Big data: Finders keepers, losers weepers?

Sax, M.

Link to publication

Citation for published version (APA):

Sax, M. (2016). Big data: Finders keepers, losers weepers? *Ethics and Information Technology*, *18*(1), 25-31. https://doi.org/10.1007/s10676-016-9394-0

CrossMark

Big data: Finders keepers, losers weepers?

Marijn Sax[1]

Abstract This article argues that big data's entrepreneurial potential is based not only on new technological developments that allow for the extraction of non-trivial, new insights out of existing data, but also on an ethical judgment that often remains implicit: namely the ethical judgment that those companies that generate these new insights can legitimately appropriate (the fruits of) these insights. As a result, the business model of big data companies is essentially founded on a libertarian-inspired 'finders, keepers' ethic. The article argues, next, that this presupposed 'finder, keepers' ethic is far from unproblematic and relies itself on multiple unconvincing assumptions. This leads to the conclusion that the conduct of companies working with big data might lack ethical justification.

Keywords Big data · Privacy · Personal data · Finders–keepers · Finders keepers ethic · Data mining · Entrepreneurship · Justice

Introduction

This article[1] deals with big data's entrepreneurial potential. With the arrival of big data, data are often lauded as 'the new oil' or as 'goldmine' from which 'nuggets of gold' can be retrieved (e.g. World Economic Forum 2011; Chen et al. 2012: 1167; McAfee and Brynjolfsson 2012: 59; Kroes

2013; Mayer-Schönberger and Cukier 2013: 16). Advanced data mining techniques allow companies[2] to generate non-trivial new insights out of existing data. Since collecting more data always translates itself into more potential new insights waiting to be extracted from the data, data hungriness is a structural condition of the big data world we have come to inhabit. This observation already goes a long way in explaining why many privacy worries are raised in the context of big data. In this article, I take these privacy worries as a point of departure, but I focus on the conduct of big data companies and, more specifically, on some of the fundamental assumptions underlying their conduct and that have remained implicit thus far.

Looking at the conduct of typical big data companies such as Acxiom, Bloomreach, Lotame, Palantir, Google, APT, Facebook, and Marketshare, we see that their success is largely dependent on their ability to generate new, non-trivial insights out of existing data. APT, for instance, mentions on its website that it offers "a cloud-based software application that efficiently analyzes promotional and financial data to generate actionable insights".[3] In a similar vein, Lotame claims that their "industry-leading analytics tools built into Lotame's Data Management Platform (DMP) enable you to uncover new insights".[4] Besides this ability to create new insights (and one may even say new data) out of

✉ Marijn Sax
M.Sax@uva.nl

1 Department of Philosophy, Capacity Group Philosophy and Public Affairs, University of Amsterdam, Oude Turfmarkt 143, 1012 GC Amsterdam, The Netherlands

---

[1] This article is partly derived from my MA thesis, which can be found here: http://dare.uva.nl/scriptie/575795.

[2] In this article I only discuss private sector applications of big data since I believe that these applications in particular share certain presupposed ethical considerations that stand in need of scrutiny. It might be possible to extend my anaysis to public sector applications as well. Due to limited space I will not attempt to do so here, although that would be interesting for further research.

[3] http://predictivetechnologies.com/products/apt-illuminate.aspx.

[4] http://www.lotame.com/platform/.

existing data, the business model of companies such as the ones mentioned is also premised on the fact that creators of these new insights may *appropriate* these new insights. The question I want to take up here is whether this *ethical* judgment is a legitimate one. Straightforward as it may sound, I argue that the legitimacy of this act of appropriating newly created insights essentially depends on the implicit acceptance of a 'finders, keepers' ethic.[5] Once this implicit acceptance of a 'finders, keepers' ethic is made explicit, it turns out that this 'finders, keepers' ethic itself depends on various implausible assumptions. As a result, it is far from obvious that the business of big data companies is legitimate from an ethical point of view. Against the background of potential threats of big data that other scholars have formulated (e.g. Crawford and Schultz 2014; Richards and King 2014; Barocas and Selbst 2016), this analysis could function as an additional basis of critique, highlighting the problematic normative presuppositions of big data companies' entrepreneurial conduct. Ultimately, my argument may have legal and political consequences concerning the regulation of big data companies.

