Is Data Labor? Two Conceptions of Work and the User-Platform Relationship

Julian David Jonker
University of Pennsylvania, USA

Some observers of the data economy have proposed that we treat data as labor. But are data contributions labor? Our folk conception of work emphasizes its importance and effort, such that work has a special interpersonal priority and deserves appreciation and compensation. The folk conception does not generally favor counting data as work, and so it serves as an error theory for reluctance to regulate data as labor. In contrast, labor regulation and policy focus on the political economy of labor, and in particular the bargaining dynamics that participants in a labor market face. Labor regulation aims to protect workers against the threats characteristically posed by such bargaining dynamics. Data-transferring interactions between internet platforms and users share similar bargaining dynamics, and so there is a promising functional similarity between labor regulation and proposals for data regulation that would facilitate data strikes and data unions.

Key Words: data governance, data as labor, surveillance capitalism, philosophy of work, data unions, data strikes

Many of the largest firms of our era are internet platforms and software service providers.[1] Their business models typically consist in providing internet services (such as cloud storage, search, and social networking) free of charge in exchange for collecting the behavioral data of users. The user data is then used to improve services so that users can be retained in an engagement cycle; or it is sold to third parties, or used to target ads at users (Hwang 2020; Zuboff 2019; Wu 2016; Lanier 2013). Since the ultimate revenue stream of internet platforms very often derives from trafficking in users' data so that ultimately users' attention can be sold to advertisers, it has become popular to say that users are not the customers of internet platforms, but rather the product.[2] Whether that is accurate or not, it

[1] The top ten companies in the S&P 500, by market capitalization, include Microsoft, Amazon, Alphabet (the parent company of Google), and Meta—which derive most of their revenue from providing internet services. In addition, the top ten also includes other information technology companies whose business model makes use of data: Apple, Nvidia (the manufacturer of chips used to train AI models), and Tesla. See "S&P 500 | S&P Dow Jones Indices" https://www.spglobal.com/spdji/en/indices/equity/sp-500/#data (August 31, 2023, accessed September 11, 2023).

[2] The slogan is hard to attribute (Oremus 2018). It was probably popularized when Jake Tapper tweeted it as a quote, attributing it to Bruce Schneier. But the idea dates back at least to the 1973 short film *Television Delivers People* by Carlota Fay Schoolman and Richard Serra.

Business Ethics Quarterly (2024), pp. 1–34. DOI:10.1017/beq.2024.25
Published by Cambridge University Press on behalf of the Society for Business Ethics.

highlights the sentiment that users are not being treated fairly when it comes to use of their data. Is the exchange in fact fair? The present paper aims to make some progress on this question by considering the economic value of users' contributions.

That the data generated by internet users has some value has become even more vivid with the rise of machine learning (ML) techniques, and the recent step change in performance by large language models, chatbots, and generative artificial intelligence (AI). The dramatic increase in the abilities of these AI systems is due not just to advances in ML techniques, but to the size of the datasets on which the ML algorithms are trained. Alongside the speculative boom in the valuations of generative AI startups there have been rising complaints by those who have provided training data without consent or compensation. A number of lawsuits involve authors claiming that generative AI providers such as OpenAI, the creator of ChatGPT, have violated their copyrights by training the model on their works (Creamer 2023; Samuelson 2023). The New York Times has also filed suit for copyright violation of its published reporting (Grynbaum and Mac 2023; Frenkel and Thompson 2023). Music labels are renegotiating their agreements to include royalties for musicians whose works have been used as training data (Kruppa 2023). And concerns about the use of recordings of actors for AI-generated replicas were at the heart of the actors' strike that paralyzed Hollywood through the summer of 2023 (Dargis 2023; Bedingfield 2023).

This rapidly evolving situation prompts the following question about data ethics and policy: what is the normative relation between an internet user and the data generated by their internet use? One proposal is that user data is a commodity, that is, a good that is relatively fungible in virtue of the fact that it has no essential normative relationship to any particular person (in contrast to a body part or an award or a service). This lies behind the once frequent claim that "data is the new oil" (The Economist 2017). But treating data as a commodity neglects the fact that data "belongs" to a particular user at least in the minimal sense that the data is about *their* behavior. This relational fact partly constitutes it as data and arguably limits its fungibility. A more novel proposal is that data should be treated as capital, with users occupying the role of investors in the business of internet platforms and deserving a financial security on the basis of their data contributions (Kim et al. 2021). Data as capital assumes that data belongs to users not just in the sense that it is about them but in the sense that they should (or do) *own* it. But this is not the only possible conception of the relation between users and their data.

A third proposal is that data is like labor, which must be combined with capital and other factors of production in order to produce surplus economic value. Similarly, an individual user's data requires the capital infrastructure of an internet platform—its compute, algorithm, and social graph—in order to produce the algorithmic predictions that make it valuable for other users, that is, valuable as more than just a fact about the contributing user. The analogy motivates the proposal to be discussed here: that data should be treated as labor ("Data as Labor"). Posner and Weyl articulate this idea when they say that "[users'] role as data producers is not fairly used or properly compensated," and that "[d]ata *work*, like 'women's work' and the cultural contributions of African Americans at one time, has been taken for granted" (2018,

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

209, emphasis added). Their proposal is based on ideas previously articulated by Jaron Lanier (2013), who envisaged users being paid for social media posts and other contributions to online culture, and further elaborated by Arrietta-Ibarra et al. (2018), who called for the creation of data unions.

How should we evaluate Data as Labor? This is the central question addressed by this paper. In order to answer it, I articulate two different ways of determining what counts as labor: a folk conception that focuses on the intuitive importance that work occupies in our personal lives and interpersonal relationships, and a political economy of labor that focuses on the characteristics of labor as a factor of production, in particular the characteristic vulnerabilities of workers that labor law and regulation often seek to address.[3] The political economy conception contrasts labor with other factors of economic production, and recognizes that what is distinctive of labor as a factor of production is that contributors of labor routinely face asymmetric bargaining dynamics. Data shares this structural asymmetry, and so the political economy conception supports Data as Labor, and recommends that we explore data regulations that are structurally analogous to labor regulations, such as those that facilitate data strikes and data unions. In contrast, the folk conception does not favor Data as Labor, since our data contributions typically share little of the personal and interpersonal importance of other work activities. The folk perspective does not undermine the case for regulating data as labor, but it does provide an error theory for why the Data as Labor agenda might initially strike some as implausible.

Questions about how to treat data and *data subjects* (i.e. the users who contribute data) are not just questions for policymakers, but for managers too. Given the global nature of data flows and the complexity of the issues raised by the internet, states have been slow to regulate the ways in which platforms and AI developers treat data (Kang 2023, Heikkilä 2023). This has left business to take a leadership role in forming policy, as exemplified by the voluntary commitments made early on by generative AI producers (Shear, Kang, and Sanger 2023). The question raised by Data as Labor is therefore one of business ethics, especially for those scholars who take seriously the call to explore the political dimension of markets (Heath, Moriarty, and Norman 2010; Smith 2019).[4]

[3] I follow Joseph Heath in using the term "political economy" to describe "the high-level question of how the economy as a whole should be organized" (Heath 2023, 4).

[4] It is worth noting that in business ethics the discussion of the digital economy in general, and especially of the present question about the status of data, is still nascent. While there is a robust discussion of technology policy and law in a wide variety of fields, and an intriguing literature on ethics arising from computer science, some describe business ethics as still catching up in its discussion of the ethical responsibilities of founders and executive managers of tech platforms (Martin 2023; see also Flyverbom, Deibert, and Matten 2019, 8). Nonetheless there is an emerging literature, though its concerns can be differentiated from the present paper. Much of the literature examines ethical issues concerning algorithms, such as whether managers may ethically use AI algorithms in decision-making (Martin and Parmar 2024), the obligation to communicate non-deceptively about algorithms (Seele and Schultz 2022; Schultz and Seele 2023), the ethics of personalized pricing (Steinberg 2020; Seele et al. 2021), addictive algorithms (Bhargava and Velasquez 2021), the right to an explanation of algorithmic decisions (Kim and Routledge 2020), the gamification of work by algorithms (Kim 2018), and how AI algorithms

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

1. THE PRACTICE OF BEHAVIORAL DATA

What does Data as Labor identify as data? Since this paper's agenda is normative, not metaphysical, it does not seek an ontological account of data for its own sake, nor does it aim to solve puzzles about the meaning or nature of data. But we do need some conception of data that will allow us to frame our normative inquiry precisely. For example, some of the philosophical literature on data has in mind the broad definition of "data" involved in scientific experimental practice (e.g. Suppes 1962; Harris 2003; Leonelli 2016). But this is too broad for our purposes.

The normative question posed by Data as Labor is instead concerned with what is often described as "big data," and even more narrowly, that subset of big data that is *data about human behavior obtained from users of internet platforms.* Some examples show the variety of the phenomenon: posting personal photos to a cloud storage site (e.g. Google Photos), or a sharing site (e.g. Flickr); generating and interacting with posts on a social media site (e.g. Reddit); filling out surveys on a crowdsourcing platform (e.g. Mechanical Turk); posting location data to a fitness app (e.g. Strava); using a search engine (e.g. Google Search) that logs interaction data; browsing a social media site (e.g. Facebook) that monitors attention and engagement; or filling out a reCAPTCHA security check that records the labeling of images.[5]

should ethically make tradeoffs (Scharding 2021). The present paper concerns data transfers between users and platforms rather than algorithms, and so falls outside of the scope of these discussions. Some papers also discuss the ethical obligations of platforms as facilitators of exchanges (Martin, Guo, and Easley 2023), though the issues do not rest directly on the acquisition of user data, as considered here. Other papers focus on the ethical and political impact of social media (Bhargava and Velasquez 2021; Whelan, Moon, and Grant 2013), but the scope of the present paper includes not just social media but all internet platforms that gather user data. Yet other papers in the emerging literature do address data rather than algorithms: for example Martin (2022) discusses the regulation of firms whose data collection puts them in a position to manipulate users; Steinberg (2022) discusses the use of surveillance technologies in the insurance industry; and Sarathy and Robertson (2003) consider the ethical management of data given concerns about privacy. But the present paper does not concern either privacy or unethical *uses* of data, but rather whether data transfers should be compensated as labor. This question falls squarely within the scope of business ethics, since it concerns both the management and regulation of an important contemporary business model. In addition, business ethics often go beyond the managerial perspective by asking which of a variety of potential stakeholders should be considered in managerial decision-making, and how they should be considered (Freeman 1984, 1994; Phillips 1997; Etzioni 1998; Phillips, Freeman, and Wicks 2003; Harrison and Wicks 2013). The agenda of Data as Labor brings data ethics squarely into this canonical framework, by asking what kind of stakeholder a data subject is.

[5] While these are all examples in which a platform obtains behavioral data through user input, this needn't be the case, and the fact that data is *behavioral* data is context-sensitive. For example, sensors that are installed in a building for the purpose of increasing energy efficiency by gathering temperature and humidity data may leak data about room occupants and their activities, and thus be used for gathering behavioral data even if they are not intended for that purpose (Morgner et al. 2017). Whether we describe this as data about human behavior will depend on whether we are concerned about the leakage, as we would be in the context of discussing privacy. The context-sensitivity of what counts as "data" is also implicated when data gathering practices discriminate by overlooking or discounting certain forms of data (e.g. Favaretto, De Clercq, and Elger 2019; Chun 2021)—though this important topic must be dealt with elsewhere. This context-sensitivity of the term "data" means that even though the forthcoming discussions of behavioral data focus on paradigmatic instances—such as the data collected from users by social media platforms when they intentionally post images and text—there will be room for debate about whether less paradigmatic interactions involve behavioral data or not.

