Political Philosophy

Dishaw, Samuel. 2025. The Right to a Justification. *Political Philosophy*, 2(2): 496–520. DOI: https://doi.org/10.16995/pp.23269

Open Library of Humanities

The Right to a Justification

Samuel Dishaw, Philosophy, University of Louvain, Belgium, samuel.dishaw@uclouvain.be

Many institutions and organizations now delegate important decisions to algorithms. These algorithms promise greater predictive accuracy, at a lower operating cost than the human decision-makers they replace. But they also have the distinctive disadvantage of being "black boxes": we lack intelligible explanations of why they arrive at the decisions they do. Those adversely affected by these decisions, it seems, may reasonably object to the opaque nature of the decision-process. My aim in this article is to explain the moral basis of this objection. The account I develop centers on the moral right to a justification. I motivate this view in part by criticizing two rival accounts, one based in the interest in self-advocacy, the other in the requirement of public reason.

The Right to a Justification

Samuel Dishaw
Philosophy, University of Louvain, Belgium

Cash transfer programs are a potentially powerful means of lifting people out of poverty. The World Bank has promoted such programs in several countries in North Africa and the Middle East. In most cases, limited budgets mean that not all eligible applicants can become beneficiaries. The World Bank must thus select among applicants by trying to estimate their income and welfare, a selection process known as "poverty targeting."[1] Such evaluations are fraught with controversy. Human error in complex assessments such as these seems unavoidable, and corruption is often a very real threat.

The World Bank, like similar institutions, has sought to allay these concerns by *automating* the pivotal steps of its decisions about whom to assist. For instance, in the cash transfer program that it recently deployed in Jordan, the World Bank delegated the selection of beneficiaries to an algorithm trained on large sets of data.[2] Each applicant was made to provide information about their household, based on fifty-seven different socio-economic indicators. The algorithm then issued a decision about who the beneficiaries should be, based on this information and on the broader patterns between welfare and these fifty-seven socio-economic indicators, which the algorithm would have previously inferred from its data set.

Outsourcing the decision to an algorithm promises not only some measure of immunity from corruption but also assessments that are, on average, more *accurate*. In all likelihood, the algorithm the World Bank relied on was more reliable in its predictions than the human officials it replaced. After all, machine learning algorithms find patterns in large sets of data that the rest of us simply cannot see.

Yet the very complexity that makes an algorithm predictively powerful can result in that decision-system being, from our point of view, a "black box."[3] In particular, two technical properties of AI systems tend to result in decisions that are opaque to us.[4] One is the sheer number of input features that the system is modelling. The World Bank's algorithm, for example, is making decisions on the basis of fifty-seven different

[1] Hillebrecht et al. 2023.

[2] Human Rights Watch 2023.

[3] Burrell 2016; Selbst & Barocas 2018.

[4] See Grant et al. 2025, pp. 60-61.

socio-economic indicators. The other is that certain systems, such as deep neural networks, model complex and often non-linear functions between these many input features and the output feature they are trying to predict. When a decision-system has these two properties, it will often end up implementing decision rules that are too complex and gerrymandered for us to be capable of understanding them.[5] And when that happens, the decision-process will be, in an important sense, opaque to us. For any particular output, we will not be in a position to know on what basis the algorithm arrived at that decision.

Let us imagine, then, that the algorithm used by the World Bank is precisely such an opaque but highly accurate decision-system.[6] In order to make optimally accurate predictions, the algorithm will be sensitive not only to the contribution of each of the fifty-seven socio-economic indicators, but also to the subtle interaction effects between (any subset of) these indicators. In these circumstances, the World Bank cannot explain why any particular applicant was excluded from the program. The World Bank will know, say, *that* the algorithm found this household to be less poor than most. But it cannot explain *why*, or on what basis, the algorithm arrived at that decision.

