

Ethical Aspects of Multi-stakeholder Recommendation Systems

Silvia Milano¹, Mariarosaria Taddeo^{1,2}, Luciano Floridi^{1,2*}

¹ Oxford Internet Institute, University of Oxford, 1 St Giles, Oxford, OX1 3JS, UK

²The Alan Turing Institute, British Library, 96 Euston Rd, London NW1 2DB, UK

* Email of correspondence luciano.floridi@oii.ox.ac.uk

Abstract

This article analyses the ethical aspects of multistakeholder recommendation systems (RSs). Following the most common approach in the literature, we assume a consequentialist framework to introduce the main concepts of multistakeholder recommendation. We then consider three research questions: who are the stakeholders in a RS? How are their interests taken into account when formulating a recommendation? And, what is the scientific paradigm underlying RSs? Our main finding is that multistakeholder RSs (MRSs) are designed and theorised, methodologically, according to neoclassical welfare economics. We consider and reply to some methodological objections to MRSs on this basis, concluding that the multistakeholder approach offers the resources to understand the normative social dimension of RSs.

Keywords

Artificial Intelligence, Digital Ethics, Multistakeholder Recommendation, Recommender Systems, Recommender Systems Ontology, Recommender Systems and Welfare, Social aspects of recommendation.

1. Introduction

Recommender Systems (RSs, also known as recommendation algorithms) play an essential role in the current online environment. They inform how web search results are generated and displayed, curate news feeds, allocate advertisement slots, and power job and dating platforms, among other things. RSs raise important issues with respect to basic ethical categories such as personal identity, fairness, accountability, and privacy (Milano, Taddeo, & Floridi, 2019). Moreover, they can shape the construction of preferences, as well as orchestrating social interactions, by virtue of controlling the presentation of options and the information exchanges that take place within the system

(Taddeo & Floridi, 2018). Therefore, studying how they function can provide conceptual tools and foundations to uncover and potentially modify social norms and to approach issues relating to value alignment and decision making in the presence of moral uncertainty.

A traditional way to conceptualise the task of a RS is that of “finding good items”, interpreted as those “things” that are most relevant or that match most accurately the preferences of the user to whom the recommendation is targeted (Jannach & Adomavicius, 2016). While this approach has been useful in the past—informing much of the computer science research in the area since the beginning—its limitations are starting to become evident. Conceptualising recommendation as essentially a problem of predicting user preferences fails to consider adequately the origin of these preferences, their malleability, and their contextual dependence. Furthermore, even when user preferences are relatively stable and well-defined, promoting the accuracy of recommendation may be at odds with ethical considerations concerning, for example, fairness and privacy (Milano et al., 2019). Recent work on accuracy-preserving fair recommendation confirms this, by taking on the challenge of limiting the loss in accuracy while striving to provide fair recommendations (Edizel, Bonchi, Hajian, Panisson, & Tassa, 2019).

In contrast to this traditional user-centred approach, which is too impoverished to account adequately for the social impacts of recommendation, a new research paradigm is emerging that explicitly models the multiple dimensions of recommender systems. In particular, (Abdollahpouri, Burke, & Mobasher, 2017) define recommender systems as *multistakeholder environments*, in which different categories of agents (users, providers, and system developers) come together. These stakeholders typically pursue contrasting interests, and the role of a well-functioning RS is to structure the interactions efficiently. Seen in this light, RSs can be understood as giving rise to institutional environments where the primary goods that are exchanged are items of information. Indeed, the environments in which RSs operate often constitute *multisided markets* (Rochet & Tirole, 2003), that is, platforms serving multiple sets of participants. The research on the economics of multisided markets is a new and exciting area, which will no doubt contribute to expanding our understanding of the technical aspects of RSs. In the rest of this article, our purpose is to contribute to the conceptual foundations of this new field by examining the ethical aspects of multistakeholder RSs. In line with the current literature (although not because we wish to endorse it as necessarily the correct or only available approach) we shall adopt a common consequentialist approach. We examine three questions:

- a) who are the stakeholders in a RS?
- b) how are their interests taken into account when formulating a recommendation? And
- c) what is the scientific paradigm underlying RSs?

The rest of the article is organised as follows. Section 2 introduces the basic elements of multisided platforms. Section 3 outlines the ethical features of RSs, drawing on the analysis previously developed in (Milano et al., 2019) and considering the advantages and disadvantages of multistakeholder evaluations of RSs in a consequentialist setting. Section 4 addresses the first of our research questions, examining the ontology of recommender systems and identifying five stakeholder categories. Section 5 turns to the second question, examining how multistakeholder RSs approach welfare evaluations. In answer to our third research question, section 6 examines how multistakeholder RSs fall methodologically within the paradigm of neoclassical welfare economics, considering some methodological objections to this approach. Section 7 concludes the article.

