{\boldsymbol{x}'}) = \varepsilon_{\mathcal{D}} d^r_{\mathrm{HamS}}(\boldsymbol{x}, \boldsymbol{x}')$ for some $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}$, and all $\mathcal{D} \in \mathscr{D}_{\boldsymbol{c}_{\mathrm{Swap}}}$). The optimality of Theorem 1 is weaker than any of these notions; all we have shown is that, for all $\delta > 0$, there exists some $\mathcal{D}_0 \in \mathscr{D}_{\boldsymbol{c}_{\mathrm{Swap}}}$ and some $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}_0$ such that $\varepsilon^{(1)}_{\mathcal{D}_0} - D_{\mathrm{Mult}}(\mathsf{P}_{\boldsymbol{x}}, \mathsf{P}_{\boldsymbol{x}'}) < \delta$. Part of the suboptimality arises from the fact that $\varepsilon^{(1)}_{\mathcal{D}}$ is a function only of p and b. We could perform a tighter analysis of the PSA by allowing $\varepsilon_{\mathcal{D}}$ to depend on $\mathcal{D}$ in more complex ways (i.e., by allowing $\varepsilon_{\mathcal{D}}$ to be a function of other properties of $\mathcal{D}$, not just b).

Proof of Theorem 3. Result (A) follows from Propositions 6 and 7. Result (B) follows from Proposition 8. Result (C) follows from Propositions 9 and 10. □

Proof of Theorem 4. Because the multiverse $\mathscr{D}_{\boldsymbol{c}_{\mathrm{Swap}}}$ partitions $\mathcal{X}_\times$, the DP constraint imposed on each universe $\mathcal{D}$ is independent of the constraint on another universe $\mathcal{D}' \neq \mathcal{D}$. Hence the PSA does indeed satisfy $\varepsilon^{(\inf)}_{\mathcal{D}}$-DP$(\mathcal{X}_\times, \mathscr{D}_{\boldsymbol{c}_{\mathrm{Swap}}}, d^r_{\mathrm{HamS}}, D_{\mathrm{Mult}})$. Clearly, $\varepsilon^{(\inf)}_{\mathcal{D}} \leq \varepsilon_{\mathcal{D}}$ holds for all $\mathcal{D}$ and all budgets $\varepsilon_{\mathcal{D}}$ for which the PSA satisfies $\varepsilon_{\mathcal{D}}$-DP$(\mathcal{X}_\times, \mathscr{D}_{\boldsymbol{c}_{\mathrm{Swap}}}, d^r_{\mathrm{HamS}}, D_{\mathrm{Mult}})$. Hence $\varepsilon^{(\inf)}_{\mathcal{D}}$ is the smallest budget for which the PSA satisfies the DP flavor $(\mathcal{X}_\times, \mathscr{D}_{\boldsymbol{c}_{\mathrm{Swap}}}, d^r_{\mathrm{HamS}}, D_{\mathrm{Mult}})$.

Moving on to the second half of the theorem, we have by Theorem 3 that

$$\varepsilon^{(1)}_{\mathcal{D}_0} - \varepsilon^{(\inf)}_{\mathcal{D}_0} = 0,$$

if $b = 0$ or $p = 0$ or $\frac{\sqrt{b+1}}{\sqrt{b+1}+1} \leq p \leq 1$. On the other hand, if $0 < p < \frac{\sqrt{b+1}}{\sqrt{b+1}+1}$ and $b > 0$, then

$$\begin{aligned} \varepsilon^{(1)}_{\mathcal{D}_0} - \varepsilon^{(\inf)}_{\mathcal{D}_0} &\leq \ln(b+1) - \frac{1}{2} \ln[d(b)/d(b-2)] \\ &= \frac{1}{2} \ln \left[(b+1)^2 \frac{\left\lfloor \frac{(b-2)!}{e} + \frac{1}{2} \right\rfloor}{\left\lfloor \frac{b!}{e} + \frac{1}{2} \right\rfloor} \right] \\ &\leq \frac{1}{2} \ln \left[\frac{(b+1)^2}{b(b-1)} \frac{1 + \frac{e}{2(b-2)!}}{1 - \frac{e}{2b!}} \right] \\ &= f(b), \end{aligned}$$

where the first line follows by Proposition 9 and the second line by the identity $d(k) = \left\lfloor \frac{k!}{e} + \frac{1}{2} \right\rfloor$. The second term inside the logarithm

$$\frac{1 + \frac{e}{2(b-2)!}}{1 - \frac{e}{2b!}}$$

has a numerator that decreases with b and a denominator that increases. Hence this term is monotonically decreasing. The first term inside the logarithm $\frac{(b+1)^2}{b(b-1)}$ has negative first derivative and hence is also decreasing. Therefore, $f(b)$ is monotonically decreasing. Moveover, $f(b)$ is positive and converges to zero because both terms inside the logarithm are greater than one and converge to one. □