Considered from the perspective of the individual hoping to become a beneficiary of the program, the opacity of the decision-making process seems objectionable, at least to some degree. Here is an applicant excluded from the cash transfer program. They are, let us imagine, the head of a large household, with many dependents, and expected the financial aid to come through in part for this reason. Having received the bad news, they ask to know on what grounds they were excluded from the program (perhaps, to make matters worse, our applicant has friends or neighbors in seemingly similar economic circumstances, whose applications nevertheless fared better). This request seems perfectly reasonable. And yet, as we have just seen, the World Bank has no adequate explanation to offer them, for the World Bank itself does not understand why the algorithm chose to exclude this person. Thus, in these circumstances, it is reasonable for the applicant to object to their being excluded from the program *with no explanation*. In other words, the opacity of the decision-making process seems objectionable in itself.

My aim in this article is to provide an account of what, exactly, is the *basis* of this moral objection to opaque decision-making. As I will understand the term, a decision-making process is “opaque” just in case the reasons on the basis of which decisions

[5] *Ibid.*; Fleisher 2022.

[6] Some imaginative stipulation is required here by the fact that the World Bank refused, when pressed, to say what sort of algorithm was used (see Human Rights Watch 2023, p. 11).

are made are not available, or accessible, to those whom the decision concerns. This means, of course, that opacity is not a unique feature of *algorithmic* decision-making, and is shared, for instance, with cases in which the reasons for a decision are kept secret.[7] I focus on algorithms in this article both because their use in decision-making is becoming rapidly widespread, and because they raise the relevant concern in its most acute version, any decision based on such algorithms being *in principle* opaque, given the inherent complexity of the mechanism in question.[8]

Here is how the article will go. In Sections I and II, I criticize two influential accounts of what is objectionable about the use of opaque algorithms in decision-making: the self-advocacy view, and the public reason view. I discuss these views not only to motivate the search for a more adequate alternative, but also to illustrate some of the difficulties attending any attempt to articulate a moral objection to opaque algorithms that is not immediately obviated by reference either to the efficiency or reliability of such algorithms.[9] The heart of the article lies in Section III, in which I introduce and develop my positive view. The view I put forward centers on the *right to a justification*, and supports the claim that there are *non-instrumental* reasons to provide individuals with the *particular* reasons that justified a decision made about them. These two features of the view will enable it to avoid the difficulties identified earlier on. In Section IV, I argue that the right to a justification provides, additionally, useful guidance when it comes to distinguishing those contexts in which the use of opaque algorithms is morally objectionable from those in which it is benign. Section V concludes.

I. SELF-ADVOCACY

In explaining what is objectionable about opaque decision-making, a natural place to begin is by looking at the *interests* of individuals which might be set back, or in some

[7] For a helpful overview of these and other senses of "opacity," see Burrel (2016).

[8] I leave it open whether the problem of understanding the decisions of an opaque algorithm might admit of a technical solution, e.g., via the creation of models that *approximate* the reasoning of an opaque algorithm sufficiently well. This is the ambition of the "explainable AI" research program in computer science (Ribeiro et al. 2016; Jacovi & Goldberg 2020). However, this research program is, among other things, a response to a perceived moral problem: opaque decision-making. To determine whether the technical solution solves the moral problem, we first need to know *what* the moral problem is. That is my topic in this article.

[9] I will simply assume that opaque algorithms are more efficient and reliable than the human decision-makers they replace. I make this assumption not because I think it holds true in all cases but simply to reduce noise, my aim being to illuminate what is problematic about the opacity of algorithmic decision-making as such.

other way threatened, by the use of opaque algorithms. One of the most influential critiques of opaque decision-making, put forward by Kate Vredenburgh, takes exactly this approach. In particular, Vredenburgh argues that the use of opaque algorithms undermines our interest in being able to engage in informed *self-advocacy*.[10]

For someone to be able to engage in self-advocacy is for that person to have a suite of abilities which empower that person to promote and protect their own interests. Two abilities which play an especially important role within Vredenburgh's account are the ability to further one's interests by *navigating* and so conforming to the rules by which decisions are made; and the ability to hold decision-makers *accountable* by rectifying possible mistakes.[11] Being able to advocate for ourselves in these ways is useful. By knowing what rules to follow, and being able to contest their misapplication, you are more likely to come out of a decision-process with the outcome you sought.