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

These examples vary in dimensions such as whether the user knows that their data is being captured, generates the data in the course of an occupational activity, or is uniquely positioned to provide the data (Li et al. 2023). Therefore the motivation of the user for participating in the data may vary. In some cases the user seeks compensation (e.g. Mechanical Turk) or to increase a reputational score (e.g. Reddit). In some cases (e.g. Google Search) the user wishes to use a service and is either willing to give up their data in exchange or does not know that they do so, or even does not wish to do so but has no realistic alternative. Though these platforms often belong to very large companies (thus "Big Tech"), the question about the status of user data does not depend on them being large or prominent or monopolistic. That said, our inquiry is lent ethical urgency by the fact that exit is often difficult because platforms exhibit network externalities (e.g. social media) and users experience high switching costs when the value of the service increases with time or usage (e.g. email, recommendation engines).

Describing our topic as *data about human behavior obtained from internet users* already suggests a *pragmatic* account of data, that is, one that focuses on interactions between internet platforms and users and the practices of generating, collecting, storing, and using data attendant upon those interactions. The aforementioned context-sensitivity of data also supports viewing data as embedded in social and economic practices, rather than as an entity found in nature independent of such practices. Indeed, data only exists as such once it has been recorded or generated, or in some other way made the subject of our data processing practices.

This pragmatic view departs from a naive conception of data as a kind of raw material. Even very sophisticated discussions of data sometimes adopt this naive conception (e.g. Silver and Silver 1989, 6; cf. Beaulieu and Leonelli 2022, 53), and it is encouraged by characterizations of user data as a commodity, that is, "the new oil" (The Economist 2017). More recent critical studies of data, and big data in particular, are alive to the fact that data are not simply raw materials or found objects, but are instead mediated by practices that are not innocent of normative and theoretical commitments (e.g. Gitelman 2013; Kitchin 2021). Consider the data that led to a high school in New York City's Upper East Side being incorrectly labeled "the saddest spot in Manhattan" (Thorp 2021, 19–23). The number of social media posts made at a particular school that mention words associated with sadness, such as "cry" or "miserable," may appear to be "raw data" about the sentiment of members of the school's community. But these data are shaped by the availability of social media, by practices concerning who gets to post what (students may be barred from posting, or may not find a particular platform fashionable, or may think it proper only to post about certain topics), and by techniques of geolocating (that may not accurately demarcate a school from neighboring residences).

The pragmatic conception of data can capture the intuitive idea that data is closer to reality than other epistemic concepts, such as information and knowledge, while avoiding the naïveté of the raw materials conception. User data has a life cycle that flows from generating and collecting data points through cleaning and preparation to validation and analysis, and involves choices about how to store and disseminate it (see e.g. Flyverbom and Madsen 2015). Focusing on this life cycle clarifies the sense

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

in which data is closer to reality than, say, the end product of the data life cycle (which may be the outputs of an ML model). As we progress through this data life cycle, the object upon which these practices operate is more likely to be described as "information" and, at the very latest stages, "knowledge" (Zins 2007; Weinberger 2011, 1–3). While there is a diverse way in which philosophers relate the terms "information," "truth," "data," and "knowledge" (e.g. Boisot and Canals 2004; Floridi 2005; Leonelli 2015; Flyverbom and Whelan 2019), what these mappings share is the understanding that data is what lies closest to the world and earliest in our practices of processing and refinement—even if it never offers an entirely transparent window upon reality that is innocent of our practices.

Nonetheless, it is still misleading to think of data as raw material, even at the earliest stages. A platform plays a crucial role in shaping user interaction. It doesn't just passively record reality; it actively molds, channels, and constrains the ways users generate data (van Dijck 2009). This important point doesn't undermine the agency of the user (or the claim that they contribute labor) any more than the point that a factory worker's activities are enabled and structured by the tools and layout of the factory undermines the fact that the activities in the factory are labor. Instead, it is important for advocates of Data as Labor to emphasize that data is not a raw material that is discovered and mined like a commodity, but instead an artifact of an ongoing interaction between user and platform. By foregrounding the practice of data generation, especially the user's agential involvement, we avoid the accusation that it is a category mistake to equate data (as thing) with labor (as activity).

2. THE FOLK CONCEPTION OF LABOR

Aware of the unruly nature of our concept of labor, most writers settle for a set of associations or contrasts rather than a definition, and some deny that a definition is possible or necessary (Svendsen 2014, 7; Applebaum 1992). Raymond Geuss also resists providing a definition, but does identify claims that recur in our discussion of labor: that it involves exertion, that it is a necessity of life, and that it produces something of objective substance apart from the activity of work itself (2021, 5). The first two of these claims are indeed repeated within long-standing philosophical discussions of labor, and are implicit in two contrasts—prominent in our folk conception of labor—between 1) labor and leisure, and 2) labor and idleness. Given its rootedness in our philosophical tradition, the folk conception must be taken seriously, though because of its conservatism it is at best the basis of an error theory for commonly experienced recalcitrance to Data as Labor.

Before exploring the two familiar contrasts just noted, I wish to set aside a contrast that was suggested by Locke, writing of "the Labour of his Body, and the Work of his Hands" ([1698] 1988, §27), and lengthily elaborated by Arendt (1958). For Arendt, *labor* encompasses physical (as opposed to intellectual) activities that attend to biological needs, such as agriculture and housekeeping. Before modernity these activities took place in a private sphere that was strictly differentiated from a public sphere reserved for political activity. After modernity they take place in a commercially driven social realm ("the social") which has displaced the truly political public

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

sphere. In contrast, *work* encompasses those activities, such as crafts and the arts, which create objects of relative permanence constituting a shared and lasting intersubjective world in which public political action can take place. The tragedy of modernity is that work (and our identity as *homo faber*) is increasingly replaced by labor (and our identity as *animal laborans*), and the truly political public sphere by the material necessities of the social. This trajectory is partly due to the demands of developing technology, since labor is required to serve the repetitive needs of machines (Arendt 1958, 147).

Whether this complex set of claims is correct is a matter of great relevance for how we assess the transformations wrought by technology upon the quality of our life and work. Indeed, an Arendtian concern that the internet economy's reliance on behavioral data and predictive algorithms diminishes the space for true human freedom is at the core of Zuboff's complaints about surveillance capitalism (2019). If we assumed this framework then, as the discussion to come will show, data contributions would be more likely to count as Arendtian labor since they lack the intellectual engagement and the kind of intersubjective permanence that characterizes Arendtian work. And since Arendt diagnoses the widening sphere of labor as one of the pathologies of modernity, this designation should humble rather than cheer the advocate of Data as Labor. Despite this sort of concern, I will set aside Arendt's distinction as a distraction, and even use "labor" and "work" interchangeably. One reason for doing so is that there are deep problems with drawing the distinction as she does, problems that are exemplified by her failure to acknowledge that forms of labor such as care for dependents may be just as crucial to the creation of an intersubjective sphere as anything she designates as work.[6]

In contrast to Arendt's distinction, the distinction between labor and leisure is so long-standing that it can seem definitional—despite Veblen's claim that leisure is a symptom of social differentiation rather than a primeval feature of human nature ([1899] 2012). The distinction likely has its roots in the ancient Greek discussion of *scholê*, which is sometimes translated as "leisure" but really connotes a time free of service to anything other than discovering the truth, and *ascholia*, which connotes a kind of distracted busy-ness. Both Plato and Aristotle esteemed *scholê* as the highest end of life, and deemed *ascholia* an unwelcome distraction (Kalimtzis 2017).[7]

In modern times the contrast has inverted, perhaps because of Protestant attitudes toward work (Weber 1976; Applebaum 1992, 321–37). Labor is now frequently

[6] See also Sheth (2002, 176–97). As Hannah Pitkin has pointed out, Arendt's criticism of Marx (namely, that his entire system is misguided because it focuses on labor rather than work) misfires precisely because Marx does not presuppose the distinction. For Marx, humans are the creatures who distinctively "create and sustain our specifically human cultures and character" through labor [*Arbeit*] (Pitkin 1998, 136). Any attempt to sustain the distinction in the context of data would require us to disentangle the biological and material aspects of data and its intersubjective and agential aspects in precisely the way that I cautioned against in Section 1. But if the reader resists this move, they may treat what follows as being about Arendtian labor without any damage to the argument.

[7] Of course, to the extent that some members of Greek society were able to pursue such an ideal it was largely because socially necessary tasks were performed by slaves and women (Applebaum 1992, 170–75).

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

characterized as both individually and socially necessary, while leisure is characterized as dispensable and self-interested, a distraction from pursuing our common good. Labor is necessary in several ways. First, most people in a modern market economy must work for the means of their subsistence, or at least in order to live a decent life (Geuss 2021, 15). Many, already devoting so much time to work, also hope that their jobs will satisfy an existential need for meaning (Schwartz 1982; Ciulla 2001; Veltman 2016; Noonan 2020). For those pursuing such an ideal, work is a vocation or a calling that is necessary for living a meaningful life (Wrzesniewski et al. 1997). At least, they expect their work choices to bring them important "goods of work," such as the sense of making a social contribution and the development of skillful abilities (Gheaus and Herzog 2016). Furthermore, labor is socially necessary in that it is required to satisfy the needs of society for vital goods and services—even taking into account ongoing advances in technology (Deranty 2022; cf. Danaher 2019). Indeed, jobs that do not appear to satisfy social needs are sometimes deemed to be work in name only, that is, they are "bullshit jobs" (Graeber 2018).

While the contrast between labor as individually and socially necessary and leisure as frivolous and dispensable does not always withstand scrutiny, it does capture a hardy sentiment that guides our folk conception. The necessity of labor has also played a role in political philosophy. Hegel writes that labor has not only material but civic importance. Labor involves practical education, meaning that the worker must develop the habit of being constantly occupied, and must also learn to work under the direction of others ([1820] 1991, §197). In addition, the division of labor involves members of society in increasing interdependence, so that work aimed at satisfying the worker's self-directed need ends up satisfying the needs of everyone (§198–99). This is not a claim about the invisible hand and the efficiency of the market, but rather about how labor is a mechanism of social unity in a plural society (Sheth 2002, 76–77; for a contemporary version of the claim, see Estlund 2005).

The opposition of labor to idleness is best exemplified by a famous passage from Marx, in which he contrasts human labor with the architectural work of the spider and the bee (Marx [1867] 1976, 284). Whereas Hegel emphasizes that labor is a matter of the worker's being directed by another, Marx emphasizes the worker's rationality. Thus "what distinguishes the worst architect from the best of the bees is that the architect builds the cell in his mind before he constructs it in wax" and thereby "realizes [*verwirklicht*] his own purpose in those materials" (284).[8] This point may appear to ground a contrast between labor and other sorts of nonpurposive activity (perhaps being famous or born with title) for which people are compensated but which should not be counted as labor.

[8] It is because of this view about the nature of labor that Marx also inverts the foregoing claim of Hegel's about interdependence and labor, pointing out that under capitalism the division of labor does not so much lead to the interdependence of workers as to their dependence upon owners of capital. It is one of the pathologies of capitalism that the worker does not enjoy "the free play of his own physical and mental powers," (Marx [1867] 1976, 284) but instead experiences their labor as extorted from them by the owner of capital (325).

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

But how should we draw this contrast exactly? We could not say that involuntary actions do not constitute labor, lest we are to deny that slaves labor. Nor should we say that the celebrity does not labor because they are being paid for mere traits rather than actions. Much of the labor that has gone unrecognized—including sex work, pregnancy and childbirth, child-rearing, and various forms of affective support in and outside the workplace (see e.g. White 2003, 108–13; Bergmann 2005; Folbre 2008, 97–102)—involve nonintentional activities and traits just as much as they do intentional activities. It is a part of the epistemic injustice of this neglect that the home manager or parent are cast as being more like Marx's spider or bee than they are like the cognizing architect. Furthermore, even well-recognized forms of labor—such as being an artist's model, a research subject, a nightclub bouncer—involve making use of and compensating the worker for traits that are not produced through purposeful activity.

A more helpful way of construing Marx's claim is that what distinguishes human labor is not purposefulness but effort, in the sense that the worker has both the ability and (at some level) the preference to have done otherwise. Human labor is distinctively rational in the sense that it involves the exercise of rational agency. It is not just work but hard work, requiring that the agent uses their body or their will, or at least their schedule, to satisfy the needs of the job. Therein lies the difference between the bee and the architect. The bee builds the honeycomb without thought and therefore without distraction, whereas even the writer who believes they can do nothing but pursue their calling to write must wrestle with each recalcitrant sentence.