2. Multisided platforms

Everyday examples of multisided markets include those constituted by credit card providers, platforms like AirBnB or Uber, music or video streaming services, e-commerce platforms such as Amazon, Taobao or eBay, or social networks like Facebook and LinkedIn. A common characteristic of multisided markets is the presence of network effects, which can positively or negatively impact each side of the market. For example, the presence of many active users on Spotify makes it more appealing for artists, including already famous ones, to have their catalogue available there (a positive cross-side network effect). However, the presence of many famous artists on Spotify, representing a wide range of music genres, makes it more difficult for new entrants to find an audience, because many users will not have heard about their music (a negative internal network effect on the providers' side). Similarly, on an e-commerce platform like Amazon, customers benefit from the presence of many sellers, and sellers also benefit from the presence of customers, though they can face intense competition.

The sides of the market are not the only stakeholders in the transactions that take place between participants in the market. These multisided platforms can produce externalities, which may remain as such for (that is, may not be internalised by) either side. For example, the rise of Amazon can negatively impact the business of other more traditional retailers, who do not participate in the e-commerce platform but whose business model is eroded. In the case of social networks, the success of a platform can have the effect of creating self-reinforcing circles where information is shared and amplified, leading to social externalities such as widespread misinformation or political polarization (Ma, n.d.).

Within this well-known setting, the task of a RS is to match users with *items*. We use the term in a general and inclusive sense: items could be products, services, content, answers to queries or anything else that can be exchanged on a platform. Since the amount of possible transactions within multisided platforms is typically enormous, the services of a RS are essential to enable users to discover items that might be relevant. Without a RS, users may face information overload, being confronted with too many options (Schwartz, 2005). So, RSs are expected to help users to make better choices, by reducing the set of options to a size that users are able to evaluate, and by highlighting new items that the users might find useful. In this way, RSs are a vital component to the functioning of many multisided markets, enabling transactions that would not take place in their absence, due to the informational limitations of each side. This gives RSs a great power to influence the market, and with this, as the saying goes, comes a great responsibility.

3. Ethical features of RS

As we noted above, RSs have impacts on the interests of multiple parties, in addition to individual users. The nature and identity of the stakeholders to the recommendation will vary depending on the context in which the recommendation is made, but in general they will include participants in a multisided platform belonging to several distinct categories (Burke & Abdollahpouri, 2017).

Considering the diversity of the stakeholders to a recommendation, there are several ways in which one can take their interests into account, when designing a RS. In (Milano et al., 2019) we argued that the main, morally relevant features of recommendation can be ordered along a two-dimensional matrix, which takes into account the quality of the actions (which can impact on stakeholder's rights) and the consequences (which can affect stakeholders' utilities) of a RS, and whether the prospected actions or consequence cause immediate or merely risk of harm. On the basis of that taxonomy, it is easy to classify different ethical challenges posed by recommender systems in terms of whether they affect the rights or the utilities of different stakeholders, and whether they cause an immediate harm as opposed to exposing a party to a risk of being harmed (Table 1). For example, providing inaccurate or irrelevant recommendations directly harms a user by reducing the utility that they derive from the recommended options, while low attention to security that could lead to leaking of sensitive information exposes users of a recommender system to violations of their rights to privacy.

Table 1 - Taxonomy of the morally relevant features of recommendation.

	Immediate Harm	Exposure to Risk
Utility	e.g. inaccurate recommendations	e.g. A/B testing
Rights	e.g. unfair treatment	e.g. leaking of sensitive information

When we consider a recommender system from the standpoint of multiple stakeholders, the dimensions of the ethical taxonomy apply to each one of them. So, when we consider the impact that a recommender can have on society as a whole, we may take into account how the recommendations it produces have an impact on the level of social utility that it achieves (for example, gains in efficiency), or the social rights that it promotes or harms. Then, we have to balance these considerations with the corresponding ones for the other relevant stakeholders.

As we noted in the introduction, for the purposes of this article we are going to look at the ethical implications of MRS from a consequentialist perspective, that is, we will focus exclusively on the first row of Table 1. This should not be read as an endorsement of consequentialism on our part. Rather, the reason for this choice is merely pragmatic: consequentialist evaluations are the ones that are more frequently incorporated in the design of artificial systems, arguably because they are easier to implement.

The metrics that are used most frequently to evaluate the performance of RSs, such as mean average precision or recall (Herlocker, Konstan, Terveen, & Riedl, 2004; Shani & Gunawardana, 2011), are grounded in a consequentialist framework, as is the traditional set up of the recommendation problem (“find the best items”), which focuses on identifying the items that are most relevant to the user who receives the recommendation; an approach that is conceptually close to revealed preference. It ties to the tradition of welfare economics, so we can use the conceptual tools of welfare economics to address questions about the ethical significance of recommendations.