This article proceeds in four steps. Firstly, the concept of big data will be introduced. Special attention will be given to big data's ability to generate non-trivial new insights out of existing data. Secondly, the 'finders, keepers' ethic will be introduced and it will be suggested that this ethic can help us understand the often implicit ethical judgments used to justify the conduct of big data companies. Thirdly, three assumptions of the 'finders, keepers' ethic will be addressed and criticized for their implausibility. Fourth and lastly, conclusions will be drawn on the basis of the preceding arguments.

Big data

Big data is a notoriously messy concept, so let me first explain what I take the concept big data to mean before proceeding. My description here is tentative and does not do justice to the great variety of academic literature dealing with big data. I choose to focus only on those aspects of big data that I need for my argument since my space is limited.

It is clear that the ability to collect and store larger volumes of data than ever before is a driving force behind the phenomenon of big data. This does not mean, however, that what makes big data big is simply a certain, large enough, volume of data. The transition from data to big data is not just a quantitative shift, it is a qualitative shift as well. Big data is not so much about amounts of data, as it is

about thinking about data, dealing with data, and approaching challenges and opportunities through the eyes of data. Mayer-Schönberger and Cukier (2013) identify three major shifts in moving from a 'normal' approach to data to a big data approach to data. The first shift constitutes a focus on creating datasets that approach $N = all$, instead of the careful creation of samples that should be representative of much larger populations. The second shift constitutes the belief that in order to achieve a (nearly) $N = all$ dataset, we should allow data from many different sources, even if the data are of dubious quality, to be included. In big data contexts, sheer size and volume are supposed to make up for messiness and low quality data. It "permit[s] us to loosen up our desire for exactitude" (Mayer-Schönberger and Cukier 2013: 13). The third shift constitutes the abandonment of the "age-old search for causality" since 'mere' correlations suffice (Mayer-Schönberger and Cukier 2013: 13).

Mayer-Schönberger and Cukier's convincing description of big data[6] clearly moves beyond a description that focuses merely on the size of datasets. What is of great importance to them is the idea that in a big data world more data from various sources, even sources of dubious quality, combined in a larger dataset will almost always lead to more powerful and valuable analyses. This insight leads to a certain 'data hungriness'; when doing big data analyses, more data on the input side is (almost) always preferable. Big data incentivizes data collection, but also data recombination[7]: "so-called big data brings together not only large amounts of data, but also various types that previously never would have been considered together" (Michael and Miller 2013: 22).

When using a big data approach to a problem, the goal is not to amass as much data as possible in order to simply paint an as accurate as possible picture. The goal is to come up with interesting and unanticipated insights that do not follow directly from the aggregated data themselves, but that need to be *extracted* or *generated* from them. As Rubinstein notes, big data "is best understood as a more powerful version of knowledge discovery in databases or data mining, which has been defined as 'the nontrivial extraction of implicit, previously unknown, and potentially useful information from data'"[8] (Rubinstein 2013: 76). As

---

[5] A word of caution: I do not claim that big data businesses do in fact subscribe to a notion of finders–keepers themselves. My claim is that the 'finders, keepers' ethic is a suitable theory to *model* their conduct convincingly and to help explain why big data businesses believe that their conduct is legitimate.

[6] The description by Mayer-Schönberger and Cukier that I draw on here has a somewhat different focus than the often-used description of big data that is based on the '3V's': volume, velocity, and variety. Many authors describe what is new about big data by referring to the vast increase in available data (volume), the increase in speed at which data are created, disseminated, and analyzed (velocity), and the wide variety of data that are combined (variety) (see for instance McAfee and Brynjolfsson (2012) for such a description of big data).

[7] The element of recombination comes close to the above mentioned 'V' of variety.

[8] Rubinstein quotes Fayyad et al. (1996) here.

an effect, storing information—even if one is not sure how useful the data are right *now*—becomes more and more interesting. Tene and Polonetsky go even one step further when they state that "the big data business model is antithetical to data minimization. It incentivizes collection of more data for longer periods of time. It is aimed precisely at those unanticipated secondary uses, the "crown jewels" of big data" (Tene and Polonetsky 2013: 259).

Data mining is the technique that can be seen as the big data analysis technique par excellence.[9] As was already mentioned, it is a technique that is aimed at *discovering* non-trivial new insights in existing datasets, insights that cannot simply be *observed* in datasets or follow automatically from datasets, but insights that have to be *extracted* or *generated* since they do not 'lie at the surface'. Crawford and Schultz (2014: 107), for example, write that "Big Data's analytics are simply too dynamic and unpredictable to determine if and when particular information or analyses will become or *generate* PII [personally identifiable information]". In order to achieve this extraction or generating[10] of new, emergent data, a combination of complex algorithms and brute computing force are used to work on the data.