Admittedly, understanding effort in this way leads to a very generous account of work. The celebrity who puts off a night in front of the television in favor of a paid appearance does something effortful only in the sense that they do not act on their instincts like the bee does—that is, they put off their immediate desires, though doing so may not be particularly difficult or involve any complicated planning or cognition. But why should we exclude comfortable jobs like this from our account of work? Only because of an overly subjective judgment about what sorts of activities are so socially valuable or individually difficult that they deserve to be counted as work.

Such judgmental tendencies suggest that what we count as labor is ultimately driven by our sense of its normative importance. When we describe an activity as labor, we are not giving it a merely descriptive classification. Instead we make a claim about the role that the activity appropriately plays in our practical lives. This normative aspect of our concept of labor makes sense of why the contrasts between labor and leisure and between effort and idleness are so resilient, even though their application in particular instances is contested. It will often arouse controversy to say that some novel or marginal activity has the same normative importance as others that have already been accepted as labor. When we argue about whether a person's engagement as an artist's model or influencer or day trader counts as work, we are arguing over whether their activities should make the same claims upon us as their activities at the steel factory.

What then are the normative upshots of treating an activity as labor? First, because labor is necessary both individually and socially, an agent's labor demands our

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

respect for their time and focus. Second, because of its effortfulness, an agent's labor demands recognition from the individual, the employer, and third parties. These normative upshots can be illustrated by contrasting our judgments about cases such as the following:

> *Joti*
>
> Joti is an investment analyst who likes to wind down by playing *Call of Duty*. One day while she is doing this her spouse calls her to come and eat dinner with their kids. The conversation goes as follows:
>
> (1) Joti: "I can't right now, I'm busy!"
>
> (2) Joti's spouse: "No you're not, you're playing video games!"

The conversation seems natural enough. This is not just because of linguistic judgments. We are also inclined to make a normative judgment about the appropriateness of Joti's reaction and her spouse's response. Indeed, Joti's spouse appears to have a knock down argument: Joti does not have a good reason not to come to dinner. Of course, we could fill out the situation in various ways in order to make Joti's reaction more appropriate: perhaps she has been working particularly hard this week and deserves to be left alone, or perhaps she and her spouse are fighting and it would be better for everyone if they did not sit at the same table. But these are countervailing reasons to be weighed against the fact that Joti is not *really* busy. She is not doing anything that has the sort of priority that work normally has. She is just playing games, and this does not, in the normal scheme of things, give her family members a reason not to interfere by calling her to dinner. Compare the following case:

> *Johan*
>
> Johan is an esports champion who plays professionally. He is practicing a tactic in *Call of Duty*—esports players must drill in much the same way that basketball players do. He is in fact performing the exact same maneuver that Joti was. And like Joti, Johan's spouse is calling him to come and eat dinner with their kids. The conversation goes as follows:
>
> (3) Johan: "I can't right now, I'm busy!"
>
> (4) Johan's spouse: "Why do you have to work at dinner time?"
>
> (5) Johan's spouse: * "No you're not, you're playing video games!"

The asterisk next to (5) denotes (as in linguistics) that the response is infelicitous. Johan is not just playing video games. He is also working, since playing video games is what he does for work. The infelicity has to do with the pragmatic effect of the sentence, that is, the fact that it is issued as a complaint. Johan's spouse can correctly describe Johan as playing video games, but cannot (in this context) appropriately *complain* that he is shirking off dinner time for no good reason. The complaint has to be something else, such as the claim in (4) that Johan is doing something wrong by working while it's dinner time.

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

These judgments make sense in light of the earlier claim that describing an activity as work or labor entails that it enjoys a certain priority over other activities. Some clarifications of this claim are in order. It is the prioritizing of an activity that characterizes it as work, rather than whatever reasons are given for prioritizing it. Thus prioritizing work is consistent with disliking it or thinking it valueless. Moreover, whether an activity is prioritized is a fact about normative attitudes, rather than about the objective value of the activity. There may be good reason not to prioritize an activity—moral reasons against engaging in it, or reasons to prioritize some more valuable activity. Finally, this claim about prioritization is what I take to be the essential theme running through our folk judgments about work as well as those of the modern Western philosophical tradition I have been describing. But it may turn out to be the basis of a misleading ethical view. The ancient Greeks may well have been wiser than us, as it may be that labor should enjoy much less priority than our society gives it. Certainly many individuals are prone to exaggerating the importance of their labor, so much so that we have names for these pathological forms of prioritization: workaholism, careerism, and inadequate work-life balance.

Moving on from these qualifications, the idea that labor enjoys normative priority explains why folk judgments also associate labor with compensation, as well as acknowledgment and praise. This association is also grounded in the philosophical tradition's claim that labor is not just voluntary or intentional but effortful or skillful. In this way the folk conception suggests another heuristic: an activity can be characterized as a job to the extent we think it deserves compensation and acknowledgment. Yet here again, the claim requires qualification. It is at best an ideal of labor that it deserves reward. "Bullshit jobs," if there are in fact any such (Soffia et al. 2022), will seem undeserving to many—the moniker confirms the ideal that a job is one that deserves its pay. Conversely, there are many activities which do not receive compensation and acknowledgment but deserve them, in particular: domestic labor, care for dependents, affective labor, and working toward positive social change. While there may be arguments from distributive justice against remunerating some of these forms of work (Bergmann 2005, 133), the widespread concern that we should count these activities as labor reflects the growing social recognition that they are just as socially necessary and effortful as other forms of work. Meanwhile, the debate about whether executive compensation is excessive very often falls back upon arguments about whether executive managers perform the kind of labor that deserves such high pay (Moriarty 2005; Magnan and Martin 2019, 91–92; Sandberg and Andersson 2022, 765–66). Desert cannot be all there is to the debates about pay, but its centrality in that debate reflects the common belief that compensation is earned by both the importance of the job and the effort it requires.

In this way the folk conception of labor suggests a set of heuristics, suitably qualified, for counting an activity as labor: that we see the activity as being sufficiently important and requiring the sort of effort such that we should give it priority in our deliberation, and deem it deserving of acknowledgment and compensation. Data as Labor does not fare well against these heuristics: data contributions do not generally have the same kind of importance as other activities regarded as labor, and they are not in general as effortful. Consider a professor

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

entering some keywords into a search engine while researching a paper. While in searching they perform labor as part of their academic job, their search query and consequent actions also provide training data to the internet platform that provides the search results. Our question is whether generating data in this way is enough to make the activity work, or equivalently: is the professor working not only for their university but also for the platform?[9] Most people would not see the mere act of data generation as enjoying the same priority over other activities that, say, Johan's esport practice does. And while forming a search query and acting upon it is certainly not reflexive or instinctual behavior, it is also unlikely to count as so skillful or effortful that it deserves praise and compensation as a socially necessary activity. The algorithm that results from collecting such data is likely to have great social value, but each individual interaction between user and platform makes only a negligible contribution to that value.[10]

The upshot of this section is not to evaluate whether the folk conception is correct or not, though if the reader accepts it then they will likely doubt that data can be counted as labor. Rather, my account of the folk conception aims to show why many will think it nonsensical, or at least unpersuasive, to claim that generating user data *is* labor. Even for such skeptics, there remains a question of whether we should treat data *as* labor. An affirmative answer will need to reach for arguments that go beyond the heuristics of the folk conception. Others will likely dismiss the folk conception as inherently conservative. The intuitions about work surveyed here have largely been forged through European philosophical traditions and the experience of industrial era labor. Why should we judge a new activity enabled by digital technology according to this standard? For this kind of skeptic, the question remains what kinds of reasons should lead us to think such a novel activity is or is not labor. In what follows, we turn to the role labor plays in political economy to find alternate reasons for treating data as labor.

[9] Some may also describe the professor as "working for an algorithm" (Curchod et al. 2020). That paper studies eBay sellers who find themselves caught within a complex field of accountability relations to both the platform, which imposes conditions on access, and to customers, who exercise power through their ability to review sellers. While the authors do not define the phrase "working for an algorithm," they seem to have in mind the work of navigating this complex field as it is constructed by the platform's algorithm. The implication is not that we must think of work differently, but that we should think of who a platform participant *works for* differently (see especially pp. 665–66). This is a fascinating phenomenon, but one that is distinct from data contribution, though the latter is almost always algorithmically mediated. Here again we want to peel the data contribution away from other aspects of an interaction that already count as work.

[10] It would be easy to generate counterexamples against the universal version of the claim—that data contributions are never labor—given the great variety of online interactions that generate data. Sometimes individuals provide data in the form of survey or experimental results to platforms such as Mechanical Turk, or as part of a program of reinforcement learning by human feedback, and these contributions have the kind of social value, skillfulness, and effortfulness that qualify them as labor—a judgment supported by the fact that they are likely to be paid by the platform (or its users) for these interactions. But the point of this section has not been to argue that there are no transfers of data that count as labor on the folk conception. It is especially not concerned with data contributions that are already treated as labor, as is the case with the examples just mentioned. Rather, we are interested in whether those data contributions that are not already treated as labor are *in general* to be thought of as labor. The burden on the advocate of Data as Labor is to justify this generic claim, rather than to produce counterexamples against the rule that we should never do so.

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

3. THE POLITICAL ECONOMY OF LABOR

Labor occupies a special role in our practical deliberations about how to coordinate and regulate society, one that contrasts with the roles of capital and land, the other primary economic factors of production (Mankiw 2015, 22–24, 373–93). The thought that there is something distinctive about labor as a factor of production is reflected in neoclassical economics as well as more critical strands of political economy, and it is presupposed by modern systems of labor law. So a different approach to Data as Labor would ask not whether the folk conception judges that the activity of generating user data *is* labor, but rather whether that activity fits the political-economic role of labor. If it does, then that provides us with an argument for regulating it as labor.[11]

The political economy of labor must start with the insights of the neoclassical theory of distribution, whose textbook treatment identifies land, labor, and capital as the principal ingredients of the production of goods and services by firms (e.g. Mankiw 2015, 22–24, 373–93).[12] According to the theory, goods and services (sometimes described as commodities) are supplied by firms and demanded by consumers, while the factors of production are demanded by firms and supplied by consumers. The relative demand for and supply of these factors determine the income they earn (respectively rent, wages, and profit and interest), and therefore the distribution of income between households. The theory of supply and demand applies universally to these factors, but the factors themselves exhibit different characteristics and relationships that affect their respective equilibria. For example, land has a relatively fixed supply, unaffected by price. Another example: the purchase of labor-saving capital reduces the marginal product of labor, leading to an exogenous reduction in the demand for labor.

[11] There is a good deal of overlap between what the folk conception views as labor and the activities that are regulated as such by labor law and policy. The overlap is limited, however, in ways that sometimes reflect the limits of the folk conception as well as of labor law and policy. As an example, the important work of changing social norms through activism, bearing witness, and everyday conversation may well deserve to be seen as a kind of labor given the criteria we find in the folk conception. But it will be hard to view these as such through the lens of political economy, given that such activity is not ordinarily part of the process of economic production. This is just one reason why the political economic view should not be taken as a definition of labor, but rather as a normative theory about why we should regulate certain productive activities (of which platform-user interaction is certainly one) as labor.

[12] Economics textbooks deem it unnecessary to define land and labor. A typical description of capital is that it is "the stock of equipment and structures used for production" (Mankiw 2015, 387). This is but one side of a crucial ambiguity (Pistor 2019, 9–10; Hodgson 2015, 173–84). In practice, economists also use the word to identify the money and financial assets supplied by households to firms in order to invest in productive equipment (Mankiw 2015, 388), tracking the original and continued popular usage of the word to mean wealth generally. This leaves open whether capital includes human capital. Mankiw assumes that it does, but points out that human capital is an investment in the productivity of a particular person that accompanies that person results in increased wages rather than dividends or interest (2015, 397). Pistor agrees, but only because for her capital is "not a thing" but a legal construction through which an asset is subjected to universal, durable, and prior claims that can be converted into money—an individual can transform her labor into capital if they can give rise to such claims, perhaps by incorporating themself (2019, 10–11; cf. Hodgson 2015, 188). While this section begins with the assumptions of economic theory, it will ultimately pay close attention to the role that law plays in constructing labor, capital, and the other factors.