The multistakeholder approach to RSs has several advantages over the more traditional user-centred approach. Epistemologically, it gives a more accurate representation of the target phenomenon, enabling one to represent features that would be masked in the user-centred approach, like the dependence between recommendation strategies and the amount and quality of information that the RS is able to collect. Thus, the multistakeholder approach enables one to draw more inferences and better predictions of how the system is likely to evolve. Normatively, the multistakeholder approach enables one explicitly to account for the interests of all relevant parties. This is important both in theory and in practice. It makes it possible to reason about the trade-

offs that RSs make between the interests of different stakeholders without introducing *ad hoc* biases, which is what normally would have happened in practice, as RSs are developed to serve particular applications. For example, it is well-known and appreciated that e-commerce websites often introduce a bias in favour of less popular products, increasing the chances that they are recommended *vis à vis* more popular items (which, according to the RS's own model, are more relevant to the user to whom the recommendation is served). This bias serves a specific interest, reducing what is known as the “long tail” of items that are seldom recommended and sold on the catalogue, a common problem for online retailers who carry large catalogues. It could be argued that, while on the user-centred approach this represents a bias, reducing the long tail may be a legitimate interest of the platform, and something that should feature more explicitly into the formulation of the RS's objective. The multistakeholder approach to RSs makes it possible to include this as a feature of the system.

In the following sections, we turn to examine the features of multistakeholder recommendation with respect to ontological, normative, and methodological issues. We examine and respond to the objections that arise in these areas, concluding that the multistakeholder approach provides an adequate framing for the general problem of generating recommendations.

4. Ontology

In this section, we focus on the specification of *who* are the relevant stakeholders and *how* the stakeholders are represented within the RS. In this context, we consider three possible objections to the multistakeholder approach to RSs that could be seen to arise from issues to do with the specification of their ontology, which include:

- 1) the problem of *fragmentation* of the stakeholders;
- 2) the choice of level of abstraction (LoA); and
- 3) the problem of delimiting the set of relevant stakeholders.

As we shall see, each of these issues needs to be considered carefully, but they are features of any system that is designed to formulate recommendations (so, even of user-centred RSs). Therefore, none of them constitute valid objections to the multistakeholder approach.

Recommender systems are designed to formulate recommendations for the users accessing the platform. In order to perform this task, they are equipped with a representation of the problem and access to data regarding users and items. This representation implies an ontology. Expanding on the analysis given by (Burke & Abdollahpouri, 2017), we identify four categories of stakeholders in a recommendation.

- i. Users.* These are the parties to whom the recommendation is targeted. For example, in a commercial application this could be a shopper accessing an e-commerce website, or a user accessing a streaming service (note that these do not always correspond to physical individuals; they can be groups, or users sharing a single account, like a family).
- ii. Providers.* These are the parties who make the options available. For example, they could be the hotels in a commercial platform specialising in hotel recommendations, or the artists who publish their work on a music recommendation system. Providers are affected by the recommendations that the system makes to users, in so far as their “items” can receive more or less attention depending on how they are recommended.
- iii. System.* This captures the interests of the platform on which the recommendations are generated. An obvious component may be for the platform to remain viable, or to be able to satisfy user requests. In some cases, this may be at odds with providing the most accurate recommendations. For example, if a user would likely abandon the platform after meeting their perfect match, this may decrease the value of perfect match recommendations to the system.
- iv. Society.* The recommendations made by a system can have systemic effects on society, for example by altering or reinforcing existing social norms. They can also give rise to externalities. For example, a wildlife reserve could be indirectly affected by the popularity of a route that is recommended often, or herd immunity across a population could be lowered due to the spread of misinformation regarding the safety of vaccines on a social platform (facilitated by a RS).

These four categories of stakeholders are present in any recommendation, although their contours may differ depending on the application. In some instances, several of the categories may overlap. For example, on a dating platform users and providers will normally coincide. One can also imagine a media RS on a platform which produces all its content, where system and providers coincide. Society usually encompasses the members of the other categories of stakeholders.