The fact that we can discover new knowledge in existing data by using data mining techniques goes a long way in explaining why big data is a phenomenon that attracts so much attention. It surrounds big data with an aura of entrepreneurship. Since "by its very nature, big data analysis seeks surprising correlations and produces results that resist prediction" (Tene and Polonetsky 2013: 261), it always remains an open question what new information will be found and who will find it at what time. Entrepreneurs who work with big data hope that they will be the first to awaken the dormant value that lies hidden in big data datasets. The often used metaphors of data as the new oil and of datasets as goldmines with nuggets of gold hidden inside those datasets are expressions of this entrepreneurial potential.

Finders–keepers

In the previous section I have argued that from a technical perspective, big data's entrepreneurial potential resides in the fact that advanced data mining techniques can extract/generate unanticipated, non-trivial, new, and (commercially) interesting insights. In this section I want to shift the focus to the domain of ethics. Big data's entrepreneurial potential is equally dependent on the *legitimacy* of the appropriation of these newly extracted/generated insights by commercial parties. If companies cannot legitimately appropriate their newly mined insights, big data's attractiveness (from the perspective of companies) drastically declines or even evaporates completely. This question of legitimate appropriation is ultimately and ethical judgment that relies on substantial normative assumptions that can and should be scrutinized.

In the following paragraphs I will reconstruct the 'finders, keepers' ethic as Israel M. Kirzner (1978) has formulated it. This will help us to better understand the presupposed ethic of big data business. With the help of Kirnzer's theory I will bring to the fore the notion of finders–keepers that appears to legitimatize the appropriation of newly generated insights by big data companies. In the next section, I will problematize the normative assumptions of finders–keepers in big data contexts.

Kirzner (1978: 9) searches for what he calls the "morality of the entrepreneurial role" as he sees himself confronted with the challenge of justifying the appropriation by entrepreneurs of profits that are derived from either new applications or clever new uses of properties of existing goods. He proposes to resolve this challenge by accepting both "a particular *ethical judgment*" and "a particular *economic insight*" (Kirzner 1978: 17). The economic insight

> is that which permits us to perceive the discovery of a hitherto unknown market use for an already owned resource or commodity as the discovery of (and consequently the spontaneous establishment of ownership in) a hitherto *un-owned* element associated with that resource or commodity (Kirzner 1978: 17).

The ethical judgment is the acceptance of a 'finders, keepers' ethic. This means precisely what it appears to mean: those who find something that is not held by anybody, are, as they found it, the legitimate owners of that which they have discovered. Kirzner, however, proposes to reconceptualize what discovering something that was previously unheld means. "In order to introduce plausibility to the notion of finders–keepers, it appears necessary to adopt the view that, until a resource has been discovered, *it has not*, in the sense relevant to the rights of access and common use, *existed at all*" (Kirzner 1978: 17). This in effect means that, under Kirzner's reconceptualization, the discoverer of an unheld resource, brings the resource into existence and must therefore be seen as the *creator*[11] of the

---

[9] Barocas and Selbst (2016), for example, seem to use the concepts 'data mining' and 'big data' almost interchangeably.

[10] From here on out I will use 'extracting new insights' and 'generating new insights' interchangeably. Ontologically speaking, one could make a difference between, on the one hand, extracting something that is already latently present but has remained 'under the surface' thus far, and, on the other hand, generating new insights ex nihilo. Although an interesting topic in its own right, I propose that either interpretation is compatible with my argument.

[11] "We see the entrepreneur as "creator" not in the sense of the physical producer, but strictly in the sense of his being the discoverer of an available opportunity" (Kirzner 1978: 19).

resource. This is a significant reconceptualization since *creation* is a substantially different act than *acquisition from nature*. The latter occurs "against the background of a *given* unheld resources (even if no one is aware of their very existence)" (Kirzner 1978: 18), meaning that acquisition constitutes a *transfer*, namely from nature to the discoverer who becomes the first holder. The justness of the transfer could then be subject to ethical scrutiny. In the case of creation, no notion of transfer is involved in the establishment of ownership over the created good: "the finder-creator has spontaneously generated hitherto non-existent resources, and is seen, therefore, as their natural owner" (Kirzner 1978: 18). If the finder has created the goods by finding them, they cannot transfer from nature to the finder for the simple fact that the goods did not exist, in the relevant sense, in nature before they were found.[12]

Entrepreneurs, however, do not—or not exclusively—appropriate unheld resources, but also acquire held resources via just transfers, apply an entrepreneurial insight to create more commercial value, and then profit from these improvements. Those are two different situations, although the way they have to be understood according to Kirzner will turn out to be remarkably similar.