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

But the political economic characterization of labor cannot stop with distinguishing features and relationships described by the neoclassical theory. As Polanyi ([1944] 2001) writes, in a market economy land and labor have been transformed into "fictitious commodities"—"fictitious" because they are not produced for sale as commodities but pre-exist the market, and "commodities" because they do come to be bought and sold on the market. As he puts it, labor "is only another name for a human activity which goes with life itself, which in its turn is not produced for sale but for entirely different reasons" ([1994] 2001, 75). The contemporary economist may reasonably complain that welding automobile chassis or designing PowerPoint presentations are not activities that go with life itself but rather are elicited only by the incentives that are so well designed by the market. But Polanyi's point is that when the market allocates such labor activities, it also allocates the "physical, psychological, and moral entity" who "happens to be the bearer of this peculiar commodity" ([1994] 2001, 76). That is, when the market allocates labor, it also regulates the free activity of human beings.

What follows from the fact that labor is not just another factor of production but a "technical term for human beings" (Polanyi [1994] 2001, 79)? The most fundamental aim of labor regulation has been blunting whatever forces might lead to human subjection, most notably slavery, but including limits on child labor, the working day and working week, conditions on workplace safety, and minimum wage laws. The appropriate extent of these limits is controversial, but their rationale is clear. The employment relation can become a means of subjection because, for one, what is sold is not just any commodity but one that is so closely tied to the seller's well-being.

A second reason that the employment relation can become a means of subjection has to do with a structural feature of the labor market: a bargaining asymmetry that, while not a necessary feature of the labor market, is so typical that it justifies broad remedial measures such as the above limits on the employment relation as well as the powers that underwrite labor organizing and collective bargaining. The asymmetry is grounded in three features of typical employment relations: 1) nearly everyone prefers to sell their labor, but 2) buyers of labor have a systematically larger threat advantage than sellers; and 3) the scope of employment authority is necessarily open-ended and open to abuse. Each in turn is grounded in economic generalities that play an important role in our best economic theory: 1) a preference for income smoothing, 2) decreasing marginal cost of production, and 3) the transaction costs rationale for the firm. These features of economic life are not exceptionless, but they are sufficiently general that they underwrite a system of labor regulation that aims to mitigate the widespread effects of the asymmetry in the face of such counterexamples as the Silicon Valley engineer and star CEO who can name the price of their labor. Let's consider each generality in turn.

1) Most people aim to smooth their income over time, preferring a regular stream of income over large losses and gains. This preference may be due to risk aversion, which also explains the general demand for insurance. But it may also be due to the fact that most people have regular payments to make, such as rent or mortgage

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

payments, grocery bills, and so on. The easiest way for most people to guarantee themselves a relatively constant income is to sell their labor. Only those lucky few who own large amounts of land or wealth have a large enough stream of interest or rent income payments to subsist without laboring. An entrepreneur may seek to subsist by starting their own business (thus providing "sweat equity," which is really labor rewarded as capital), but most businesses fail, and even those that don't are vulnerable to making the occasional loss, which the entrepreneur must bear alongside other suppliers of capital. By contrast, employees have a contractual right to be paid no matter what their employer's statement of profit and loss. So most people (except a student with board or a person of independent means) will prefer the prospect of a fixed wage over the entrepreneurial gamble (Heath 2023).

2) Production processes typically face decreasing marginal costs, to a large extent because of returns to specialization, but also because as a firm grows it is more likely to invest in capital goods that increase efficiency. Most industries are somewhat concentrated for this reason, and so, unlike in the idealized world of the economics textbook, there is seldom an indefinite number of buyers of labor in a market. In fact, the average US labor market is highly concentrated amongst buyers of labor (Azar, Marinescu, and Steinbaum 2022). Moreover, a worker ordinarily cannot efficiently withdraw from the labor market and work for themselves, since they cannot produce the same goods as efficiently as a firm that exploits its economies of scale. Since it is generally much easier for a firm to continue somewhat efficiently when a worker leaves than for a worker to continue to earn income upon leaving a firm, most workers have little advantage in the way of a credible threat to exit when bargaining over the terms of employment.

3) It is now well established that firms exist in order to overcome transactions costs, such as the difficulties of coordinating production and avoiding holdouts (Coase 1937; Alchian and Demsetz 1972; Jensen and Meckling 1976; Williamson 1973, 2010). Firms therefore involve cooperative relations that can be contrasted with the competitive relations of the market (even taking into account the prevalence of the kinds of coordination between firms observed by Richardson 1972). The form of cooperation we find within the firm is typically somewhat hierarchical, and to a significant extent employers overcome the relevant transactions costs by exercising authority over their employees. While the scope and economic importance of employer authority is controversial, some go so far as to describe it as a kind of "private government" (Anderson 2017). What is clear is that employer authority is open-ended, even while constrained by the employment agreement. In part this is because language, and therefore legal documents like the employment contract, are inevitably open-textured and so do not determine every issue within their scope (Hart [1961] 2012). But most importantly, employment contracts must be incomplete in order to allow firms to be flexible in the face of changing market conditions and in order to enable employers to solve the transactions costs problems that give rise to firms in the first place.

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

Together, these three features of the general market for labor establish a bargaining asymmetry that can have unfortunate consequences for workers. Most individuals must participate in the labor market in order to adequately smooth their income, and most such sellers of labor have less threat potential than buyers. Thus employers tend to set the terms of the employment relationship. These terms tend to be open-ended in order to overcome the transactions costs that undermine production outside of the firm, and employers exercise greater authority than is typical in other cases of arm's length contracting, such as are found in markets for commodities. The result is that workers have characteristic vulnerabilities to harsh terms of employment and abuse of employer discretion. Much of labor law can be interpreted as an attempt to either mitigate the bargaining asymmetry that results in such vulnerability, by permitting or even enabling workers to organize across firms or even industries and bargain collectively; or to mitigate the vulnerability by establishing minimum conditions of employment, such as minimum wage laws, workplace safety conditions, antidiscrimination and anti-harassment measures, and regulation of working hours.

The problems that labor law sets out to fix contrast starkly with the problems that the laws of capital formation—in the form of corporate governance laws and securities regulations—aim to fix. Corporate governance law and securities regulation both aim to protect investors, that is, owners of capital, but they intervene at different stages (Park 2017). Securities regulation protects investors when they buy ownership stakes in companies, and seeks to ensure transparency about what they are buying. It does so primarily through disclosure requirements and restrictions on fraud, manipulation, and insider trading—all of which aim to mitigate the characteristic information asymmetry between corporate insiders and traders making use of externally available information (Goshen and Parchomovsky 2006). This information asymmetry is a distinct problem from the bargaining asymmetry to which labor law attends. Both are asymmetries that affect bargaining outcomes, but information asymmetries are best mitigated by disclosure while bargaining power asymmetries are best mitigated by strengthening bargaining power.

In contrast to securities regulation, corporate governance law protects investors while they are owners, and seeks to mitigate the agency problem that arises when owners delegate decision-making power to managers (Jensen and Meckling 1976). The characteristic threat is that corporate insiders, particularly managers, expropriate outsider investors, either by appropriating profits to themselves, or selling corporate assets or additional equity stakes so as to dilute the rights of current investors (La Porta et al. 2000). This threat is again distinct from the problem targeted by labor law. In the case of corporate governance, insiders exercise power over outside investors because they have decision-making power and asymmetric information about how they exercise it. In contrast, employers exercise power because they have the more credible threat when bargaining with prospective and current employees, for the reasons discussed above. Thus corporate governance focuses not on the bargaining power of shareholders, but on reporting rules, fiduciary obligations to shareholders, and prohibitions on self-dealing by managers—remedies for the characteristic

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

principal-agent problems and information asymmetries that arise between corporate insiders and outsiders.

The distinctive economic characteristics of labor and capital do not entail that they are mutually exclusive. Indeed, several important factors of production can be characterized as located along a spectrum between classical forms of labor and capital, thus meriting multiple or hybrid forms of regulation. Consider, as an example, the case of human capital, that is, the skills and intellectual abilities that improve the productive capacity of a worker. Human capital accompanies the worker, and so where there is no slavery it cannot be owned by the firm. Yet firms invest in human capital by hiring skilled workers or investing in training. When a firm engages in such activity, it faces an information asymmetry insofar as it has less information about the worker's abilities, motivation, and loyalty than the worker does. It will therefore attempt to protect its investment through contractual terms such as non-compete clauses, which are justified (to the extent that they are) by the moral hazard and signaling problems facing participants in the human capital market (Frye 2020)—problems that are characteristic of the political economy of capital rather than of labor.

Our question about data contributions is whether they share enough features of the political economy of labor that they should be treated as labor by the law. Given the typical asymmetry of bargaining power between an employer and its workers, and the way this asymmetry distorts the market mechanism, labor policy aims at bolstering worker bargaining power. This is why labor law enables and protects activities that seem antithetical to the competitive market, such as unionizing, strike action, and collective bargaining. To the extent a different factor of production is subject to this characteristic bargaining asymmetry of labor, it favors remedying the bargaining asymmetry by making use of similar strategies. If we think that data contributions are characterized by asymmetric bargaining power of the kind found in the labor market, we should embrace proposals such as the creation of data unions (Arrieta-Ibarra et al. 2018) and data "strikes" and similar collective actions by data contributors (Vincent, Hecht, and Sen 2019; Kulynych et al. 2020; Vincent et al. 2021). This line of argument would be compatible with the claim that data also shares some aspects of the political economy of capital (as presupposed by Kim et al. 2021 in their argument that data subjects should be treated as investors), and should be placed on a spectrum between capital and labor just as human capital is. But in what follows I will suggest that the labor-like aspects of data contributions are a more fundamental problem, which motivates our focus on the claim that data contributors deserve similar regulatory protections to workers.

4. THE POLITICAL ECONOMY OF DATA

Data contributions exhibit some of the features of the political economy of labor just described. The problems that arise with platforms' storage and usage of personal data are largely due to the poor bargaining position of data contributors. For example, data contributors are typically unable to control the subsequent uses of

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

their data by a platform, since they are unlikely to be in a position to bargain for terms of service that guarantee such control.

This situation may also seem reminiscent of the political economy of capital. That is, a user of a social media website like Facebook is often unaware of what Meta will do with their data (will they allow a third party to use it in psychological profiling for an electoral campaign?), and this can seem analogous to a shareholder of Meta being unaware of how the company's managers deploy their capital. But in the latter case the information asymmetry problem that arises is inherent to the separation of ownership and control that is fundamental to the structure of the modern firm, in which shareholders must delegate decision-making power to managers. In contrast, the information asymmetry that arises in the case of the user-platform relation is not inherent to that relation, but rather is due to the fact that users with low bargaining power grant to platforms the power to reuse their data. An alternative arrangement in which platforms must provide detailed information about data usage and ask users for permission for further reuse is compatible with the user-platform relationship, and might be the outcome of a bargaining situation in which users have more power. So while problems that are analogous to the political economy of capital arise in the case of data contributions, the labor-like bargaining situation appears to be the more fundamental problem.

The position of data subjects is indeed much like that of labor market participants, which we characterized by the facts that (L1) nearly everyone sells their labor, but (L2) sellers of labor have a systematically smaller bargaining advantage than buyers have; and (L3) the resulting employment relationship inevitably involves open-ended authority that is open to abuse. Similarly, (D1) nearly all internet users expose their data online, but (D2) individual users are dispensable and have a systematically smaller bargaining advantage than the platforms they interact with; and (D3) the resulting relationship between user and platform is inevitably open-ended and open to abuse. Let's consider these by turn.