Proposition 6. *Suppose $p = 0$ and $|\mathscr{H}|, |\mathscr{S}| \geq 2$. Then there exists $\mathcal{D}_0 \in \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}$ such that $C(\boldsymbol{x}) \neq C(\boldsymbol{x}')$ for some $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}_0$. Hence, the PSA does not satisfy $\varepsilon_{\mathcal{D}}$-DP($\mathcal{X}_\times, \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}, d^r_{\text{HamS}}$, D_{MULT}) for any finite $\varepsilon_{\mathcal{D}_0}$ and any such $\mathcal{D}_0$.*

Proof. First we show that such a universe $\mathcal{D}_0 \in \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}$ exists. Given $|\mathscr{H}|, |\mathscr{S}| \geq 2$, the data sets $[(m, h, s), (m, h', s')]$ and $[(m, h, s'), (m, h', s)]$ (for any choice of $m \in \mathscr{M}$, $h \neq h' \in \mathscr{H}$ and $s \neq s' \in \mathscr{S}$) are in the same universe $\mathcal{D}_0 \in \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}$ and satisfy $C(\boldsymbol{x}) \neq C(\boldsymbol{x}')$.

Let $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}_\times$ be data sets that are in the same universe $\mathcal{D}_0$. Suppose $C(\boldsymbol{x}) \neq C(\boldsymbol{x}')$. If $p = 0$ then the permutation σ sampled by the PSA must be the identity. Thus, $\mathsf{P}_{\boldsymbol{x}'}(C(\sigma(\boldsymbol{x}')) = C(\boldsymbol{x})) = 0$ but $\mathsf{P}_{\boldsymbol{x}}(C(\sigma(\boldsymbol{x})) = C(\boldsymbol{x})) = 1$. Since $d^r_{\text{HamS}}(\boldsymbol{x}, \boldsymbol{x}') < \infty$, the DP condition

$$\mathsf{P}_{\boldsymbol{x}}(C(\sigma(\boldsymbol{x})) = C(\boldsymbol{x})) \leq \exp[d^r_{\text{HamS}}(\boldsymbol{x}, \boldsymbol{x}')\varepsilon_{\mathcal{D}_0}]\mathsf{P}_{\boldsymbol{x}'}(C(\sigma(\boldsymbol{x}')) = C(\boldsymbol{x})),$$

cannot be satisfied by a finite $\varepsilon_{\mathcal{D}_0}$. □

Proposition 7. *Suppose $p = 1$ and $|\mathscr{H}|, |\mathscr{S}| \geq 2$. Then there exists $\mathcal{D}_0 \in \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}$ with $n_{m_0 h_0} = n_{m_0 h'_0} = n_{m_0 s_0} = n_{m_0 s'_0} = 1$ for some $m_0 \in \mathscr{M}$, $h_0 \neq h'_0 \in \mathscr{H}$, and $s_0 \neq s'_0 \in \mathscr{S}$. Hence the PSA does not satisfy $\varepsilon_{\mathcal{D}}$-DP($\mathcal{X}_\times, \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}, d^r_{\text{HamS}}, D_{\text{MULT}}$) for any finite $\varepsilon_{\mathcal{D}_0}$ and any such $\mathcal{D}_0$.*

Proof. The universe given in the proof of Proposition 6 satisfies the property: $n_{m_0 h_0} = n_{m_0 h'_0} = n_{m_0 s_0} = n_{m_0 s'_0} = 1$ for some $m_0 \in \mathscr{M}$, $h_0 \neq h'_0 \in \mathscr{H}$ and $s_0 \neq s'_0 \in \mathscr{S}$.

Now take any $\mathcal{D}_0 \in \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}$ that satisfies this property. Then there exists $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}_0$ that differ by a single swap between (m_0, h_0, s_0) and (m_0, h'_0, s'_0)—that is,

$$\boldsymbol{x} = [(m_0, h_0, s_0), (m_0, h'_0, s'_0), \boldsymbol{x}_{3:n}],$$
$$\boldsymbol{x}' = [(m_0, h_0, s'_0), (m_0, h'_0, s_0), \boldsymbol{x}_{3:n}],$$

where $\boldsymbol{x}_{3:n} = [(M_i, H_i, S_i), i = 3, \ldots, n]$. Then $n^{\boldsymbol{x}}_{m_0 h_0 s_0} = n^{\boldsymbol{x}}_{m_0 h'_0 s'_0} = 1$ and $n^{\boldsymbol{x}}_{m_0 h s_0} = n^{\boldsymbol{x}}_{m_0 h_0 s} = 0$ for all $h \neq h_0$ and all $s \neq s_0$. Since no records can be fixed by σ when $p = 1$, we have $n^{\sigma(\boldsymbol{x})}_{m_0 h_0 s_0} = 0$ for any possible σ and hence $\mathsf{P}_{\boldsymbol{x}}(C(\sigma(\boldsymbol{x})) = C(\boldsymbol{x})) = 0$ but $\mathsf{P}_{\boldsymbol{x}}(C(\sigma(\boldsymbol{x}')) = C(\boldsymbol{x})) > 0$. □