Entrepreneurs' main activity consists in finding and exploiting market opportunities, and this usually happens when they discover a new marketable property or application of a known and held resource or commodity. According to Kirzner, the 'economic insight' still applies in this case: "the discovery of a hitherto *unknown market use* for an already-owned resource or commodity constitutes the discovery of a hitherto *un-owned* element associated with that resource or commodity" (Kirzner 1978: 18). Put differently, the owner of a resource can only be the owner of those properties and potential applications of a resource that the owner is *explicitly* aware of. This is in stark contrast to the view that ownership means ownership of *all* a resource's or commodity's properties and applications, even the latent ones that have yet to be discovered.

To help us better understand the connection between Kirzner's theory and big data entrepreneurship, it may be instructive to introduce an example at this point (one that I borrowed from Kirzner): the example of oranges and orange juice. Image an entrepreneur who can buy oranges on the market for €5, who knows she can convert those

oranges into orange juice for €4 (costs of the conversion process of oranges to orange juice), and who also knows that consumers on the market are willing to pay €12 for the orange juice. The entrepreneur who discovers this market opportunity[13] can make a nice profit of (€12 − (€5 + €4)) = €3. The idea here is that the entrepreneur has *created*—ex nihilo—the new use for oranges and has therefore *created* the additional value of €3. In other words, the additional value of €3 was not, in any relevant sense, present in the oranges before the entrepreneur's intervention. This also means that the newly created value was not transferred from the original holder of the oranges to the entrepreneur, since this created value came into existence after the entrepreneur acquired the oranges and applied her insights to the product.

Now, Kirzner's answer to the question whether the appropriation of the fruits of the entrepreneurial insights is just, is simple. Exactly because we are dealing with ex nihilo creation by the entrepreneur *after the initial transaction*, "this additional $3 value may well be held *never to have been possessed by the seller at all*" (Kirzner 1978: 20). That part of the transaction which allows the entrepreneur to be an entrepreneur was, properly speaking, never part of the transaction. The concept transaction implies that the element of the good that is exploited by the entrepreneur to allow for her profitable insight was first held by the original holder and later, after the transaction, by the entrepreneur. But this is not the case, because the entrepreneur created the additional value ex nihilo, meaning that the initial seller never possessed it to begin with. As a result, Kirzner concludes that "justice *requires* that the "creator" be recognized as the owner of what he has "created": to deny the "creator" title would be to inflict injustice on him" (Kirzner 1978: 24).

Big data and finders–keepers

My claim now is that this notion of finders–keepers appears to be presupposed by those companies working with big data. These companies use, just like the orange juice entrepreneur, specific resources to create something new out of these resources. In the case of big data, (personal) data are used to extract non-trivial new information out of the given data via the technique of (predictive) data mining. The big data entrepreneurs then appropriate the (fruits of) the newly discovered insights. It is the 'gold' that is so emphasized by commentators. And just like in the case of

---

[12] The reader should note, however, that according to Kirzner finders–keepers does not fit every single case; there are legitimate cases of acquisition from nature (as opposed to creation), namely cases in which *everyone* is fully aware of the existence of the resource or goods. Kirzner's example to illustrate such cases: "The first man to land on Mars can hardly claim title to it as its "creator." In order to establish just ownership in an unheld resource the existence of which everyone is fully aware, it is certainly necessary to follow the criteria considered appropriate to just acquisition from nature" (Kirzner 1978: 18).