(D1) I'll limit the argument to active internet users—though arguably everyone uses the internet, whether they are active users or not, in virtue of the fact that they use medical, transport, and government administrative services that store data about them online. Those who are active users typically interact with a variety of platforms for search, shopping, news, file storage, streaming film and music, and social networking that are free or cheaper than they would otherwise be because their services are subsidized by advertising. But advertisers are attracted to these platforms because of the promise that users can be identified, profiled, and more profitably targeted. This promise relies on the collection of vast amounts of data about both individual users and the traits they share with others (Hwang 2020). So the economic relationship between an internet user and a platform is very often based on the user's consensual transfer of data.[13]

[13] Even users who pay for services often transfer data as part of the relationship, and in these cases the transfers of data can be a significant part of the platform's business model. And even in cases in which the user

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

(D2) It is true that users have the ability to exit by avoiding particular services, as large numbers of users sometimes do in protest (Lopez 2017; Li et al. 2019). Nonetheless, an individual user's suspension of data contributions to a platform does very little to affect the platform's algorithm or training dataset. Machine-learning algorithms make use of vast amounts of data—"big data" is a relative rather than absolute term, indicating that a dataset contains close to all the relevant data rather than a sample (Mayer-Schönberg and Cukier 2013, 29). But internet platforms exhibit significant network externalities, so that their usage follows a power law i.e. "winner takes most" (Faloutsos, Faloutsos, and Faloutsos. 1999). These platforms exercise network power: monopolistic power that is derived from network effects (Grewal 2008). As a result, the marginal value of any individual's data to one of the platforms is very close to zero, since it is likely to account for a tiny amount of the accuracy of the algorithm and the popularity of the platform. So there is no credible threat of exit that an individual user poses to a platform, and the user has very little bargaining power over terms of service.

(D3) Data has "option value" for a platform, meaning that it is valuable for its potential to be re-used for other purposes, or merged with other datasets for additional value (Mayer-Schönberger and Cukier 2013, 104). Therefore a platform is likely to retain data that it has obtained where it can, and most platforms impose terms that allow it to reuse this data in the future or share it with third parties (Venturini et al. 2016, 75–79). This opens users up to vulnerabilities characteristic of open-ended relationships. In particular, there is the danger that data will be leaked, or reused in ways that do not benefit users and to which they would not otherwise have consented. The Cambridge Analytica scandal was a memorable illustration of the general point, with data collected from Facebook for academic purposes subsequently repurposed for the commercial purpose of profiling potential swing voters by the Trump campaign (Cadwalladr 2018).

This analysis illuminates some other pathologies of the data economy. *Ubiquitous surveillance* results when users accept terms that allow platforms to track them across the internet by installing cookies and aggregating data shared by other services. It is often assumed that the reason for this is that users don't understand the terms of service to which they assent (e.g. Fiesler and Bruckman 2014; Robinson and Zhu 2020). While the EU privacy regime presupposes that notice and consent requirements—in particular with respect to third-party services that track online behavior by way of cookies—will lead to greater user autonomy, the result has

does not willingly (as far as assent to terms of service counts as willing) part with their data, they might "leak" data: for example, an adversary can infer one's sexual orientation or party affiliation from one's Facebook likes (Kosinski, Stillwell, and Graepel 2013), or identify a suspect by using DNA uploaded by relatives to a genealogy website (Kearns and Roth 2019, 54–56), or learn novel data by aggregating supposedly separate sources of online information (Krishnamurthy and Wills 2009) and de-anonymizing records in large datasets (Narayanan and Schmatikov 2008).

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

instead been user frustration with notices, and continued acquiescence in the existing surveillance regime (Utz et al. 2019). This is because users' ignorance is not the real source of the problem. The problem is instead that, even when users do know and understand the proffered terms of service, they have little power to negotiate over these terms, and they find "privacy self-management" elusive (Solove 2012).

An emerging problem of large language models and other generative AI systems is their *appropriation of intellectual property as training data* (Samuelson 2023). The problem arises because trainers of machine learning models have been able to access large amounts of intellectual property online for free. The copyright in blog posts, social media comments, uploaded photos, and other user-created content vests in the creator upon creation. But the terms of service between user and platform typically licenses or even transfers copyright in this material to the platform, often without the full knowledge of the user (Fiesler, Lampe, and Bruckman 2016). The unfavorability of such terms of service again result from the bargaining power asymmetry between users and platform.

The analogy between the structural positions of data and labor is also supported by Zuboff's (2019) work on "surveillance capitalism" (or more generally, "data capitalism," see West 2019). Zuboff makes the explicit claim that internet user's data contributions are not labor because there is no actual payment for their data (2019, chapter 3). The claim is consistent with the folk conception of labor, and perhaps it is correct insofar as we seek to impose a definition of labor. But our question about the political economy of data asks how we should manage and regulate the current business model of internet platforms, given the structural position of users. Here Zuboff's critique is consistent with the claims of Data as Labor, as she characterizes internet users as occupying a position that is structurally equivalent to labor from a Marxist perspective. Users are the source of raw materials for the "behavioral value reinvestment cycle" (Zuboff 2019, chapter 3), in which data is collected from users and then used to train the algorithms of platform services for the sake of creating improved products and services that are likely to engage users so that more data can be collected. The business model of a platform is to sell "behavioral futures" (predictions of user behavior) to advertisers and others who wish to engage users. But, as Zuboff notes, this means that users are not just a source of raw material, they are the source of the surplus that platforms collect as profits (2019, chapter 5).

Described in this way, data contributors occupy the same role that traditional forms of labor do in the Marxist model. Marxists think that labor is necessarily exploited within a capitalist system (Cohen 1979, chapter 7; Elster 1985, chapter 4). The reason for this is that the surplus value that workers are unable to capture, due to their weak bargaining power, is the source of the capital accumulation that increases the returns to capital and net present value of capital to its owners. Zuboff's argument suggests that the weak position of data subjects is similarly the source of the capital accumulation providing the large returns to internet platforms under surveillance capitalism. This diagnosis of the position of workers and data contributors, with its emphasis on exploitation as the source of profitability, is stronger than the one I have

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

presented, but it is compatible with my claim that the political economy of data is analogous to that of labor.

5. POLICY ALTERNATIVES

The foregoing argument is that the political economy of data—the way in which power is allocated when it is a factor of production—is similar to that of labor, and that any similarities to the political economy of capital are in fact grounded in its labor-like position. Does this analogy entail that data should inherit the protections of labor law, such as the power to strike and to form unions (Vincent, Hecht, and Sen 2019; Arrietta-Ibarra et al. 2018)? Policy alternatives must be evaluated comparatively and with respect to the policy goal of mitigating the problems that arise from user-platform bargaining power asymmetry, such as unfavorable terms of service, ubiquitous surveillance, data leakage, and misappropriation of intellectual property.[14]

Consider then, from the political economy perspective outlined above, the major solutions that have been suggested for managing the user-platform relationship:

- *Data Strikes and Data Leverage.* Vincent et al. (2021) propose three forms of "data leverage" in which people use the dependence of platforms upon their data to urge platforms to change their policies: data strikes, that is, withholding data by suspending use of a service or using it only through a privacy filter (see also Arietta-Ibarra 2018, Li et al. 2018, Vincent, Hecht, and Sen 2019); data poisoning, that is, inserting inaccurate or harmful data into a data-dependent algorithm; and conscious data contribution to better services so as to enhance their competitive advantage. Data strikes are the most reminiscent of the traditional tools of labor bargaining, but all three face an equally familiar challenge: collective action at scale is required in order to exert any real pressure on the platform. In addition, as the authors point out, while users may suspend use of a platform, in some cases they can only effectively strike by having the platform delete their data and refrain from using models trained on their data (Vincent et al. 2021, 222). This requires regulatory support in the form of a right to be forgotten, as found in Article 17 of the EU's General Data Protection Regulation, as well as enforcement that data-dependent models do not engage in "data laundering," that is, continuing to use supposedly deleted data (2021, 222). But such a right does not support the collective action problem, which requires in addition that users be able to coordinate with each other on the terms and scope of a data strike.

[14] There are many other pathologies that are associated with the data economy, such as filter bubbles and disinformation (Sunstein 2017, but cf. Kitchens, Johnson, and Gray 2020), rising anxiety and depression (Twenge, Martin, and Campbell 2018), and the legitimacy challenges of content moderation practices that have sometimes been described as "shadow banning" (Savoleinen 2022, cf. Nicholas 2023 decrying this terminology)—but these are not meant to be explained by Data as Labor, and it is quite likely that they will require different policy fixes that attend to questions about addictive products and free speech and information policy.

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

- *Data Unions.* Data unions have been proposed by advocates of Data as Labor (Posner and Weyl 2018), and one data union has already been established in the Netherlands (De Datavakbond 2018; Sterling 2018). A union that represents a sufficient number of data users of a particular platform could coordinate adversarial pressure through data strikes and other forms of data leverage, and could also facilitate cooperative bargaining and make credible demands of a platform. But the union would need to be large enough, and have enough commitment from its members, to credibly threaten to withdraw data. It would also need to be supported by a regulatory framework that enables it to enforce such threats, such as a right to be forgotten and restrictions on data laundering.

- *Data Trusts and Data Stewardship.* A data trust is an independent body that stewards data on behalf of data subjects, deciding who has access to it, the conditions of access, and who may benefit from access (Hardinges et al. 2019) while exercising a fiduciary responsibility toward data subjects (Delacroix and Lawrence 2019). It is occasionally likened to a union, in that it has the power to stand up for the rights of its members (Ruhaak 2021). One difference between a data union and a data trust, at least insofar as a data trust uses the legal analogy of a traditional trust, is that the data trust represents data rather than data subjects. That is, it governs a discrete body of data and the accompanying rights, rather than the ongoing relation between a data subject and a platform, raising questions about how effective it might be at advocating for changes in its future treatment of data subjects. Another difference is that the data trust can take on a wide variety of governance models, and so democratic or cooperative governance by users is not entailed where the trust is not explicitly understood as a "civic" data trust (McDonald 2019). Data trusts face similar challenges to data unions: they are unlikely to be effective without legal infrastructure enabling them to enforce the suspension of data use (Delacroix and Lawrence 2019, 243–44).

- *Personal Data Ownership or Equity-like Claims.* The GDPR is widely seen as a move toward data ownership since it recognizes rights of portability and exclusion that would ordinarily accompany ownership. Some politicians and regulators have called more explicitly for regulating data through traditional forms of ownership rights (Padova 2021, 43). The proposal that we see data as capital may also appear to entail something like personal ownership: in fact, the primary advocates of data as capital recommend an equity-like claim on the basis of what they see as the data subject's underlying property interest in data (Kim et al. 2021, 80–85). Yet as we have seen through our bargaining analysis, individual ownership rights cannot be used to exert any power over platforms anymore than individual workers' contractual rights and self-ownership rights—at least, not without institutional facilitation of coordination and enforcement of the type provided by labor law. And given that ownership is alienable, then without the stewardship of a trust or the coordination of a union it is likely that users will simply sell or give away their property rights, resulting in a situation much like the status quo. This much is confirmed by the fact that the GDPR departs from the ownership model in according inalienable rights to data subjects (Custers and Malgieri 2022). As

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

Yann Padova, Secretary General of France's data protection agency CNIL, has put it, the basic mistake is that recommendations in favor of ownership "[assume] a debatable equivalence between the data subject's control over the use made of his or [her] data and their ownership" (Padova 2021, 66). Ownership does not entail control, and control may be facilitated by measures other than ownership. The Data as Labor perspective suggests what those measures might look like.