Proposition 8. *Suppose that $0 < p < 1$ and $|\mathscr{H}|, |\mathscr{S}| \geq 2$. Then there exists $\mathcal{D}_0 \in \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}$ and $m_0 \in \mathscr{M}$ such that $n_{m_0} \geq 2$ and $n_{m_0 h}, n_{m_0 s} \in \{0, 1\}$ for all $h \in \mathscr{H}$ and $s \in \mathscr{S}$. A necessary condition for the PSA to satisfy $\varepsilon_{\mathcal{D}}$-DP($\mathcal{X}_\times, \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}, d^{hh}_{\text{HamS}}, D_{\text{MULT}}$) is that $\varepsilon_{\mathcal{D}_0} \geq \ln o$ for any such $\mathcal{D}_0$.*

Proof. Let $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}_0$ with $d^r_{\text{HamS}}(\boldsymbol{x}_{m_0}, \boldsymbol{x}'_{m_0}) = 2$ and $d^r_{\text{HamS}}(\boldsymbol{x}_m, \boldsymbol{x}'_m) = 0$ for all $m \neq m_0$. (Such a pair of data sets exist because $n_{m_0} \geq 2$.) Reorder the records in $\boldsymbol{x}'$ so that there exists a permutation ρ which deranges exactly two records and satisfies $\rho(\boldsymbol{x}') = \boldsymbol{x}$. (Such a permutation exists by Lemma 5.)

Because $n_{m_0h}, n_{m_0s} \in \{0,1\}$ for all $m \in \mathscr{M}$ and $s \in \mathscr{S}$, there are no vacuous swaps in the m_0 stratum. That is, $g(\boldsymbol{x}_{m_0}) \neq \boldsymbol{x}_{m_0}$ for all permutations g that are not the identity id. Hence $G_{\boldsymbol{x}_{m_0} \to \boldsymbol{x}_{m_0}} = \{\mathsf{id}\}$. Thus,

$$\begin{aligned}\frac{\mathsf{P}_{\boldsymbol{x}}(C(\sigma(\boldsymbol{x})) = C(\boldsymbol{x}))}{\mathsf{P}_{\boldsymbol{x}'}(C(\sigma(\boldsymbol{x}')) = C(\boldsymbol{x}))} &= \frac{\mathsf{P}_{\boldsymbol{x}}(C(\sigma_{m_0}(\boldsymbol{x}_{m_0})) = C(\boldsymbol{x}_{m_0}))}{\mathsf{P}_{\boldsymbol{x}'}(C(\sigma_{m_0}(\boldsymbol{x}'_{m_0})) = C(\boldsymbol{x}_{m_0}))} \\ &= \frac{\mathsf{P}_{\boldsymbol{x}}(\sigma_{m_0} = \mathsf{id})}{\mathsf{P}_{\boldsymbol{x}'}(\sigma_{m_0} = \rho)} \\ &= o^{-2}.\end{aligned}$$

Hence, $\mathsf{P}_{\boldsymbol{x}'}(C(\sigma(\boldsymbol{x}')) = C(\boldsymbol{x})) \leq \exp[d^r_{\text{HamS}}(\boldsymbol{x}, \boldsymbol{x}')\varepsilon_{\mathcal{D}_0}]\mathsf{P}_{\boldsymbol{x}}(C(\sigma(\boldsymbol{x})) = C(\boldsymbol{x}))$ if and only if $\varepsilon_{\mathcal{D}_0} \geq \ln o$. □

Proposition 9. *Suppose that $0 < p < 1$ and $|\mathscr{H}|, |\mathscr{S}| \geq 4$. Then there exists $\mathcal{D}_0 \in \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}$ that has the following properties:*

$$\max_h n_{m_0h} \leq \frac{b}{2} - 1 \text{ and } \max_s n_{m_0s} \leq \frac{b}{2} - 1, \tag{F.1}$$

for some $m_0 \in \mathscr{M}$ with $n_{m_0} = b$, and there exists $h_1 \neq h_2 \in \mathscr{H}$ and $s_1 \neq s_2 \in \mathscr{S}$ such that

$$n_{m_0h_1} = n_{m_0h_2} = n_{m_0s_1} = n_{m_0s_2} = 1. \tag{F.2}$$

A necessary condition for the PSA to satisfy $\varepsilon_{\mathcal{D}}$-DP$(\mathcal{X}_\times, \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}, d^r_{\text{HamS}}, D_{\text{Mult}})$ is that

$$\varepsilon_{\mathcal{D}_0} \geq 0.5 \ln[d(b)/d(b-2)] - \ln(o),$$

for any $\mathcal{D}_0$ satisfying the above properties.