[13] Notice that one does not need to assume that the entrepreneur is the first to discover that there is juice in an orange. One only needs to acknowlegde that the entrepreneur discovered the *market opportunity* of squeezing the juice out of the orange and selling it for a profit.

the oranges, we can ask whether the big data entrepreneur can *legitimately* appropriate (the fruits of) these new insights. Kirzner's answer can still apply here. As long as the big data entrepreneur gets a hold of the original (personal) data in a just way, the entrepreneur is free to apply entrepreneurial insights and appropriate the additional value that she creates. Indeed, justice even *requires* that the entrepreneur is the legitimate owner of these new insights that are extracted/generated from the original data by the entrepreneur. Just like the original holder of the oranges was never the owner of the property of the oranges that allowed the entrepreneur to make orange juice out of the oranges, so the data subjects, whose data are used, were never the owners of those valuable insights that lie hidden in the data and that the big data entrepreneurs manage to extract. The data subjects providing the data cannot, in providing the data, be explicitly aware of the specific valuable insights that are hidden in their data. To see why, remember that these insights are in fact *new* non-trivial data, *created* out of the original data. The very nature of big data analysis is such that the newly mined insights do not follow directly from the original data, meaning that the original data subjects cannot, by definition, be aware of what emergent data can be extracted/generated from their personal data prior to the actual extraction via data mining. Due to this lack of explicit knowledge of all the unpredictable new insights that can be extracted from their personal data, the original data subjects can, under the 'finders, keepers' ethic, not be seen as the legitimate owners of these newly mined insights. The big data companies are the finders-creators of these new insights and their appropriation of the fruits of these new insights is therefore legitimate when the 'finders, keepers' ethic is accepted.

Problematic assumptions of finders–keepers in big data contexts

In the previous section I have argued that big data companies must presuppose a 'finders, keepers' ethic to explain why their appropriation of the new, valuable insights they manage to extract out of existing (personal) data can be seen as legitimate. In this section I will describe, in a very tentative manner, three assumptions of the 'finders, keepers' ethic that are especially problematic in big data contexts: (1) the presumed 'divisibility' of personal data; (2) the legitimacy of the original acquisition of personal data; and (3) the historical conception of justice that underlies finders–keepers. All three assumptions are problematic due to their insensitivity to the specificity of what kind of things personal data are and the functioning of personal data in big data contexts. As the discussion of these problematic

assumptions shows, explicating the normative basis of big data entrepreneurship allows for new types of critique on the conduct of big data companies.

Divisibility of personal data

As I have argued, the 'finders, keepers' ethic depends on the idea that within the same goods, some of the properties can be owned by the original holder, while other properties, namely those allowing for applications the original holder is not explicitly aware of, are unheld at the very same time and can thus, after discovery, be appropriated by the finder-creator. This introduces a certain kind of *divisibility* to goods which is necessary for finders-keepers to function adequately. In the case of inanimate objects this theory may be plausible—although even in those cases the divisibility of objects might feel highly artificial. But even if we assume, for the sake of argument, that this divisibility is plausible and accepted by everyone in the case of inanimate objects, it still does not follow that it is, by extension, equally plausible to think of personal data in a similar fashion. Granted, we often do speak of personal data as something—a resource, a thing—that can be owned, but does that automatically mean that personal data are to be understood as nothing more than inanimate objects?

I believe that the relationship between a person and her data is not exactly the same as the relationship between a person and a quotidian object (a phone, an orange, etc.) she owns. Floridi expresses this suspicion very accurately:

> [O]ne may still argue that an agent "owns" his or her information, […] in the precise sense in which an agent *is* her or his information. "My" in "my information" is not the same "my" as in "my car" but rather the same "my" as in "my body" or "my feelings": it expresses a sense of constitutive *belonging*, not of external *ownership*, a sense in which my body, my feelings and my information are part of me but are not my (legal) possession (Floridi 2005: 195).

If we understand the relation between an individual and her personal data the way Floridi does, it becomes immediately clear that it is far from unproblematic to conceive of personal data as if they were like oranges and orange juice. If Floridi's understanding of personal data is plausible, and I believe it is, then it can explain why the idea of divisibility—something finders–keepers needs—is much less convincing in relation to personal data than it is in relation to inanimate objects. Floridi notes that the 'my' in 'my information' "expresses a sense of constitutive belonging" (Floridi 2005: 195). This remark expresses the idea that your identity as a person is always necessarily

constituted—at least partly—by your information (either information about you, or information that you happen to 'possess'), seeing the person "as an informational entity" (Floridi 2005: 194) or "the nature of a person as being constituted by that person's information" (Floridi 2005: 195). This, in effect, means that unwanted meddling with one's personal data constitutes "changes in one's own identity as an informational entity" (Floridi 2005: 195). Based on Floridi's characterization of personal data, one could argue that thinking about personal data exactly like one thinks of oranges is to make a *category mistake*.[14] As an effect, additional arguments are needed to extend this idea of divisibility from inanimate objects to personal data.