- *Open Data and Data Sharing.* The idea of open data is that data should be available to be shared by all, rather than being monopolized. It is primarily a recommendation for governments to make their data accessible to the public in the interests of promoting transparency and preventing corruption (McDermott 2010, Zuiderwijk and Janssen 2014). It has also been occasionally contemplated as a recommendation for private firms, as a way of increasing competition and preventing the economic domination of large platforms (Kitchin 2014, chapter 3; Padova 2021, 54–56). Perhaps this could be of indirect benefit to data subjects insofar as the participation of a wider array of firms and states in the data economy could lead to a wider range of terms of usage, some of them more favorable to data subjects than is currently the case. But it is unclear why we should be so optimistic about the results of more competition at the level of providers of algorithms. Indeed, the analysis of the previous section shows that the bargaining asymmetry is not due to monopolistic organization of the data economy, but rather to structural features of data as a factor of production. Even where there are many firms competing for users' data and patronage, it will still be the case that any particular platform need not fear uncoordinated action by individual users. In addition, economic modeling shows that the effects of data sharing on data collection are suboptimal relative to data unions, since firms incur the costs of original data acquisition but cannot internalize all of the benefits (Cong and Mayer 2023). This means there would be less data collection from data subjects, and therefore less exploitation, but at the cost of fewer of the existing benefits of the status quo (e.g. improved internet platform services) as well as none of the envisaged benefits of data unions (e.g. compensation, increased control over data usage).

- *Antitrust.* Antitrust aims to break up collusive industries and prevent anticompetitive behavior, and it has been a favored remedy amongst policymakers who believe that internet platforms have gotten too big and wield their power ("platform power") unfavorably against users (Khan 2017, 2019). But as in the case of data sharing, it is unclear why this is the right fix for the particular kind of asymmetry facing users. By promoting competition between platforms there may well be the conceptual possibility of more favorable terms for users, but the structural bargaining power characterized in this paper is not due to a platform having a monopolistic or oligopolistic position in a market or industry. It is true that the asymmetry is partly due to the network effects of a platform, and this network effect would be mitigated by "breaking up" a platform that participates in several markets at once. But network effects extend wherever an internet platform is able to use the data of its users to improve its services, and these would obtain no matter how competitive a market is made, short of prohibiting the use of user data

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

entirely. So again, as in the case of data sharing, this recommendation may improve the terms for data subjects but only at the cost of the benefits made available by platforms.

- *Universal Basic Income (UBI).* UBI involves a regular modest payment by the state to all citizens independent of welfare eligibility or need (Standing 2010, 3). It has often been proposed as a remedy for the loss of work that is expected to follow significant automation, as well as a way to redistribute the benefits of automation fairly (Danaher 2019, 101). As such it may be thought that it is the proper recommendation of Data as Labor, given the important role data is playing in accelerating automation. Under UBI we would all be seen as contributors to and beneficiaries of machine-learning algorithms and their productive efficiency, with the state being responsible for payments because platforms have no incentive to return surplus value to us (at least in the absence of any of the above solutions such as data unions). This may be the simplest solution of all those considered, since it would avoid difficult questions such as who to count as a data subject (for the sake of data unions, data trusts, and data ownership) or how to define a market (for the sake of data sharing and antitrust)—though it raises challenges from the perspective of intergenerational justice (should the current generation be paid to compensate for advantages accruing to future generations?), and economists' concerns that UBI reduces investment in future generations (Daruich and Fernández 2021).[15]

A major disadvantage of the UBI proposal is that it places a large burden on the state to remedy the consequences of private actions by internet platforms. Even if the tax were drawn entirely from platform profits, administering tax and redistribution efforts takes a complicated and costly administrative effort. And it is probably not politically feasible to draw such an income from profits: first, profits fluctuate in a way that is politically unappealing for income grants; second, corporations are likely to change accounting methods and investment choices in order to avoid the tax. UBI also does not directly mitigate the platform-user bargaining power asymmetry. If we assume that without the bargaining asymmetry users would demand compensation for their data, then we can see UBI as a substitute for such compensation. But it does

[15] The idea that basic income may compensate unrecognized labor has also arisen in the domain of care work. There the central problem is that care for dependents is often conducted by adult women in families, in particular wives, but that their work is neglected as such and goes without adequate recognition and compensation. One apparent solution is for the state to compensate such work through a grant to families. But whether this is an adequate solution or not depends upon whether justice also requires that the expected gender roles change (e.g. Gheaus 2008; Hirschmann 2016), and whether it requires that the state be neutral between households with dependants and households without (Bergmann 2005, 143). Once we have established a criterion for justice, an additional difficulty is establishing how a basic income can be expected to impact gender norms and family structure (O'Reilly 2008). Because of these specificities, the debate about care work is not a perfect analogy for that about data regulation, and it counsels us to be wary of thinking of UBI as a simple solution and to consider second order effects. One such second order effect is the incentive structure for data acquisition—here the economic work on data unions suggest that they would make for more economically optimal decisions by platforms and data subjects (Cong and Mayer 2023).

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

not in fact change the asymmetry, and so it is unlikely to improve the favorability of other terms of the user-platform relationship.

Which of the above is the best policy strategy will depend ultimately on the empirical consequences as well as political feasibility of particular implementations. I have speculated above about how these may pan out, but further research that is beyond the scope of this paper will be determinative. Here I have primarily aimed to show that Data as Labor is a promising way to understand the political economy of data and evaluate the policy responses. Supposing the above speculative evaluation is correct, and the most suitable response to the bargaining power asymmetry that characterizes the data subject-platform relationship involves such labor-like institutions as data unions and data strikes, then the question that remains for policy scholarship is what sort of regulatory reforms would facilitate data unions and data strikes. That further question must be cognizant of important differences between the internet and the workplace, such as the fact that data strikes are only possible given a robust right to be forgotten or a right to data portability that is accompanied by robust enforcement in order to avoid data laundering.

CONCLUSION

Is data labor? The question is intelligible once we consider that data should be viewed pragmatically in terms of the interactions between users and platforms that give rise to practices of collecting, analyzing, and using user data. But what is labor? One way to answer this question is definitional, and applies the principles implicit in our folk conception of labor. The folk conception emphasizes individual and social importance and effortfulness of labor, traits which justify the kind of priority we give our work and our assumption that it deserves praise and compensation. Whether data contributions meet these criteria will depend upon the context and type of data, but many of the paradigm cases that Data as Labor has in mind, such as the behavioral data gathered through interactions with social media, are unlikely to strike us as labor-like.

An alternate way to understand labor is functional, attending to the political economic role in order to understand why certain productive activities are regulated as labor rather than as other factors of production. The political economic role of data is indeed much like that of labor, since both are characterized by bargaining power asymmetry. In particular, data transfers inevitably arise from internet usage, but the value of a particular user's data is low while the value of a platform to a user is higher the more users it has. Thus, a user has little bargaining advantage relative to the platform they interact with, much as an employee has little bargaining power in interactions with the owners of capital. This analysis recommends that we treat data as labor from a regulatory perspective, and suggests that we borrow from the institutions of labor law in looking for ways to mitigate the pathologies of the user-platform relationship. In particular, the idea of a data union or data strike is worth further investigation, particularly with respect to the regulatory infrastructure these ideas require.

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

I have not investigated here whether there are further analogies between data and labor. One avenue for further research would be whether the theory of alienation can be applied to data in the same way as it is often applied to labor. Alienation is best understood as a subjective experience of loss of meaning or something else crucial to agential freedom (Jaeggi 2014). The internet user whose data has been acquired and used to shape a music recommendation algorithm finds an aspect of their identity abstracted, aggregated with the statistical preferences of unknown others, and externalized in the recommendations of the algorithm. This user may be forgiven for feeling a loss of enchantment about the idea of the self, anxiety about their own individuality, or simply a loss of control over their preferences and other agential attitudes. But it is also possible that this is simply an overly fearful response to novel technology that is destined to become an unremarkable part of the fabric of our lives. The data/labor analogy can prompt this sort of question, but an illuminating answer will depend on further examining our actual experiences of the data economy in light of a full theory of alienation.

Acknowledgments

This paper was presented at the 2022 Society for Business Ethics Annual Conference in Seattle, and some unrecognizable prior iterations of some of its themes were presented at the 2018 Amsterdam Privacy Conference, the 2019 Association for Practical and Professional Ethics Annual Conference in Baltimore, the 2021 workshop on Labor Justice and Transformation of Work held online by the Universitat Pompeu Fabra Law and Philosophy Research Group, and the 2022 workshop on New Foundations of Labor Law held in Toronto. I thank all these audiences, as well as Eric Posner, for their detailed comments and criticisms. I am especially grateful to Alejo Sison for encouraging me to write the current version of the paper and graciously allowing me to submit it to this journal. The paper was vastly improved by the generous and incisive recommendations of Frank den Hond and three anonymous reviewers.

REFERENCES

Alchian, Armen A., and Harold Demsetz. 1972. "Production, Information Costs, and Economic Organization." *American Economic Review* 62: 777–95.

Anderson, Elizabeth. 2017. *Private Government*. Princeton, NJ: Princeton University Press.

Applebaum, Herbert A. 1992. *The Concept of Work: Ancient, Medieval, and Modern*. Albany: State University of New York Press.

Arendt, Hannah. 1958. *The Human Condition*. Chicago: University of Chicago Press.

Arrietta-Ibarra, Imannol, Leonard Goff, Diego Jiménez-Hernández, Jaron Lanier, and E. Glen Weyl. 2018. "Should We Treat Data as Labor? Moving Beyond 'Free'." *AEA Papers and Proceedings* 108: 38–42.

Azar, José, Ioana Marinescu, and Marshall Steinbaum. 2022. "Labor Market Concentration." *Journal of Human Resources* 57: S167–S199.

Beaulieu, Anne, and Sabina Leonelli. 2022. *Data and Society: A Critical Introduction*. London: SAGE.

Bedingfield, Will. 2023. "Why Hollywood Really Fears Generative AI." *Wired*, June 2. https://www.wired.com/story/hollywood-actors-artificial-intelligence-performance/.

Bergmann, Barbara R. 2005. *The Economic Emergence of Women*. 2nd ed. New York: Palgrave Macmillan.

Bhargava, Vikram R., and Manuel Velasquez. 2021. "Ethics of the Attention Economy: The Problem of Social Media Addiction." *Business Ethics Quarterly* 31 (3): 321–59.

Boisot, Max, and Agustí Canals. 2004. "Data, Information, and Knowledge: Have We Got It Right?" *Journal of Evolutionary Economics*, 14: 43–67.

Cadwalladr, Carole. 2018. "'I Made Steve Bannon's Psychological Warfare Tool': Meet the Data War Whistleblower." *The Guardian*, March 18. https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump.

Chun, Wendy H. K. 2021. *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition*. Cambridge, MA: MIT Press.

Ciulla, Joanne. 2001. *The Working Life: The Promise and Betrayal of Modern Work*. New York: Three Rivers.

Coase, Ronald H. 1937. "The Nature of the Firm." *Economica* 4 (16): 386–405.

Cohen, Gerald Allan. 1979. *Karl Marx's Theory of History*. Princeton, NJ: Princeton University Press.

Cong, Lin William, and Simon Mayer. 2023. "Data Union and Regulation in a Data Economy." NBER Working Paper 30881, National Bureau of Economic Research, Cambridge, MA. http://www.nber.org/papers/w30881.

Creamer, Ella. 2023. "Authors File a Lawsuit against OpenAI for Unlawfully 'Ingesting' Their Books." *The Guardian*, July 5. https://www.theguardian.com/books/2023/jul/05/authors-file-a-lawsuit-against-openai-for-unlawfully-ingesting-their-books.

Curchod, Corentin, Gerardo Patriotta, Laurie Cohen, and Nicolas Neysen. 2020. "Working for an Algorithm: Power Asymmetries and Agency in Online Work Settings." *Administrative Science Quarterly* 65 (3): 644–76.

Custers, Bart, and Gianclaudio Malgieri. 2022. "Priceless Data: Why the EU Fundamental Right to Data Protection Is at Odds with Trade in Personal Data." *Computer Law & Security Review* 45: 105683.

Danaher, John. 2019. *Automation and Utopia: Human Flourishing in a World without Work*. Cambridge, MA: Harvard University Press.

Dargis, Manohla. 2023. "Do Studios Dream of Android Stars?" *New York Times*, September 14. https://www.nytimes.com/2023/09/14/movies/artificial-intelligence-hollywood.html.

Daruich, Diego, and Raquel Fernández. 2021. "Universal Basic Income: A Dynamic Assessment." NBER Working Paper 27351, National Bureau of Economic Research, Cambridge, MA. http://www.nber.org/papers/w27351.