We will use the following two lemmata in the proof of Proposition 9.

Lemma 7. *For any $\boldsymbol{x}$ and any permutation g,*

$$d^r_{\text{HamS}}(\boldsymbol{x}, g(\boldsymbol{x})) \leq k_g,$$

where k_g is the number of records that are deranged by g.

Proof. For every record (M_i, H_i, S_i) permuted by g, the counts in the fully saturated contingency table can change by at most 2: the count $n_{M_iH_iS_i}$ will decrease by (at most) 1 and the count $n_{M_i,H_i,S_{g(i)}}$ will increase by (at most) 1. Thus, in sum, the counts n_{mhs} can change by at most $2k_g$. That is,

$$\ell^r_1(\boldsymbol{x}, g(\boldsymbol{x})) = \sum_{m,h,s} \left| n^{\boldsymbol{x}}_{mhs} - n^{g(\boldsymbol{x})}_{mhs} \right| \leq 2k_g.$$

The desired result then follows by Lemma 2. □

Lemma 8. *Suppose that $\mathcal{D}_0 \in \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}$ satisfies Equation F.1. Then there exists $\boldsymbol{x} \in \mathcal{D}_0$ and a derangement g of $\boldsymbol{x}_{m_0}$ such that*

$$d^r_{\text{HamS}}(\boldsymbol{x}_{m_0}, g(\boldsymbol{x}_{m_0})) = b.$$

(In fact, such an $\boldsymbol{x}$ and g exist if and only if $\mathcal{D}_0$ satisfies Equation F.1.)

Proof. We suppress the subscript m_0 in $\boldsymbol{x}_{m_0}$ throughout the proof.

We begin by consider the cases $b = 1$ and $b = 0$ individually. Equation F.1 implies that $b \neq 1$. Similarly, no derangement of $\boldsymbol{x}$ exists when $b = 1$. In the case of $b = 0$, the result is also trivial. Hence we may assume throughout that $b \geq 2$.

"⇒": Suppose that $\mathcal{D}_0$ does not satisfy Equation F.1. Then $n_{m_0} < b$ or there exists (WLOG) a swapping category s_0 such that $n_{m_0 s_0} \geq b/2$. In the first case, any permutation g of $\boldsymbol{x}$ deranges at most n_{m_0} records and hence $d^r_{\text{HamS}}(\boldsymbol{x}, g(\boldsymbol{x})) < b$ by Lemma 8. By the pigeonhole principle, the second case implies every derangement g of $\boldsymbol{x}_{m_0}$ must send a record with swapping value s_0 to a record that also has value s_0. Yet the counts n_{mhs} are unaffected by permutations of records within the same swapping category s. Hence

$$\sum_{h,s} \left| n^{\boldsymbol{x}}_{m_0 hs} - n^{g(\boldsymbol{x})}_{m_0 hs} \right| \leq 2(k_g - 1) < 2b,$$

(where the first inequality follows by the reasoning in the proof of Lemma 7). The desired result then follows by Lemma 2.

"⇐": Assume for now that b is even. By Equation F.1, there exists $\boldsymbol{x} \in \mathcal{D}_0$ whose records are ordered so that every odd record has a different $\boldsymbol{V}_{\text{Swap}}$ and $\boldsymbol{V}_{\text{Hold}}$ compared to the subsequent record. That is, $H_i \neq H_{i+1}$ and $S_i \neq S_{i+1}$ for all odd i. (One can construct $\boldsymbol{x}$ by picking any $\boldsymbol{x}' \in \mathcal{D}_0$, ordering the records of $\boldsymbol{x}' \in \mathcal{D}_0$ so that the values of $\boldsymbol{V}_{\text{Hold}}$ differ between consecutive records, and then permuting $\boldsymbol{V}_{\text{Swap}}$ so that their values also differ between consecutive records.)

Construct g by swapping odd and even records:

$$g(i) = \begin{cases} i+1 & \text{if } i \text{ odd}, \\ i-1 & \text{if } i \text{ even}. \end{cases}$$

Then $k_g = b$ and $d^r_{\text{HamS}}(g(\boldsymbol{x}), \boldsymbol{x}) = b$.

Now suppose that b is odd. Then Equation F.1 implies that there exists $\boldsymbol{x} \in \mathcal{D}_0$ such that

1) $H_i \neq H_{i+1}$ and $S_i \neq S_{i+1}$ for all odd $i < n_{m_0}$; and
2) $H_{n_{m_0}} \notin \{H_{n_{m_0}-1}, H_{n_{m_0}-2}\}$ and $S_{n_{m_0}} \notin \{S_{n_{m_0}-1}, S_{n_{m_0}-2}\}$.