At this point, the objection might be raised that big data analyses do not even need personal data to be effective. Completely anonymized data can also do the trick in some instances. If this is the case, the objector could claim that my argument, which is based on the 'specialness' of *personal* data, fails. In response, I would like to draw attention to different ways to define and understand the term 'personal data'. An often-used definition is the one found in the European Union's Data Protection Directive (95/46/EC), namely "any information relating to an identified or identifiable natural person" (article 2 (a)). This definition hinges on the question whether a piece of information or data can be explicitly related back to a person. If this standard definition is adopted, my argument my indeed seem dubious. However, in light of Floridi's remarks and big data's ability to generate inherently unpredictable outcomes that can influence the standing of data subjects significantly, I would like to suggest that a broader notion of personal data is appropriate. Even data that cannot be directly related to natural persons can be used, in big data contexts, to generate insights that can nonetheless have a significant impact on the lives and self-understanding of persons. Think for instance of discriminatory targeting practices as described by Turow (2011) that need not necessarily be based on personal data in the *legal* sense of the word to still have those discriminatory effects. I want to propose that in those cases where, legally speaking, anonymized and therefore non-personal data are used, there is still something personal about the data in a moral sense. Because these data can still have a significant influence on the lives and self-understandings of persons and are, seen from this perspective, still constitutive of personhood, I believe it makes sense to say that these data are still, in a moral sense, personal. As a result, it is still unconvincing to assume, without argument, that these data can be treated as just any quotidian object.

---

[14] This line of reasoning can be extend even further. The category mistake is also the foundation of the idea that personal data can be understood and treated as *commodities*. However, treating personal data as mere commodities is highly problematic from a moral point of view (see Roessler (2015) for such an analysis).

Acquisition of personal data

The *acquisition* of personal data—understood in the broader sense advocated above—by big data companies has not been problematized thus far. It has simply been assumed that big data companies acquire personal data in a just manner on the market, by way of transactions based on mutual consent. The idea that personal data are usually acquired in a just manner by big data companies because individuals consent to it may seem plausible. In reality, however, this position is quite hard to maintain. The idea that these transactions of personal data are based on informed consent, and that this informed consent is truly *informed* consent, is not very convincing in the face of the apparent failures of the informed consent model.

Zuiderveen Borgesius (2014) investigates the actual functioning of the informed consent model for the placement of cookies on computers and concludes that informed consent mechanisms are not strong enough to protect individuals. In a similar vein, Hoofnagle and Urban (2014) contend that informed consent mechanisms assume man to be a pure homo economicus: "Companies, long encouraged by regulators, issue privacy policies for consumers to read and act upon. In theory, consumers read these notices and make decisions according to their overall preferences, including preferences about privacy, price, service offering, and other attributes" (Hoofnagle and Urban 2014: 261–262). But for informed consent to function properly, this model of man as a perfect homo economicus must be somewhat adequate, and it is far from obvious that it is. Solove (2013: 1883) calls this informed consent based approach 'privacy self-management' and states that "empirical evidence and social science literature demonstrates that people's actual ability to make such informed and rational decisions does not even come close to the vision contemplated by privacy self-management".

The problems of the informed consent model can potentially erode the legitimacy of the original acquisition of the personal data that are used by big data companies. This, in turn, raises the question whether the appropriation of newly mined insights can be just if the data entrepreneurs work with to generate these insights were not acquired justly.

Historical conception of justice

Finders–keepers presupposes a historical conception of justice (Kirzner 1978: 9). A clear formulation of this historical conception can be found in Nozick (1974: 151–153). A historical conception of justice evaluates outcomes by focusing exclusively on two questions: (1) was the original acquisition just, and (2) were all the subsequent transfers just. If both conditions are satisfied, then outcomes must necessarily be just. As an effect, outcomes cannot be evaluated *in their own right*.