De Datavakbond. 2018. "Het is officieel." May 24. https://datavakbond.nl/het-is-officieel/.

Delacroix, Sylvie, and Neil D. Lawrence. 2019. "Bottom-up Data Trusts: Disturbing the 'One Size Fits All' Approach to Data Governance." *International Data Privacy Law* 9 (4): 236–52.

Deranty, Jean-Philippe. 2022. "Post-Work Society as an Oxymoron: Why We Cannot, and Should Not, Wish Work Away." *European Journal of Social Theory* 25 (3): 422–39.

Elster, John. 1985. *Making Sense of Marx*. Cambridge: Cambridge University Press.

Estlund, Cynthia. 2005. *Working Together: How Workplace Bonds Strengthen a Diverse Democracy*. New York: Oxford University Press.

Etzioni, Amitai. 1998. "A Communitarian Note on Stakeholder Theory." *Business Ethics Quarterly* 8 (4): 679–91.

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

Faloutsos, Michalis, Petros Faloutsos, and Christos Faloutsos. 1999. "On Power-Law Relationships of the Internet Topology." *ACM SIGCOMM Computer Communication Review* 29 (4): 251–62.

Favaretto, Maddalena, Eva de Clercq, and Bernice Simone Elger. 2019. "Big Data and Discrimination: Perils, Promises and Solutions: A Systematic Review." *Journal of Big Data* 6: 12.

Fiesler, Casey, and Amy Bruckman. 2014. "Copyright Terms in Online Creative Communities." In *CHI'14 Extended Abstracts on Human Factors in Computing Systems* (CHI EA'14*)*, 2551–56. New York: Association for Computing Machinery.

Fiesler, Casey, Cliff Lampe, and Amy S. Bruckman. 2016. "Reality and Perception of Copyright Terms of Service for Online Content Creation." In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW'16*)*, 1450–61. New York: Association for Computing Machinery.

Floridi, Luciano. 2005. "Is Semantic Information Meaningful Data?" *Philosophy and Phenomenological Research* 70 (2): 351–70.

Flyverbom, Mikkel, Ronald Deibert, and Dirk Matten. 2019. "The Governance of Digital Technology, Big Data, and the Internet: New Roles and Responsibilities for Business." *Business and Society* 58: 3–19.

Flyverbom, Mikkel, and Anders K. Madsen. 2015. "Sorting Data Out: Unpacking Big Data Value Chains and Algorithmic Knowledge Production." In *Die Gesellschaft der Daten: Über die Digitale Transformation der Sozialen Ordnung*, edited by Florian Süssenguth, 123–44. Bielefeld, Germany: transcript.

Flyverbom, Mikkel, and Glen Whelan. 2019. "Digital Transformations, Informed Realities, and Human Conduct." In *Human Rights in the Age of Platforms*, edited by Rikke Frank Jørgensen, 53–72. Cambridge, MA: MIT Press.

Folbre, Nancy. 2008. *Valuing Children: Rethinking the Economics of the Family*. Cambridge, MA: Harvard University Press.

Freeman, R. Edward. 1984. *Strategic Management: A Stakeholder Approach*. Boston: Pitman.

Freeman, R. Edward. 1994. "The Politics of Stakeholder Theory: Some Future Directions." *Business Ethics Quarterly* 4 (4): 409–21.

Frenkel, Sheera, and Stuart A. Thompson. 2023. "'Not for Machines to Harvest': Data Revolts Break Out against AI." *New York Times.* July 1. https://www.nytimes.com/2023/07/15/technology/artificial-intelligence-models-chat-data.html.

Frye, Harrison. 2020. "The Ethics of Noncompete Clauses." *Business Ethics Quarterly* 30 (2): 229–49.

Geuss, Raymond. 2021. *A Philosopher Looks at Work*. Cambridge: Cambridge University Press.

Gheaus, Anca. 2008. "Basic Income, Gender Justice and the Costs of Gender-Symmetrical Lifestyles." *Basic Income Studies* 3 (3): 8.

Gheaus, Anca, and Lisa Herzog. 2016. "The Goods of Work (Other Than Money!)." *Journal of Social Philosophy* 47 (1): 70–89.

Gitelman, Lisa. 2013. *"Raw Data" Is an Oxymoron*. Cambridge, MA: MIT Press.

Goshen, Zohar, and Gideon Parchomovsky. 2006. "The Essential Role of Securities Regulation." *Duke Law Journal* 55: 711–82.

Graeber, David. 2018. *Bullshit Jobs: A Theory*. New York: Simon & Schuster.

Grewal, David S. 2008. *Network Power: The Social Dynamics of Globalization*. New Haven, CT: Yale University Press.

Grynbaum, Michael M., and Ryan Mac. 2023. "The Times Sues OpenAI and Microsoft over A.I. Use of Copyrighted Work." *New York Times*, December 27. https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html.

Hardinges, Jack, Peter Wells, Alex Blandford, Jeni Tennison, and Anna Scott. 2019. *Data Trusts: Lessons from Three Pilots.* Open Data Institute, April 14. https://theodi.org/news-and-events/blog/odi-data-trusts-report/.

Harris, Todd. 2003. "Data Models and the Acquisition and Manipulation of Data." *Philosophy of Science* 70 (5): 1508–17.

Harrison, Jeffrey S., and Andrew Wicks. 2013. "Stakeholder Theory, Value, and Firm Performance." *Business Ethics Quarterly* 23 (1): 97–124.

Hart, H.L.A. (1961) 2012. *The Concept of Law*. 3rd ed. Oxford: Clarendon Press.

Heath, Joseph. 2023. *Ethics for Capitalists: A Systematic Approach to Business Ethics, Competition, and Market Failure*. Altona, MB: Friesen Press.

Heath, Joseph, Jeff Moriarty, and Wayne Norman. 2010. "Business Ethics and (or as) Political Philosophy," *Business Ethics Quarterly* 20 (3): 427–52.

Hegel, G. W. F. (1820) 1991. *Elements of the Philosophy of Right*. Edited by Allen W. Wood, translated by H. B. Nisbet. Cambridge: Cambridge University Press.

Heikkilä, Melissa. 2023. "How Judges, Not Politicians, Could Dictate America's AI Rules." *MIT Technology Review*, July 17. https://www.technologyreview.com/2023/07/17/1076416/judges-lawsuits-dictate-ai-rules/.

Hirschmann, Nancy. 2016. "The Sexual Division of Labor and the Split Paycheck." *Hypatia* 31(3): 651–67.

Hodgson, Geoffrey M. 2015. *Conceptualizing Capitalism: Institutions, Evolution, Future*. Chicago: University of Chicago Press.

Hwang, Tim. 2020. *Subprime Attention Crisis: Advertising and the Time Bomb at the Heart of the Internet*. New York: Farrar, Straus and Giroux.

Jaeggi, Rahel. 2014. *Alienation*. Translated by F. Neuhouser and A.E. Smith. New York: Columbia University Press.

Jensen, Michael C., and William H. Meckling. 1976. "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure." *Journal of Financial Economics* 3: 305–60.

Kalimtzis, Kostas. 2017. *An Inquiry Into the Philosophical Concept of Scholê: Leisure as a Political End*. New York: Bloomsbury Academic.

Kang, Cecilia. 2023. "In U.S., Regulating A.I. Is in Its 'Early Days'." *New York Times*, July 21. https://www.nytimes.com/2023/07/21/technology/ai-united-states-regulation.html.

Kearns, Michael, and Aaron Roth. 2019. *The Ethical Algorithm*. New York: Oxford University Press.

Khan, Lina. 2017. "Amazon's Antitrust Paradox." *Yale Law Journal* 126: 710–805.

Khan, Lina. 2019. "The Separation of Platforms and Commerce." *Columbia Law Review* 119 (4): 973–1098.

Kim, Tae Wan. 2018. "Gamification of Labor and the Charge of Exploitation." *Journal of Business Ethics* 152 (1): 27–39.

Kim, Tae Wan, Jooho Lee, Joseph Xu, and Bryan Routledge. 2021. "Data and Manure: Are Data Subjects Investors?" *Berkeley Business Law Journal* 18: 65–89.

Kim, Tae Wan, and Bryan Routledge. 2020. "Why a Right to an Explanation of Algorithmic Decision-Making Should Exist: A Trust-Based Approach." *Business Ethics Quarterly* 32 (1): 75–102.

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

Kitchens, Brent, Steven L. Johnson, and Peter Gray. 2020. "Understanding Echo Chambers and Filter Bubbles: The Impact of Social Media on Diversification and Partisan Shifts in News Consumption." *MIS Quarterly* 44 (4): 1619–49.

Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: SAGE.

Kitchin, Rob. 2021. *Data Lives: How Data Are Made and Shape Our World*. Bristol: Bristol University Press.

Kosinski, Michal, David Stillwell, and Thore Graepel. 2013. "Private Traits and Attributes Are Predictable from Digital Records of Human Behavior." *Proceedings of the National Academy of Sciences* 110 (15): 5802–805.

Krishnamurthy, Balachander, and Craig E. Wills. 2009. "On the Leakage of Personally Identifiable Information via Online Social Networks." In *Proceedings of the 2nd ACM workshop on Online Social Networks (WOSN'09)*. New York: Association for Computing Machinery.

Kruppa, Miles. 2023. "How Frank Sinatra and Yo Gotti Are Influencing the Future of Music on YouTube." *Wall Street Journal*, August 21. https://www.wsj.com/tech/how-frank-sinatra-and-yo-gotti-are-influencing-the-future-of-music-on-youtube-971db915.

Kulynych, Bogdan, Rebekah Overdorf, Carmela Troncoso, and Seda Gürses. 2020. "POTs: Protective Optimization Technologies." *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (FAT '20), 177–88. New York: Association for Computing Machinery.

Lanier, Jaron. 2013. *Who Owns the Future?* New York: Simon & Schuster.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny. 2000. "Investor Protection and Corporate Governance." *Journal of Financial Economics* 58: 3–27.

Leonelli, Sabina. 2015. "What Counts as Scientific Data? A Relational Framework." *Philosophy of Science* 82: 810–21.

Leonelli, Sabina. 2016. *Data-centric Biology: A Philosophical Study*. Chicago: University of Chicago Press.

Li, Hanlin, Bodhi Alarcon, Sara Milkes Espinosa, and Brent Hecht. 2018. "Out of Site: Empowering a New Approach to Online Boycotts." *Proceedings of the ACM on Human-Computer Interaction* 2: 1–28. New York: Association for Computing Machinery.

Li, Hanlin, Nicholas Vincent, Janice Tsai, Jofish Kaye, and Brent Hecht. 2019. "How Do People Change Their Technology Use in Protest?" *Proceedings of the ACM on Human-Computer Interaction* 3: 1–22. New York: Association for Computing Machinery.

Li, Hanlin, Nicholas Vincent, Stevie Chancellor, and Brent Hecht. 2023. "Dimensions of Data Labor: A Road Map for Researchers, Activists, and Policymakers to Empower Data Producers." In *FAccT'23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 1151–61. New York: Association for Computing Machinery.

Locke, John. (1698) 1988. *Two Treatises of Government*. Student ed. Edited by Peter Laslett. Cambridge: Cambridge University Press.

Lopez, German. 2017. "Why People Are Deleting Uber from Their Phones after Trump's Executive Order." *Vox*, January 29. https://www.vox.com/policy-and-politics/2017/1/29/14431246/uber-trump-muslim-ban.

Magnan, Michel, and Dominic Martin. 2019. "Executive Compensation and Employee Remuneration: The Flexible Principles of Justice in Pay." *Journal of Business Ethics* 160: 89–105.

Mankiw, N. Gregory. 2015. *Principles of Microeconomics*. 7th ed. Stamford, CT: Cengage Learning.

Martin, Kirsten. 2022. "Manipulation, Choice, and Privacy." *North Carolina Journal of Law & Technology* 23 (3): 452–524.

Martin, Kirsten. 2023. "Who Counts in Business Ethics." *Business Ethics Quarterly* 33 (1): 216–43.