Why is this true? We already know that 1) must be true by the proof for even b. Suppose that 2) is not true for any $\boldsymbol{x}$. Then it must not be true for any $\boldsymbol{x}'$ that are just reorderings of the records of $\boldsymbol{x}$. Hence, for every adjacent pair $(i, i+1)$ (with $i < n_{m_0}$ odd), we must have $H_{n_{m_0}} \in \{H_i, H_{i+1}\}$ or $S_{n_{m_0}} \in \{S_i, S_{i+1}\}$. Yet this would contradict Equation F.1.

Construct g by swapping odd and even records, bar the final three records, which are permuted. That is,

$$g(i) = \begin{cases} i+1 & \text{if } i < n_{m_0} \text{ odd}, \\ i-1 & \text{if } i < n_{m_0} - 1 \text{ even}, \\ n_{m_0} & \text{if } i = n_{m_0} - 1, \\ n_{m_0} - 2 & \text{if } i = n_{m_0}. \end{cases}$$

As before, $k_g = b$ and $d^r_{\text{HamS}}(g(\boldsymbol{x}), \boldsymbol{x}) = b$. □

Proof of Proposition 9. Fix some $\mathcal{D}_0 \in \mathscr{D}_{\mathbf{c}_{\text{Swap}}}$, which satisfies the Equations F.1 and F.2. Such a universe exists when $|\mathscr{H}|, |\mathscr{S}| \geq 4$ because, for example,

$$\boldsymbol{x} = [(m_0, h_1, s_1), (m_0, h_2, s_2), (m_0, h_3, s_3), (m_0, h_3, s_3), (m_0, h_4, s_4), (m_0, h_4, s_4)],$$

satisfies these properties.

Let $\varepsilon_0 = 0.5 \ln[d(b)/d(b-2)] - \ln(o)$. We want to prove that

$$\frac{\mathsf{P}_{\boldsymbol{x}}(C(\sigma(\boldsymbol{x})) = C(\boldsymbol{z}))}{\mathsf{P}_{\boldsymbol{x}'}(C(\sigma(\boldsymbol{x}')) = C(\boldsymbol{z}))} = \exp[d^r_{\text{HamS}}(\boldsymbol{x}, \boldsymbol{x}')\varepsilon_0], \tag{F.3}$$

for some $\boldsymbol{x}, \boldsymbol{x}', \boldsymbol{z} \in \mathcal{D}_0$.

We will construct $\boldsymbol{x}$ and $\boldsymbol{x}'$ so that they are identical except within the matching category m_0. Then by independence between matching categories,

$$\frac{\mathsf{P}_{\boldsymbol{x}}(C(\sigma(\boldsymbol{x})) = C(\boldsymbol{z}))}{\mathsf{P}_{\boldsymbol{x}'}(C(\sigma(\boldsymbol{x}')) = C(\boldsymbol{z}))} = \frac{\mathsf{P}_{\boldsymbol{x}}(C(\sigma_{m_0}(\boldsymbol{x}_{m_0})) = C(\boldsymbol{z}_{m_0}))}{\mathsf{P}_{\boldsymbol{x}'}(C(\sigma_{m_0}(\boldsymbol{x}'_{m_0})) = C(\boldsymbol{z}_{m_0}))}.$$

This justifies dropping the subscript m_0 from $\boldsymbol{x}_{m_0}$ and ignoring records with matching categories not equal to m_0 throughout the remainder of the proof.

We construct $\boldsymbol{x}$ as follows: The first two records of $\boldsymbol{x}$ are (m_0, h_1, s_1) and (m_0, h_2, s_2). The remainder of the records satisfy Equation F.1. Hence construct the remainder of $\boldsymbol{x}$ according to the procedure given in the proof of Lemma 8. Let $\boldsymbol{x}'$ be the same as $\boldsymbol{x}$, except interchange the values of the swapping variable of the first two records. That is, $\boldsymbol{x}' = [(m_0, h_1, s_2), (m_0, h_2, s_1), \boldsymbol{x}_{3:n}]$.

Lemma 8 implies there exists a permutation g_0, which fixes the first two records and deranges the remaining records such that

$$d^r_{\mathrm{HamS}}(\boldsymbol{x}, g_0(\boldsymbol{x})) = b - 2.$$

Moreover, for $g'_0 = g_0 \circ (12)$, we have

$$d^r_{\mathrm{HamS}}(\boldsymbol{x}', g'_0(\boldsymbol{x}')) = b.$$

Set $\boldsymbol{z} = g_0(\boldsymbol{x}) = g'_0(\boldsymbol{x}')$.