Having identified the main categories of stakeholders in a recommendation, let us now turn to considering how RSs represent them. The ontology of a RS typically includes a set of options (the items that can be recommended), a set of users (the receivers of the recommendation), and, less frequently—depending on the level of abstraction¹ (LoA, Floridi, 2008) that is chosen—

¹ A level of abstraction can be imagined as an interface that enables one to observe or describe some aspects of a system analysed, while making other aspects opaque or indeed invisible. For example, one may analyse a house at the LoA of a buyer, of an architect, of a city planner, of a plumber, and so on. LoAs are common in computer science, where systems are described at different LoAs (computational, hardware, user-centred etc.). Note that LoAs can be combined in more complex sets, and can be, but are not necessarily hierarchical, with higher or lower “resolution” or granularity of information.

a representation of other stakeholders. However, as with any model, RSs represent some target real-world phenomena, but do not coincide or necessarily share all the relevant features of their target. This is a normal feature of models, but in the specific case of multistakeholder RSs, it gives rise to three potential issues, which we now examine in turn.

A first issue that we note is that the ontology implicit in the system's design may not correspond perfectly to its target. This is evident in how the system models users. In the system's perspective, each user entry is a separate individual, however this may often not be the case in the real world. Several users may share a single account (think, for example, members of a family sharing a Netflix account), or a single user may have more than one account, using each in different circumstances. In both cases, a single user entry in the recommender system's perspective does not map to a physical individual. This sort of *fragmentation and mismatch* between the RS's implied ontology and the target system causes issues for the ethical evaluation of RSs. Without a reliable way to identify the actual stakeholders in a recommendation and match them with what is represented by the system, it is difficult to give a reliable evaluation of the impacts of a RS.

A second issue that needs to be considered when designing and evaluating a recommender system is what the right LoA should be. For example, it is natural to think that the right LoA for a recommender system that is designed to make film recommendations is targeted at individual users. However, if the system is being used by a family or a group of friends, the right LoA for the user category will not be individuals, but groups. Shifting the LoA for the user category from individuals to groups generates new research questions (Lev & Tennenholtz, 2017). The choice of LoA is also relevant to the evaluation of a recommender system, in ways that are not always obvious if one focuses on the LoA that seems most natural from the design perspective. For example, an e-commerce recommender system that is designed to recommend consumers which goods to purchase may naturally be built around the task of predicting the consumers' preferences. Individual consumers might prefer cheaper products to more expensive ones, and not be very interested in the amount of packaging or whether the products need to be shipped from distant locations. However, these variables can have negative environmental effects, so if one shifted the LoA to (include also) the social preferences for environmental outcomes, then the predictions and recommendations of the system, in the aggregate, may not correspond to the social preferences. This general pattern is evident in public-good problems, where the provision of a public good (which is in the collective interest) conflicts with individual rationality assumptions. Note that another instance of this problem in recommender systems is the tension between optimal exploration and exploitation (Immorlica, Mao, Slivkins, & Wu, 2019). In this case, when one looks at the problem from the point of view of individual users, the strategy that maximises users' utility

may be to exploit the information that the system already holds, to recommend an item that has the highest expected probability of satisfying the user's preferences. However, on an aggregate level, it may be that exploration is the strategy that has the greatest expected utility, giving rise to positive network effects on the multisided platform in which the RS operates.

The issues of fragmentation and of choice of LoA that we have just described are not specific to multistakeholder RSs. These issues apply equally to the user-centred approach: fragmentation is as much of a problem for a RS that is designed to respond to individual users, since its presence corrupts the integrity of the data that the RS holds about users. Similarly, the choice of LoA is crucial to user-centred RSs, as the users to whom they are designed to cater may correspond to identified groups of individuals, as in the case of group recommendations. The third and last issue that we consider, however, is more specific to multistakeholder recommendation and concerns the problem of *delimiting the set of relevant stakeholders*. Identifying all the relevant stakeholders in a recommendation is often unlikely, if not utterly unfeasible. In the most straightforward of cases, there may be consequences of a (pattern of) recommendation that affect parties in ways that are difficult to anticipate: for example, recommending a certain song may affect the popularity of a musical genre, influencing what is in fashion, and indirectly influencing broader cultural trends. This problem is well known in normative ethics, where consequentialist theories—particularly act utilitarianism—have trouble in giving concrete guidance in practical situations, because it is always difficult, if not impossible, to foresee and balance the consequences of all the possible courses of action. In this sense, the problem is probably not solvable in its generality. Following an approach that is common in applied ethics, as well as in economics, we should be clear about how we are going to identify practically the stakeholders in the recommendations, and the way in which their interests are measured (which we consider in the next section). This modelling stage should involve a justification of the chosen LoA, setting clear boundaries to the recommendation problem that we are trying to address. This is necessary in order to make possible a formal representation of the problem, but at the same time it should be clear that the LoA will leave out some information, and remains open to further scrutiny.