Martin, Kirsten, Guo Hong, and Robert Easley. 2023. "When Platforms Act Opportunistically: The Ethics of Platform Governance." Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4202821.

Martin, Kirsten, and Bidhan Parmar. 2024. "AI and the Creation of Knowledge Gaps: The Ethics of AI Transparency." Working Paper. https://papers.ssrn.com/abstract=4207128.

Marx, Karl. (1867) 1976. *Capital*, Vol. 1. Translated by Ben Fowkes. London: Penguin.

Mayer-Schönberger, Viktor, and Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston, MA: Houghton Mifflin Harcourt.

McDonald, Sean Martin. 2019. "Reclaiming Data Trusts." *Centre for International Governance Innovation*, March 5. https://www.cigionline.org/articles/reclaiming-data-trusts/.

McDermott, Patrice. 2010. "Building Open Government." *Government Information Quarterly* 27 (4): 401–13.

Morgner, Philipp, Christian Müller, Matthias Ring, Björn Eskofier, Christian Riess, Frederik Armknecht, and Zinaida Benenson. 2017. "Privacy Implications of Room Climate Data." In *Computer Security – ESORICS 2017*, part 2, 324–43, edited by Simon N. Foley, Dieter Gollmann, and Einar Snekkenes. Cham, Switzerland: Springer.

Moriarty, Jeff. 2005. "Do CEOs Get Paid Too Much?" *Business Ethics Quarterly* 15 (2): 257–81.

Narayanan, Arvind, and Vitaly Shmatikov. 2008. "Robust De-anonymization of Large Sparse Datasets." In 2008 *IEEE Symposium on Security and Privacy*, 111–25. Los Alamitos, CA: IEEE Computer Society.

Nicholas, Gabriel. 2023. "Sunsetting 'Shadowbanning.'" Information Society Project WIII essay series, Yale Law School, New Haven, CT. https://law.yale.edu/sites/default/files/area/center/isp/documents/sunsettingshadowbanning_ispessayseries_2023.pdf.

Noonan, Jeff. 2020. "Luddites, Labor, and Meaningful Lives: Would a World without Work Really Be Best?" *Journal of Social Philosophy* 51 (3): 441–56.

O'Reilly, Jacqueline. 2008. "Can a Basic Income Lead to a More Gender Equal Society?" *Basic Income Studies* 3 (3): 9.

Oremus, Will. 2018. "Are You Really the Product?" *Slate*, April 27. https://slate.com/technology/2018/04/are-you-really-facebooks-product-the-history-of-a-dangerous-idea.html.

Padova, Yann. 2021. "Data Ownership and Data Sharing: And What about Privacy?" *Lex Electronica* 21: 38–73.

Park, James J. 2017. "Reassessing the Distinction between Corporate and Securities Law." *UCLA Law Review* 64: 116–82.

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

Phillips, Robert. 1997. "Stakeholder Theory and a Principle of Fairness." *Business Ethics Quarterly* 7 (1): 51–66.

Phillips, Robert, R. Edward Freeman, and Andrew C. Wicks. 2003. "What Stakeholder Theory Is Not." *Business Ethics Quarterly* 13 (4): 479–502.

Pistor, Katharina. 2019. *The Code of Capital: How the Law Creates Wealth and Inequality*. Princeton, NJ: Princeton University Press.

Pitkin, Hannah Fenichel. 1998. *The Attack of the Blob: Hannah Arendt's Concept of the Social*. Chicago: University of Chicago Press.

Polanyi, Karl. (1944) 2001. *The Great Transformation*. 2nd ed. Boston: Beacon Press.

Posner, Eric A., and E. Glen Weyl. 2018. *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton, NJ: Princeton University Press.

Richardson, G.B. 1972. "The Organisation of Industry." *The Economic Journal* 82: 883–96.

Robinson, Eric P., and Yicheng Zhu. 2020. "Beyond 'I Agree': Users' Understanding of Web Site Terms of Service." *Social Media + Society* 6 (1): 1–13.

Ruhaak, Anouk. 2021. "How Data Trusts Can Protect Privacy." *MIT Technology Review*, February 24. https://www.technologyreview.com/2021/02/24/1017801/data-trust-cybersecurity-big-tech-privacy/.

Samuelson, Pamela. 2023. "Generative AI Meets Copyright." *Science* 381: 158–61.

Sandberg, Joakim, and Alexander Andersson. 2022. "CEO Pay and the Argument from Peer Comparison." *Journal of Business Ethics* 175: 759–71.

Sarathy, Ravi, and Christopher J. Robertson. 2003. "Strategic and Ethical Considerations in Managing Digital Privacy." *Journal of Business Ethics* 46, 111–26.

Savoleinen, Laura. 2022. "The Shadow Banning Controversy: Perceived Governance and Algorithmic Folklore." *Media, Culture & Society* 44 (6): 1091–109.

Scharding, Tobey K. 2021. "Recognize Everyone's Interests: An Algorithm for Ethical Decision-Making about Trade-Off Problems." *Business Ethics Quarterly* 31 (3): 450–73.

Schwartz, Adina. 1982. "Meaningful Work," *Ethics* 92 (4): 634–46.

Schultz, Mario D., and Peter Seele. 2023. "Towards AI Ethics' Institutionalization: Knowledge Bridges from Business Ethics to Advance Organizational AI Ethics." *AI and Ethics* 3: 99–111.

Seele, Peter, Claus Dierksmeier, Reto Hofstetter, and Mario D. Schultz. 2021. "Mapping the Ethicality of Algorithmic Pricing: A Review of Dynamic and Personalized Pricing." *Journal of Business Ethics* 170: 697–719.

Seele, Peter, and Mario D. Schultz. 2022. "From Greenwashing to Machinewashing; A Model and Future Directions Derived from Reasoning by Analogy." *Journal of Business Ethics* 178: 1063–89.

Shear, Michael D., Cecilia Kang, and David E. Sanger. 2023. "Pressured by Biden, A.I. Companies Agree to Guardrails on New Tools." *New York Times*, July 21. https://www.nytimes.com/2023/07/21/us/politics/ai-regulation-biden.html.

Sheth, Falguni. 2002. *Labor, Work, and Citizenship: A Study in the Meaning and Implications of the Concept of Work in Hegel, Marx, Arendt, and Kittay*. PhD diss., New School University, New York.

Silver, Gerald. A. and Myrna L. Silver. 1989. *Systems Analysis and Design*. Reading, MA: Addison Wesley.

Smith, Jeffrey. 2019. "Navigating Our Way between Market and State." *Business Ethics Quarterly* 29 (1): 127–41.

Soffia, Magdalena, Alex J. Wood, and Brendan Burchell. 2022. "Alienation Is Not 'Bullshit': An Empirical Critique of Graeber's Theory of BS Jobs." *Work, Employment and Society* 36 (5): 816–40.

Solove, Daniel J. 2012. "Privacy Self-Management and the Consent Dilemma." *Harvard Law Review* 126: 1880–903.

Standing, Guy. 2010. *Basic Income: A Guide for the Open-Minded*. New Haven, CT: Yale University Press.

Steinberg, Etye. 2020. "Big Data and Personalized Pricing." *Business Ethics Quarterly* 30 (1): 97–117.

Steinberg, Etye. 2022. "Run for Your Life: The Ethics of Behavioral Tracking in Insurance." *Journal of Business Ethics* 179 (3): 665–82.

Sterling, Toby. 2018. "Facebook Users Unite! 'Data Labour Union' Launches in Netherlands." *Reuters*, May 23. https://www.reuters.com/article/us-netherlands-tech-data-labour-union/facebook-users-unite-data-labour-union-launches-in-netherlands-idUSKCN1IO2M3/.

Sunstein, Cass R. 2017. *#Republic: Divided Democracy in the Age of Social Media*. Princeton, NJ: Princeton University Press.

Suppes, Patrick. 1962. "Models of Data." In *Logic, Methodology, and Philosophy of Science: Proceedings*, edited by Ernest Nagel, Patrick Suppes, and Alfred Tarski, 252–61. Stanford, CA: Stanford University Press.

Svendsen, Lars Fredrik. 2014. *Work*. London: Routledge.

The Economist. 2017. "The World's Most Valuable Resource Is No Longer Oil, but Data." *The Economist*, 5 May. https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.

Thorp, Jer. 2021. *Living in Data*. New York: Farrar, Straus and Giroux.

Twenge, Jean M., Gabrielle N. Martin, and Keith W. Campbell. 2018. "Decreases in Psychological Well-Being among American Adolescents after 2012 and Links to Screen Time during the Rise of Smartphone Technology." *Emotion* 18 (6): 765–80.

Utz, Christine, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz. 2019."(Un)informed Consent: Studying GDPR Consent Notices in the Field." In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security* (CCS'19), 973–90. New York: Association for Computing Machinery.

van Dijck, José. 2009. "Users Like You? Theorizing Agency in User-Generated Content." *Media, Culture & Society* 31 (1): 41–58.

Veblen, Thorstein. (1899) 2012. *The Theory of the Leisure Class*. Cleveland: Duke Classics.

Veltman, Andrea. 2016. *Meaningful Work*. Oxford: Oxford University Press.

Venturini, Jamila, Luiza Louzada, Marilia Maciel, Nicolo Zingales, Konstantinos Stylianou, and Luca Belli. 2016. *Terms of Service and Human Rights: An Analysis of Online Platform Contracts*. Rio de Janeiro: Editora Revan.

Vincent, Nicholas, Brent Hecht, and Shilad Sen. 2019. "'Data Strikes': Evaluating the Effectiveness of a New Form of Collective Action against Technology Companies." In *The World Wide Web Conference* (WWW '19), 1931–43. New York: Association for Computing Machinery.

Vincent, Nicholas, Hanlin Li, Nicole Tilly, Stevie Chancellor, and Brent Hecht. 2021. "Data Leverage: A Framework for Empowering the Public in Its Relationship with Technology Companies." *Proceedings of the 2021 Conference on Fairness, Accountability, and Transparency* (FAccT '21), 215–27. New York: Association for Computing Machinery.

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press

Weber, Max. 1976. *The Protestant Ethic and the Spirit of Capitalism*. Translated by Talcott Parsons. New York: Scribner's.

Weinberger, David. 2011. *Too Big to Know*. New York: Basic Books.

West, Sarah Myers. 2019. "Data Capitalism: Redefining the Logics of Surveillance and Privacy." *Business and Society* 58: 20–41.

Whelan, Glen, Jeremy Moon, and Bettina Grant. 2013. "Corporations and Citizenship Arenas in the Age of Social Media." *Journal of Business Ethics* 118 (4): 777–90.

White, Stuart. 2003. *The Civic Minimum: On the Rights and Obligations of Economic Citizenship*. Oxford: Oxford University Press.

Williamson, Oliver. 1973. "Markets and Hierarchies: Some Elementary Considerations." *American Economic Review* 63: 316–25.

Williamson, Oliver. 2010. "Transaction Cost Economics: The Natural Progression." *American Economic Review* 100: 673–90.

Wrzesniewski, Amy, Clark McCauley, Paul Rozin, and Barry Schwartz. 1997. "Jobs, Careers, and Callings: People's Relations to Their Work." *Journal of Research in Personality* 31(1): 21–33.

Wu, Tim. 2016. *The Attention Merchants*. New York: Alfred A. Knopf.

Zins, Chaim. 2007. "Conceptual Approaches for Defining Data, Information, and Knowledge." *Journal for the American Society of Information Science and Technology* 58: 479–93.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capital*. New York: Public Affairs.

Zuiderwijk, Anneke, and Marijn Janssen. 2014. "Open Data Policies, Their Implementation and Impact: A Framework for Comparison." *Government Information Quarterly* 31 (1): 17–29.

. . .

JULIAN DAVID JONKER (jonker@wharton.upenn.edu) is an assistant professor in legal studies and business ethics at the Wharton School of the University of Pennsylvania. He obtained his Bachelor of Business Science, Bachelor of Laws, and Master of Philosophy at the University of Cape Town, and his PhD in Philosophy at the University of California, Berkeley.

https://doi.org/10.1017/beq.2024.25 Published online by Cambridge University Press