As Vredenburgh points out, the use of opaque algorithms directly undermines our ability to advocate for ourselves. If the rules by which decisions are made are opaque, you can hardly navigate those rules by conforming your behavior to them. Nor could you possibly contest a decision you believed was mistaken, by appealing to the rule you think was misapplied in your case, for you do not know *what* rule was meant to be applied in the first place. In short, when decisions are made by opaque algorithms, our ability to advocate for ourselves becomes severely limited.

Moreover, on Vredenburgh's account, the interest in informed self-advocacy is a "morally significant" interest, in the sense that it is an interest sufficient in importance to be worthy of protection by way of a right.[12] We thus have a right to the necessary means of engaging in informed self-advocacy. Since we need to understand why decisions have been made, in order to navigate rules and contest their misapplication, it follows, on her view, that individuals have a right to an explanation—a right grounded in the interest in informed self-advocacy.

I will not dispute the importance of our interest in being able to advocate for ourselves. I think it is clear that we do have such an interest. However, I do not think that

[10] Vredenburgh 2022.

[11] *Ibid.*, pp. 212-13. A third ability Vredenburgh mentions is the ability to have one's interests "represented": to comment on, and ideally provide input into, the content of the rules themselves, which requires rules to be publicly displayed. I discuss this publicity condition separately, in Section II. In any event, the argument I press in this section does not turn on what particular abilities we include within the interest in self-advocacy, but rather on a structural feature of that interest.

[12] See Raz (1986, ch. 7) for this idea that rights serve to protect sufficiently weighty interests.

this interest identifies the correct basis of a supposed right to an explanation, or of the moral objection to the use of opaque algorithms which that right is meant to support.

The problem is that the interest in self-advocacy, important though it may be, is a purely *instrumental* interest. It is an interest we have in possessing the means of advocating our other interests. Because the interest in self-advocacy is instrumental, the use of opaque algorithms may in fact align with it, by promoting those interests in virtue of which the interest in self-advocacy matters in the first place.

To see why the interest in self-advocacy is instrumental, notice that your interest in self-advocacy could not be the only interest that you have. This distinguishes it from our non-instrumental interests. Perhaps you have an interest in French literature. In principle, this could be the only (non-instrumental) interest that you have. It would be strange for that to be your only concern, but not conceptually incoherent. By contrast, your interest in self-advocacy could not, for *conceptual* reasons, be the only interest that you have. The interest in self-advocacy *just is* an interest in having various abilities that are a means of securing or promoting the other, more basic, interests that you have.

Like any other instrumental good, the ability to engage in self-advocacy thus derives its importance from the importance of these other, more basic interests that it serves to promote. To the extent that I have an interest in being able to advocate for myself in the context of some decision-process, this is because I have an interest in the goods that a favorable outcome would provide. Self-advocacy matters because of the more basic interests it enables us to advocate for.

Because the interest in self-advocacy is an essentially instrumental interest, the use of opaque algorithms will in fact align with that interest more often than it initially seemed. That is, opaque algorithms may be more effective at promoting those other interests upon whose importance the interest in self-advocacy ultimately depends. In such circumstances, the use of opaque algorithms will in fact be preferable from the point of view of self-advocacy itself.

Concretely, those who champion the use of opaque algorithms might put the point this way. Automating decisions to an algorithm is a *cost-effective* way of allocating resources, at no loss to accuracy. It is much cheaper to have an algorithm decide where to allocate international aid, than it is to pay several people to do that job. Some of these savings might be redirected back into the amount of resources available for distribution. For instance, the World Bank could add to the program's financial envelope a fraction of the expenses they save by eliminating human labor from the decision-making process. It would now be in the interest of any applicant to be judged by an opaque algorithm rather than a person. The total amount of resources distributed by the algorithm would be slightly larger, and so each applicant's antecedent prospect of receiving a cash transfer would be slightly greater.[13]

In this imagined version of the case, the use of an opaque algorithm is preferable in light of the deeper interests that self-advocacy is supposed to serve. After all, the interest in self-advocacy is instrumental. It is an interest in having the means of promoting our other, more basic interests. In the present case, being able to engage in self-advocacy would matter as a means of securing a favorable outcome, namely, receiving financial aid. But if the use of an opaque algorithm means a larger envelope of available funds, its use will promote precisely those interests for whose sake self-advocacy matters in the first place.