5. RSs and Welfare

Once the relevant stakeholders have been identified, the next step is to consider their interests, both in isolation and in relation to each other. This raises multiple related issues. First, the ontology of the system may not correspond to anything definite in the real world, making it difficult to evaluate recommendations in terms of its utility (which is usually conceived as tied to a person). Second, utility comparisons and trade-offs are difficult to define due to the problems of

overlapping and incomparability. And finally, if recommender systems can influence and shape the preferences of their users, this introduces possible feedback effects in their evaluation.

In the simple user-centred approach, aggregating the interests of multiple parties is not a central issue because the RS is designed to serve individual users. Some issues still arise from the fact that RSs may be designed to serve recommendations to groups of users, such as family holiday planning, or home video recommendations. In other cases, RSs may have to make trade-offs between exploration and exploitation strategies. For example, a film recommendation system may have the choice to recommend item A or item B to a user, where A has been seen and highly rated by several other users, while B is a new item on the catalogue. Recommending A might be a safer choice, but recommending B would lead the RS to acquire more valuable information, which could contribute to better recommendations down the line. However, the user-centred approach is too limited, as we have seen, because it obscures some relevant features of recommendations. The multistakeholder approach addresses this problem, but now the issue of aggregating the welfare of multiple stakeholders becomes central.

This task is complicated by the fact that (due to the features of the RS ontology discussed in the previous section) stakeholders could be grouped in different ways, which could lead to potentially different evaluations. Moreover, based on the categories of stakeholders that we outlined in the previous section, an individual person could belong to more than one category. For example, in two-sided recommendations, such as dating platforms, users and providers coincide. In most cases, users, providers, and other actors are also at the same time members of society, and as such may share a collective interest in achieving specific aggregate outcomes. Classic examples where this gives rise to complex issues are public-goods problems, recalled in the previous section. For instance, e-commerce recommendation of a single-use plastic product may maximise the user's utility, but produce a negative outcome in the aggregate. The RS thus faces a trade-off between providing an accurate recommendation, or a paternalistic one that serves the collective interests. Reliance on simple accuracy metrics for evaluating recommendations to individual users would completely miss this issue. Attention to the choice of LoA could be used to address this issue, by ensuring that the stakeholders at each LoA are constituted by disjoint sets of individuals. However, this approach may still be too restrictive, as it would generally exclude the possibility to model the trade-offs between individual and collective interests (as the category of *society* typically subsumes other stakeholder categories).

A second problem faced by multistakeholder RSs is that of grounding comparisons of utility. While user-centred RSs only deal with one category of stakeholders, whose utility is measured using a homogeneous metric, multistakeholder RSs have to compare different

stakeholders whose utility is not measured uniformly within the system. This is an added complication for the design of multistakeholder RSs, but it cannot be eliminated. User-centred approaches avoid this difficulty only by displacing the problem, as the interests of other stakeholders are often incorporated as biases in the RS (e.g., Amazon favouring less popular items on the catalogue) in a way that is not transparent.

6. Methodological objection to RSs

The paradigm within which MRSs are located is that of neoclassical economics, which relies on the conceptual apparatus of revealed preference and individual rationality. As Hausman writes, “[...] most economists take welfare to coincide with the satisfaction of preference” (Hausman, 2018). This is also the predominant approach in recommender systems, which are traditionally designed to (seek to) predict and then satisfy user preferences (Jannach & Adomavicius, 2016). Under this interpretation, maximising utility is tied to the satisfaction of stakeholders’ preferences. This approach raises several issues, some of which are familiar from debates in the philosophy of economics, and others that are more specific to the framework of recommender systems. A fundamental methodological issue that faces RSs is to understand whether and to what extent they can infer users’ “true utility”, when they are responsible for shaping the choice environment and could therefore manipulate preferences. Drawing on the literature on revealed preference theory in the philosophy of economics, in this section we focus on the most serious methodological objections that can be moved against RSs, pointing to the influence of framing effects, the possible incoherence of individual preferences, and the presence of feedback loops. All of these issues put into doubt whether RSs can legitimately infer, let alone satisfy, the preferences of their users in ways that are any better than trivial guesses (“if a likes x and y is similar to x then a may like y as well”).

Individual preferences are often incoherent, based on misinformed judgements, or otherwise mistaken (Thaler & Sunstein, 2008). The link between preference satisfaction and the maximization of utility is therefore questionable, because satisfying “dirty” preferences (as they are often referred to) does not lead generally to better outcomes for the individuals. For example, if a user has an apparent preference for sugary snacks over fruit, then giving them recommendations for sugary snacks satisfies their manifested preferences. However, it may not promote the user’s health, which is also something that they care about (or ought to), and so the recommendation may in fact have negative effect on the user (e.g., by promoting unhealthy behaviour, reinforcing unbalanced choices, or putting the user in the situation of having to resist a tempting option).