Now we will prove Equation F.3 holds for these choices of $\boldsymbol{x}, \boldsymbol{x}'$ and $\boldsymbol{z}$. We have

$$\frac{\mathsf{P}_{\boldsymbol{x}}(C(\sigma(\boldsymbol{x})) = C(\boldsymbol{z}))}{\mathsf{P}_{\boldsymbol{x}'}(C(\sigma(\boldsymbol{x}')) = C(\boldsymbol{z}))} = \frac{\sum\limits_{\boldsymbol{z}' \text{ re-ordering of } \boldsymbol{z}} \mathsf{P}_{\boldsymbol{x}}(\sigma(\boldsymbol{x}) = \boldsymbol{z}')}{\sum\limits_{\boldsymbol{z}' \text{ re-ordering of } \boldsymbol{z}} \mathsf{P}_{\boldsymbol{x}'}(\sigma(\boldsymbol{x}') = \boldsymbol{z}')}.$$

Fix some $\boldsymbol{z}'$, which is a reordering of $\boldsymbol{z}$—i.e., some $\boldsymbol{z}'$ with $C(\boldsymbol{z}') = C(\boldsymbol{z})$. We will show that $\frac{\mathsf{P}_{\boldsymbol{x}}(\sigma(\boldsymbol{x})=\boldsymbol{z}')}{\mathsf{P}_{\boldsymbol{x}'}(\sigma(\boldsymbol{x}')=\boldsymbol{z}')} = \exp(2\varepsilon_0)$, when assuming that one of the numerator or the denominator is nonzero (which implies the other is also nonzero, since $\boldsymbol{x}$ and $\boldsymbol{x}'$ differ by a single swap). Since both the numerator and denominator are nonzero when $\boldsymbol{z}' = \boldsymbol{z}$, this result will prove Equation F.3.

We know that $d^r_{\mathrm{HamS}}(\boldsymbol{z}, \boldsymbol{z}') = 0$ and $d^r_{\mathrm{HamS}}(\boldsymbol{x}', \boldsymbol{z}) = b$. Then using the triangle inequality (twice, once for $\leq$ and once for $\geq$), $d^r_{\mathrm{HamS}}(\boldsymbol{x}', \boldsymbol{z}') = b$. Lemma 7 implies that $k_g = b$ for all $g \in G_{\boldsymbol{x}' \to \boldsymbol{z}'}$.

By the same reasoning, $d^r_{\mathrm{HamS}}(\boldsymbol{x}, \boldsymbol{z}') = b - 2$. This implies $k_g \geq b - 2$ for all $g \in G_{\boldsymbol{x} \to \boldsymbol{z}'}$ by Lemma 7. We now show that, in fact, $k_g = b - 2$. By construction,

$$n^{\boldsymbol{x}}_{m_0 h_1 s_1} = n^{\boldsymbol{x}}_{m_0 h_2 s_2} = 1 \text{ and } n^{\boldsymbol{x}}_{m_0 h_1 s} = n^{\boldsymbol{x}}_{m_0 h_2 s} = n^{\boldsymbol{x}}_{m_0 h s_1} = n^{\boldsymbol{x}}_{m_0 h s_2} = 0,$$

for all $h \notin \{h_1, h_2\}$ and $s \notin \{s_1, s_2\}$. These equations also hold for $\boldsymbol{z}$ and hence also for $\boldsymbol{z}'$. Thus, all $g \in G_{\boldsymbol{x} \to \boldsymbol{z}'}$ must fix the first two records and hence $k_g \leq b - 2$.

In the proof of Theorem 1, we showed that $\mathsf{P}_{\boldsymbol{x}}(\sigma = g)$ only depends on k_g and, furthermore, that

$$\frac{\mathsf{P}_{\boldsymbol{x}}(\sigma = g)}{\mathsf{P}_{\boldsymbol{x}'}(\sigma = g')} = \frac{(1-p)^2 d(b)}{p^2 d(b-2)},$$

when $k_g = b - 2$ and $k_{g'} = b$. Thus,

$$\frac{\mathsf{P}_{\boldsymbol{x}}(\sigma(\boldsymbol{x}) = \boldsymbol{z}')}{\mathsf{P}_{\boldsymbol{x}'}(\sigma(\boldsymbol{x}') = \boldsymbol{z}')} = \frac{\sum_{g \in G_{\boldsymbol{x} \to \boldsymbol{z}'}} \mathsf{P}_{\boldsymbol{x}}(\sigma = g)}{\sum_{g' \in G_{\boldsymbol{x}' \to \boldsymbol{Z}'}} \mathsf{P}_{\boldsymbol{x}'}(\sigma = g')} = \frac{(1-p)^2 d(b)}{p^2 d(b-2)} = \exp(2\varepsilon_0),$$

since $k_g = b - 2$ for all $g \in G_{\boldsymbol{x} \to \boldsymbol{z}'}$ and $k_{g'} = b$ for all $g' \in G_{\boldsymbol{x}' \to \boldsymbol{z}'}$. □

Proposition 10. *Suppose that* $0 < p \leq 0.5$ *and* $|\mathscr{H}|, |\mathscr{S}| \geq 2$. *Then there exists* $\mathcal{D}_0 \in \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}$ *such that* $b = 2$ *and*

$$n_{m_0 h_1} = n_{m_0 h_2} = n_{m_0 s_1} = n_{m_0 s_2} = 1, \tag{F.4}$$

for some $m_0 \in \mathscr{M}$ *with* $n_{m_0} = b$ *and some* $h_1 \neq h_2$ *and* $s_1 \neq s_2$.