I have just described a way in which the deployment of an opaque algorithm might be in the interest of those who are judged by it. But we might still worry whether we should ever be confident that this is the case. After all, what *evidence* would an applicant have that the use of an opaque algorithm was in their interest, in the absence of receiving an explanation? From the applicant's epistemic perspective, the algorithm is just a black box, which may or may not serve their deeper interests. This suggests that the interest in self-advocacy might still militate against the use of opaque algorithms after all. If you do not know on what basis a system is making its decision, why should you believe that its use will promote your interests?

The answer is that the relevant evidence need not take the form of an explanation.[14] I may have good evidence that a decision-process would serve my interests, even if I lack an explanation of how these decisions will be made. In the case I have described, for instance, the relevant evidence is provided by the fact that more resources will be available for distribution under the opaque decision-process. This fact provides, to each applicant, good evidence that delegating the decision to an opaque algorithm better serves their fundamental interests. In other words, even if the algorithm

[13] This same line of reasoning can be deployed to justify preferring opaque algorithms over other, hybrid decision-making processes that rely on algorithms while also keeping human agents "in the loop." For instance, on Vredenburgh's (2022) own positive proposal, decision-makers should create simplified models of their algorithms, explain their decisions in terms of these models, and be held accountable by decision-subjects if and when mistakes have been made. But again, an organization like the World Bank will point out that all of these interventions and interactions come at a cost, a cost the saving of which would be in the interest of decision subjects themselves.

[14] This distinction is one that Vredenburgh (2022, pp. 218-19) herself makes in the course of defending her view.

itself is opaque, the fact that its use is in your interest need not be similarly epistemically inaccessible.[15]

All else equal, of course, it is better if some of the benefits of automating the decision-process are redistributed to the applicants themselves. Nevertheless, this alone does not seem to absolve the World Bank from its duty to explain its decisions to those on the losing end of them. Intuitively, those excluded from the financial aid program can still demand an explanation, and object to the opacity of the decision-process if no explanation is forthcoming. The problem is that it is not clear how to make sense of that claim, in terms of the interest in self-advocacy. For here is a case in which the use of an opaque algorithm precisely aligns with that interest. This suggests that the interest in self-advocacy is not, in fact, the correct basis of the moral objection to the opacity of a decision-process as such.[16]

We reached this conclusion by way of two simple observations. The first is that the interest in self-advocacy is an instrumental interest. It matters to us to be able to advocate for our interests, as a means of promoting those interests. The second observation is that opaque algorithms are not only more accurate than their human counterparts, but also more cost-effective. By eliminating human labor from the decision-making process, we free up resources available for distribution. Putting those two points together, we get that using opaque algorithms to allocate resources will often be an effective means of promoting those very interests in virtue of which the interest in self-advocacy matters in the first place. In other words, opaque algorithms will often be a better means of promoting these interests than informed self-advocacy itself.

Even when this is true, it will often remain reasonable for people to demand an explanation about why an algorithm arrived at an adverse decision in their case. Thus, if we are to understand what is objectionable about opaque decision-making, we must look beyond self-advocacy. We must identify the basis for a person's entitlement to an

[15] I return to this point and elaborate upon it shortly, in Section II.

[16] In order for the process to be fair, it will be important that applicants are able to engage in informed self-advocacy at least to *some* degree. For instance, it will be important that as many potential applicants as possible are informed about the opportunity of applying for funds; that they in fact have the means of applying (e.g., access to a phone; see Human Rights Watch 2023, p. 5); and that they have sufficient guidance to know how to submit a valid application. My point here is simply that the interest in self-advocacy would not support a right to an *explanation* specifically, since a cost-efficient, opaque decision-process might very well be what better promotes that interest.

explanation in itself, beyond the instrumental contribution of such an explanation to that person's other interests.