In response to the issues raised by dirty preferences, (Thaler & Sunstein, 2008) propose nudging as an approach to framing choice situations so that individuals are encouraged to choose the options that would in fact be best for them, that is, what they would prefer, if they had the facts right. This strategy is often criticised for being paternalistic, and for relying on the implicit assumption that there is a set of “laundered” preferences that are consistent and that accurately represent the individual’s fundamental preferences (Floridi, 2016; Infante, Lecouteux, & Sugden, 2016).

The issue of dirty preferences is relevant to recommender systems, which are open to the same criticism usually levelled at choice-framing in behavioural economics. In this context too, the choice of recommender systems to present specific items for a recommendation, in so far as it is based on the prediction of individual preferences, risks being paternalistic. Moreover, the sheer amount and complexity of the possible options makes it difficult, in many applications, to identify a plausible sense in which one could define an individual’s laundered preferences. In the discussion offered by (Hausman, 2016), for example, some promising methods for the identification of laundered preferences rely on the availability of comparisons between multiple choice framings and insights on the background knowledge, on the one hand, and cognitive ability of the individual to whom the preferences are imputed, on the other. However, these conditions are much harder to achieve for recommender systems.

To complicate matters further, the underlying preferences may be sensitive to the way they are measured. This, as we saw before, is known to give rise to framing effects on individual preferences, as detailed in the work of behavioural economists (Kahneman & Tversky, 1979). To use a trivial example, whether a child likes a vegetable may be influenced by the way it is presented. Similarly, the success of a recommendation may depend on factors that affect the preferences of the receiver of the recommendation, such as the appeal of how an option is presented, how the recommendation is explained, and the degree of trust that the user has in the recommender itself. The presence of feedback loops creates challenges in the evaluation of recommendations. It is possible that a recommender system that identifies such loops will exploit them, in order to achieve the maximum reward from the recommendations. This is what allegedly is the case, for example, in the case of YouTube recommendations (Chaslot, 2019).

Feedback loops challenge the idea that there is a set of underlying laundered preferences, which accurately correspond to the users’ interests. The fact that user preferences are susceptible to feedback loops, on the contrary, indicates that preferences are contextual, malleable, and may be influenced by external factors. This observation, together with the idea that utility corresponds to the satisfaction of preferences, makes it difficult to assess whether the stakeholders’ interests

are genuinely served (instead of being shaped) by a recommendation. On the one hand, if the system is contributing to shape the preferences, this could be used to promote preferences that are intuitively irrelevant or even bad, either for the individuals, or in terms of the social consequences to which they give rise. On the other hand, if the resulting preferences are genuinely held by the individuals, then—by the lights of the preference satisfaction interpretation of utility—satisfying them would still be in the best interest of the stakeholders.

Finally, a fundamental issue that is connected with the interpretation of utility as preference satisfaction is that, as we have observed above, the users in a RS's ontology may not correspond to individual persons. So, even setting aside the problem of dirty preferences, it is questionable whether we can genuinely treat the preferences ascribed to the users (as represented by the system) as belonging to actual individuals. Where this is not the case, we have the problem that the satisfaction of preferences will not correspond to an increase in utility for any one individual. For example, several individuals may share access to a single Netflix account, and watch films together or individually. From the point of view of the system, the interactions and recommendations are attached to a single user, when in reality they involve several individuals. The issue then is that whatever the system identifies as the user's preferences, they do not correspond to the preferences of any physical individual, and therefore satisfying them does not directly amount to maximising any person's utility.

These objections are not specific to the multistakeholder approach, but rather point to deep methodological commitments of RSs, including user-centred approaches. The appropriate responses will need to take into account the specific circumstances in which RSs operate. This will require further investigation. Here, it may be worth offering a final comment in relation to the issue of dirty preferences and framing effects. We noted that, while this raises serious concerns, the possibility that RSs could nudge users towards specific sets of options should not in principle be opposed. Given the size and structure of digital platforms, RSs are essential to the functioning of markets and cannot be dispensed with, given the unmanageable information burdens on users that would immediately ensue. As is often observed, any choice architecture has potential to create framing effects, so RSs will give rise to some of these. Since RSs are here to stay, it is important that we recognise their tendency to give rise to framing effects and make them the object of careful auditing, instead of trying to eliminate them, which would not be a feasible strategy. Similarly, the nature of preferences that the RSs identify should be the object of continuous monitoring. We should expect that RSs will have a tendency to pick up dirty preferences, as this is inevitable given that they rely on proxies for inferring user preferences. Since this is a feature that cannot be eliminated, the focus should be put on minimising its negative impact. The most effective way to

do so is to make recommendations more transparent. Explanations for the system's inferences and recommendations could be used to empower users and other stakeholders to use the RS as a tool to support their interests, instead of interacting with it in a passive way (Karakayali, Kostem, & Galip, 2018; Seaver, 2018).