This conception of justice is problematic in big data contexts since an exclusive focus on the original acquisition of data and the subsequent transfers of data does not allow us to deal adequately with the challenges big data presents us with. To see why, one should notice that the current model of "[p]rivacy self-management addresses privacy in a series of isolated transactions guided by particular individuals. Privacy costs and benefits, however, are more appropriately assessed cumulatively and holistically" (Solove 2013: 1881). As was already shown, one of the unique aspects of big data is that outcomes are inherently unpredictable. Therefore, an exclusive focus on individual transactions, without focus on the actual aggregated outcomes these transactions can lead to, will necessarily miss something important. An historical conception of justice neglects data subjects' structural inability—and the general impossibility—to foresee the future outcomes of data mining. Not being able to evaluate these unpredictable outcomes in their own right is a serious problem for any analysis of big data that wants to focus on the desirability of certain applications and their outcomes.

Conclusion

In this article I have suggested that if one observes the ways in which big data companies work with data, an implicitly presupposed 'finders, keepers' ethic can be uncovered. This ethic serves the purpose of legitimizing the appropriation of newly mined, potentially profitable insights by these big data companies. However, because this hidden normative manipulative basis of big data entrepreneurship has remained implicit thus far, no explicit arguments have been provided in favor of it. This is problematic, for the plausibility of finders–keepers in the context of big data is far from self-evident. The assumption that just because a company has managed to extract new insights out of existing data warrants the appropriation and use of these insights, disregards a whole range of vital questions about the nature of personal data and the way these data can have a serious impact on the lives of people. As a result, this article suggests that the practices of big data companies by which they generate new insights out of existing personal data lack ample justification and should be subject to more intense ethical scrutiny.

References

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review, 104* [forthcoming]. Draft 8/15/2015. Retrieved November 21, 2015, from http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2477899.

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence analytics: From big data to big impact. *MIS Quarterly, 36*(4), 1165–1188.

Crawford, K., & Schultz, J. (2014). Big data and due process: Toward a framework to redress predictive privacy harms. *Boston College Law Review, 55*(1), 93–128.

Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery: An overview. In U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, & R. Uthurusamy (Eds.), *Advances in knowledge discovery and data mining* (pp. 1–34). Menlo Park: AAAI.

Floridi, L. (2005). The ontological interpretation of informational privacy. *Ethics and Information Technology, 7*(4), 185–200.

Hoofnagle, C. J., & Urban, J. M. (2014). Alan Westin's privacy homo economicus. *Wake Forest Law Review, 49*, 261–317.

Kirzner, I. M. (1978). Entrepreneurship, entitlement, and economic justice. *Eastern Economic Journal, 4*(1), 9–25.

Kroes, N. (2013). The big data revolution. Speech delivered on March 26, 2013 at the EIT Foundation Annual Innovation Forum, Brussels. Retrieved November 21, 2015, from http://europa.eu/rapid/press-release_SPEECH-13-261_en.htm.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data. A revolution that will transform how we live, work, and think*. New York: Houghton Mifflin Harcourt.

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review, 90*(10), 59–68.

Michael, K., & Miller, K. W. (2013). Big data: New opportunities and new challenges. *Computer, 46*(6), 22–24.

Nozick, R. (1974). *Anarchy, state, and utopia*. New York: Basic Books.

Richards, N. M., & King, J. H. (2014). Big data ethics. *Wake Forest Law Review, 49*, 393–432.

Roessler, B. (2015). Should personal data be a tradable good? On the moral limits of markets in privacy. In B. Roessler & D. M. Mokrosinska (Eds.), *Social dimensions of privacy: Interdisciplinary perspectives* (pp. 141–161). Cambridge: Cambridge University Press.

Rubinstein, I. S. (2013). Big data: The end of privacy or a new beginning? *International Data Privacy Law, 3*(2), 74–87.

Solove, D. J. (2013). Introduction: Privacy self-management and the consent dilemma. *Harvard Law Review, 126*, 1880–1903.

Tene, O., & Polonetsky, J. (2013). Big Data for All: Privacy and User Control in the Age of Analytics. *Northwestern Journal of Technology and Intellectual Property, 11*(5), 239–273.

Turow, J. (2011). *The daily you: How the new advertising industry is defining your identity and your worth*. New Haven, CN: Yale University Press.

World Economic Forum. (2011). *Personal data: The "New Oil" of the 21st century*. Retrieved November 21, 2015, from http://www.weforum.org/sessions/summary/personal-data-new-oil-21st-century.

Zuiderveen Borgesius, F. J. (2014). *Improving privacy protection in the area of behavioral targeting*. Dissertation, University of Amsterdam. Retrieved November 21, 2015, from http://dare.uva.nl/record/1/434236.