A necessary condition for the PSA to satisfy $\varepsilon_{\mathcal{D}}$*-DP*$(\mathcal{X}_{\times}, \mathscr{D}_{\boldsymbol{c}_{\text{Swap}}}, d^r_{\text{HamS}}, D_{\text{MULT}})$ *is that*

$$\varepsilon_{\mathcal{D}_0} \geq 0.5 \ln[d(b)/d(b-2)] - \ln(o),$$

for any such $\mathcal{D}_0$.

Proof. Because $|\mathscr{H}|, |\mathscr{S}| \geq 2$, any data set of the form $[(m_0, h_1, s_1), (m_0, h_2, s_2)]$ satisfies Equation F.4. Moreover, $\boldsymbol{x}$ is in some universe $\mathcal{D}_0$, thereby proving the first half of the proposition. The second half of the proposition follows by the same reasoning as the proof of Proposition 9 applied to $\boldsymbol{x} = [(m_0, h_1, s_1), (m_0, h_2, s_2)]$ and $\boldsymbol{x} = [(m_0, h_1, s_2), (m_0, h_2, s_1)]$. □

Appendix G. Zero-Concentrated Differential Privacy

The normalized Rényi metric D_{NoR} is defined as:

$$D_{\text{NoR}}(\mathsf{P}, \mathsf{Q}) = \sup_{\alpha > 1} \frac{1}{\sqrt{\alpha}} \max\Big\{\sqrt{D_\alpha(\mathsf{P}||\mathsf{Q})}, \sqrt{D_\alpha(\mathsf{Q}||\mathsf{P})}\Big\},$$

where D_α is the Rényi divergence of order α:

$$D_\alpha(\mathsf{P}||\mathsf{Q}) = \begin{cases} \frac{1}{\alpha - 1} \ln \int \left[\frac{d\mathsf{P}}{d\mathsf{Q}}\right]^\alpha d\mathsf{Q}, & \text{if } \mathsf{P} \text{ is absolutely continuous wrt. } \mathsf{Q}, \\ \infty & \text{otherwise.} \end{cases}$$

Here $\frac{d\mathsf{P}}{d\mathsf{Q}}$ is the Radon-Nikodym derivative of P with respect to Q.

The term 'zCDP' refers to the class of DP flavors whose output premetric is the normalized Rényi metric, much as 'pure DP' refers to the class of flavors whose output premetric is the multiplicative distance (Equation 3.2). Note that we reparameterize ρ so that D_{NoR} is a metric (Bailie et al., 2026b). This is similar to the parameterization of zCDP given in Canonne et al. (2022) and Kairouz et al. (2021). The standard formulation of zCDP, as originally given in Bun and Steinke (2016), uses the square of the normalized Rényi metric as its output premetric. Consequently, the standard parameterization of zCDP is equal to ρ^2 under our formulation of zCDP.