7. RSs as social planners

The multistakeholder approach to RSs is a significant improvement on the previous user-centred approach, both on an epistemological and on a normative perspective. This shift in perspective raises several interesting and new research problems, which we have started to articulate in this paper. Before concluding, we would like to draw attention to some fundamental theoretical features of RSs and situate them within a broader perspective.

As we observed at the beginning of this article, RSs typically operate within multisided platforms, where they direct the flow of information between a multitude of participants on different sides of the platform. In this sense, they function as *social planners*—a term that intuitively describes the role that these systems have in facilitating transactions—which can be leveraged to influence the rewards that accrue to the different parties.

It is an open question whether and to what extent the presence of specifically designed RSs increases efficiency in multisided platforms, but it is certainly plausible that it will do so to some extent. In some cases, the effects of the presence of RSs have already been felt and documented for their transformative impact, such as in the music industry (Eriksson, Fleischer, Johansson, Snickars, & Vonderau, 2019). The example of Spotify is particularly interesting in this regard. As Eriksson et al. document, Spotify's business model is not so much based on selling music, but on generating data about its users, including the amount and genres of music that they listen to, their physical locations and browsing habits, which can be sold to advertisers. Spotify's pro-rata system for paying royalties to artists has also had deep effects on the music industry, with critics pointing out how it makes it more difficult for niche musicians to earn enough through their royalties (Luckerson, 2019). The reason for this is that, in Spotify's model, the money received from individual subscriptions is pooled in a pot, which is then divided up among all the artists in proportion to the number of streams. This means that each user's subscription is effectively divided up among all the artists, including ones that the user never listened to, not in proportion to the number of users that listened to them, but to the number of times that their tracks were streamed. Under this model, most small artists, especially those in niche music genres, are left with a very small share of the royalties. Moreover, although every user pays the same monthly subscription for the service, the pro-rata system assigns proportionally more importance to users

who total more hours of listening. This, as some critics have pointed out, is an unfair element of Spotify's system.

The multistakeholder approach that we have presented in this article offers the tools to analyse the Spotify case. The stakeholders in Spotify's RS clearly include listeners (or users), but also the artists whose tracks are on the system's catalogue, as well as Spotify itself, whose interest is not just to recommend music, but to gain valuable data about its users. Additionally, there is a social dimension present in the way in which Spotify's pro-rata system dis-incentivises the production of certain genres of music. Making changes to the algorithm would affect the interests of all these stakeholders. For instance, changing the way in which the system keeps tracks of which songs a user has listened to would make a difference to how royalties are distributed (currently, a track is marked as "listened to" after 30 seconds), as well as to the way users are categorised.

Spotify's case is just one example, and a useful reminder of the fact that RSs are becoming ubiquitous in multisided platforms, where their services are necessary to filter the huge amount of information available, while also having the power to exert a deep influence on entire industries. (Krukowski & Thompson, 2019) note how, despite the fact that more tracks are available thanks to streaming services than ever before in history, the design of RSs such as Spotify's are actually restricting the options available to conservative, in this sense making them more "conservative". This is in the interest of the system, in so far as its bottom line is not to sell music, but to collect valuable data on its subscribers. Predictable, conservative users are arguably better for targeting advertisement than listeners whose tastes and interests could be changing more frequently. This effect is likely to appear in other industries as well, including e-commerce and news recommendation. We need to be vigilant of this and apply MRSs so that they fulfil their aim to support human decision making, instead of risking turning individual users into ever more predictable (and manipulable) data sources.

8. Conclusion

In this paper, we have set out to examine the ethical aspects of RSs from a consequentialist perspective. Our analysis indicates that the emerging multistakeholder approach to RSs has significant advantages over the more traditional user-centred approach, but it also raises many new interesting research questions. We have highlighted what we see as the main foundational issues in the formulation of the ontology implicit in the design of RS, the way in which they impact the welfare of different stakeholders, as well as their role as social planners in the context of multisided markets with imperfect information. RSs are a ubiquitous feature of digital environments, because they respond to a pressing need to reduce information overload, and facilitate interactions on

multisided platforms with large numbers of participants. The question to ask, therefore, is not *whether* they should be developed, but rather *how* they can be improved.