Appendix H. Proof and Discussion of Theorem 2

Proof of Theorem 2. We first analyze the TopDown Algorithm (TDA) for producing the P.L. 94-171 Redistricting Summary File (PL). Abowd et al. (2022) prove that the mechanism $\boldsymbol{T}_{hh}$ that produces the household Noisy Measurement File (NMF) satisfies ρ-DP$(\mathcal{X}_{\text{CEF}}, \{\mathcal{X}_{\text{CEF}}\}, d^{hh}_{r_{bs}}, D_{\text{NoR}})$, where $\rho^2 = 0.07$ and $d^{hh}_{r_{bs}}$ is the input premetric corresponding to bounded differential privacy (DP) on household-records. But $(\mathcal{X}_{\text{CEF}}, \{\mathcal{X}_{\text{CEF}}\}, d^{hh}_{r_{bs}}, D_{\text{NoR}})$ and $(\mathcal{X}_{\text{CEF}}, \{\mathcal{X}_{\text{CEF}}\}, d^{hh}_{\text{HamS}}, D_{\text{NoR}})$ are equivalent DP flavors (see Bailie et al., 2026b). Hence $\boldsymbol{T}_{hh}$ satisfies ρ-DP$(\mathcal{X}_{\text{CEF}}, \mathscr{D}_{\boldsymbol{c}_{\text{TDA}}}, d^p_{\text{HamS}}, D_{\text{NoR}})$ by the second half of Proposition 5 with $\rho^2 = 0.07$. We can similarly conclude that $\boldsymbol{T}_p$ satisfies ρ-DP$(\mathcal{X}_{\text{CEF}}, \mathscr{D}_{\boldsymbol{c}_{\text{TDA}}}, d^p_{\text{HamS}}, D_{\text{NoR}})$ with $\rho^2 = 2.56$. Then by composition, the mechanism $\boldsymbol{T}_{ph} = [\boldsymbol{T}_p, \boldsymbol{T}_{hh}]$ has protection loss budget (PLB) $\rho^2 = 0.07 + 2.56 = 2.63$. Proposition 1 implies the invariants $\boldsymbol{c}_{\text{TDA}}(\boldsymbol{x}_p, \boldsymbol{x}_{hh})$—considered as a data release mechanism—satisfies ρ-DP$(\mathcal{X}_{\text{CEF}}, \mathscr{D}_{\boldsymbol{c}_{\text{TDA}}}, d^p_{\text{HamS}}, D_{\text{NoR}})$ with $\rho^2 = 0$. Therefore, the composed mechanism $\boldsymbol{T} = [\boldsymbol{T}_{ph}, \boldsymbol{c}_{\text{TDA}}]$ has budget $\rho^2 = 2.63$. The second step of the TDA is postprocessing on $\boldsymbol{T}$ and hence has the same budget.

The argument for producing the Demographic and Housing Characteristics File is almost analogous. The composed mechanism $\boldsymbol{T}_{ph} = [\boldsymbol{T}_p, \boldsymbol{T}_{hh}]$ has budget $\rho^2 = 7.70 + 4.96 = 12.66$. Now the second step of the TDA also uses the PL $\boldsymbol{P}$. Hence, this second step is postprocessing on the composed mechanism $[\boldsymbol{T}_{ph}, \boldsymbol{P}, \boldsymbol{c}_{\mathrm{TDA}}]$. This composed mechanism has budget $\rho^2 = 12.66 + 2.63 + 0 = 15.29$.

The second half of the theorem follows from Proposition 1. (Hence it can be generalized from $(\mathcal{X}_{\mathrm{CEF}}, \mathscr{D}_{\boldsymbol{c}'}, d^p_{\mathrm{HamS}}, D_{\mathrm{NoR}})$ to any DP flavor $(\mathcal{X}_{\mathrm{CEF}}, \mathscr{D}, d_{\mathcal{X}}, D_{\mathrm{Pr}})$ satisfying the assumptions of this proposition.) □

The second step of the TDA requires access to both the NMFs $[\boldsymbol{T}_p(\boldsymbol{x}_p), \boldsymbol{T}_{hh}(\boldsymbol{x}_{hh})]$ and the invariant statistics $\boldsymbol{c}_{\mathrm{TDA}}(\boldsymbol{x}_p, \boldsymbol{x}_{hh})$ computed on the Census Edited File (CEF). Under the DP flavor $(\mathcal{X}_{\mathrm{CEF}}, \{\mathcal{X}_{\mathrm{CEF}}\}, d_{\mathcal{X}}, D_{\mathrm{Pr}})$, the invariant statistics $\boldsymbol{c}_{\mathrm{TDA}}(\boldsymbol{x}_p, \boldsymbol{x}_{hh})$ cannot be released with finite budget. So the second step of the TDA is not actually postprocessing under this flavor—it is only postprocessing when conditioning on the invariants. In fact, the second half of Theorem 2 shows that any argument that relies on the TDA's second step being postprocessing must necessarily use a DP flavor that conditions on the invariants $\boldsymbol{c}_{\mathrm{TDA}}$.

To avoid inflating the PLB by a factor of 99,999, it is necessary to use person-records as the resolution of the Hamming distance in the TDA's DP specification. While the household mechanism $\boldsymbol{T}_{hh}$ satisfies $(\mathcal{X}_{\mathrm{CEF}}, \mathscr{D}_{\boldsymbol{c}_{\mathrm{TDA}}}, d^{hh}_{\mathrm{HamS}}, D_{\mathrm{NoR}})$, the sensitivity of the person-level query $\boldsymbol{Q}_p$ due to a single change in a household-record is 99,999. This is because the maximum possible household size in the CEF is 99,999 (Population Reference Bureau & U.S. Census Bureau's 2020 Census Data Products and Dissemination Team, 2023). This means $\boldsymbol{T}_p$ can satisfy $(\mathcal{X}_{\mathrm{CEF}}, \mathscr{D}_{\boldsymbol{c}_{\mathrm{TDA}}}, d^{hh}_{\mathrm{HamS}}, D_{\mathrm{NoR}})$ only if the PLB is amplified by 99,999.