References

- Abdollahpouri, H., Burke, R., & Mobasher, B. (2017). Recommender Systems as Multistakeholder Environments. *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization - UMAP '17*, 347–348.
<https://doi.org/10.1145/3079628.3079657>
- Burke, R., & Abdollahpouri, H. (2017). Patterns of Multistakeholder Recommendation. *ArXiv:1707.09258 [Cs]*. Retrieved from <http://arxiv.org/abs/1707.09258>
- Chaslot, G. (2019). The Toxic Potential of YouTube’s Feedback Loop | WIRED. Retrieved July 29, 2019, from <https://www.wired.com/story/the-toxic-potential-of-youtubes-feedback-loop/>
- Edizel, B., Bonchi, F., Hajian, S., Panisson, A., & Tassa, T. (2019). FaiRecSys: Mitigating algorithmic bias in recommender systems. *International Journal of Data Science and Analytics*, 1–17. <https://doi.org/10.1007/s41060-019-00181-5>
- Eriksson, M., Fleischer, R., Johansson, A., Snickars, P., & Vonderau, P. (2019). *Spotify Teardown: Inside the Black Box of Streaming Music*. Cambridge, MA: MIT Press.
- Floridi, L. (2016). Tolerant Paternalism: Pro-ethical Design as a Resolution of the Dilemma of Toleration. *Science and Engineering Ethics*, 22(6), 1669–1688.
<https://doi.org/10.1007/s11948-015-9733-2>
- Hausman, D. M. (2016). On the Econ within. *Journal of Economic Methodology*, 23(1), 26–32.
<https://doi.org/10.1080/1350178X.2015.1070525>
- Hausman, D. M. (2018). Philosophy of Economics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018). Retrieved from <https://plato.stanford.edu/archives/fall2018/entries/economics/>

- Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating Collaborative Filtering Recommender Systems. *ACM Trans. Inf. Syst.*, 22(1), 5–53.
<https://doi.org/10.1145/963770.963772>
- Immorlica, N., Mao, J., Slivkins, A., & Wu, Z. S. (2019). Bayesian Exploration with Heterogeneous Agents. *ArXiv:1902.07119 [Cs]*. Retrieved from <http://arxiv.org/abs/1902.07119>
- Infante, G., Lecouteux, G., & Sugden, R. (2016). Preference purification and the inner rational agent: A critique of the conventional wisdom of behavioural welfare economics. *Journal of Economic Methodology*, 23(1), 1–25. <https://doi.org/10.1080/1350178X.2015.1070527>
- Jannach, D., & Adomavicius, G. (2016). Recommendations with a Purpose. *Proceedings of the 10th ACM Conference on Recommender Systems - RecSys '16*, 7–10.
<https://doi.org/10.1145/2959100.2959186>
- Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263–291. <https://doi.org/10.2307/1914185>
- Karakayalı, N., Kostem, B., & Galip, I. (2018). Recommendation Systems as Technologies of the Self: Algorithmic Control and the Formation of Music Taste. *Theory, Culture & Society*, 35(2), 3–24. <https://doi.org/10.1177/0263276417722391>
- Krukowski, D., & Thompson, E. (2019). *Ways of Hearing*. Cambridge, Massachusetts: MIT Press.
- Lev, O., & Tennenholtz, M. (2017). Group Recommendations: Axioms, Impossibilities, and Random Walks. *Electronic Proceedings in Theoretical Computer Science*, 251, 382–397.
<https://doi.org/10.4204/EPTCS.251.28>
- Luckerson, V. (2019, January 16). Is Spotify's Model Wiping Out Music's Middle Class? Retrieved October 15, 2019, from The Ringer website:
<https://www.theringer.com/tech/2019/1/16/18184314/spotify-music-streaming-service-royalty-payout-model>

- Ma, C. (n.d.). The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation. Retrieved October 31, 2019, from The Computational Propaganda Project website: <https://comprop.oii.ox.ac.uk/research/cybertroops2019/>
- Milano, S., Taddeo, M., & Floridi, L. (2019). Recommender Systems and their Ethical Challenges. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3378581>
- Rochet, J.-C., & Tirole, J. (2003). Platform Competition in Two-Sided Markets. *Journal of the European Economic Association*, 1(4), 990–1029.
<https://doi.org/10.1162/154247603322493212>
- Schwartz, B. (2005). *The Paradox of Choice: Why More Is Less* (New edition edition). New York, NY: Harper Perennial.
- Seaver, N. (2018). Captivating algorithms: Recommender systems as traps. *Journal of Material Culture*, 1359183518820366. <https://doi.org/10.1177/1359183518820366>
- Shani, G., & Gunawardana, A. (2011). Evaluating Recommendation Systems. In F. Ricci, L. Rokach, B. Shapira, & P. B. Kantor (Eds.), *Recommender Systems Handbook* (pp. 257–297). https://doi.org/10.1007/978-0-387-85820-3_8
- Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. *Science*, 361(6404), 751–752. <https://doi.org/10.1126/science.aat5991>
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven & London: Yale University Press.