II. PUBLIC REASON

The problem for the self-advocacy account arose from the fact that it treats explanations as instrumentally valuable, and so treats the opacity of decisions as merely instrumentally bad. A different and perhaps more promising approach would thus be to explain what is wrong with opaque decision-making by pointing out some way in which such decisions are non-instrumentally bad. One account which attempts to do precisely that is the view that opaque decisions lack legitimacy because they fail the test of *public reason*.

The requirement of public reason says that the rules that govern political life need to be justified on the basis of reasons that are public and acceptable to all reasonable people. Public reason thus requires that rules are open to criticism, and that the reasons put forward in support of these rules are ones that any reasonable citizen could accept.[17]

This requirement provides a straightforward way of objecting to opaque algorithms. By definition, the rules that an opaque algorithm executes are unintelligible to the human mind. How could opaque rules be acceptable to all reasonable people, when people can't even understand those rules in the first place? It thus seems that decisions made on the basis of opaque algorithms could not possibly meet the standards of public reason.[18] In short, the diagnosis is that the use of opaque algorithms is *incompatible* with the ideal of public reason, which sets constraints on the sorts of rules that may acceptably be employed to govern social life. Since opaque algorithms seem to fail the test of public reason almost by definition, this argument can thus seem to pose a significant challenge to their use.

I will not dispute that the ideal of public reason sets important requirements on institutions. However, I do not think that it provides especially strong reasons against using opaque algorithms within such institutions. In particular, it seems to me that the argument from public reason overlooks an important distinction between two kinds

[17] Rawls 1996. See also Quong 2011.

[18] Binns 2018; Maclure 2021. Vredenburgh (2022) also emphasizes the importance of rules being open to criticism, insofar as this provides individuals with an opportunity to represent their interests to policy makers. Of course, on her view this requirement too is ultimately grounded in the interest in self-advocacy, and so is vulnerable to the worries discussed in Section I.

of rules. Once we bring that distinction into view, the alleged tension between opaque algorithms and the ideal of public reason becomes much less obvious.

The distinction I have in mind is the following. In addition to rules on the basis of which we make decisions, we also have rules about what decision processes to use in the first place. That is, we can distinguish between rules we follow in deciding what to do, and rules we follow in choosing when and how to go about making decisions. Rules of the second kind are "higher-order" rules, in the following sense: when we follow them, we are implementing a decision about whether and how to make decisions.

Here's an example: you might have a rule against buying plane tickets, or making any important financial decision, late in the evening. Perhaps you have this rule because you have noticed that it often turns out badly when you make such decisions at that time of day. This is a second-order rule. When you adopt this rule, you have made a decision about whether and how to make a certain kind of decision. The particular rule I have just mentioned is negative: it instructs you *not* to make certain financial decisions, under certain circumstances. But higher-order rules can also be positive. For instance, you might adopt the rule of delegating your flight itineraries to your travel-savvy friend. This too is a second-order rule. When you adopt it, you are making a decision about how to make decisions about flight itineraries.

I take it that the phenomenon I have just described is a familiar feature of our individual lives. We don't always rush head on into deliberation, applying whatever rules seem right to us at the time. We also pause and ask ourselves, "How should I decide what do here?" In some cases, we might conclude that the best way to decide is to let someone else decide for us.

What holds within our individual lives also holds, in this respect, for social and political institutions. Just as you and I have rules about how to make decisions, so too do our institutions. For instance, we have rules for delegating certain specific policies (e.g., energy policy) to scientific experts. Even when we do not fully delegate those decisions to experts, we may nevertheless give their expert opinion significant weight in deciding which energy policies to adopt.

Crucially, the reasons behind the judgments or decisions of experts will not always be intelligible to the general public.[19] This gap in scientific understanding is precisely why experts are needed in the first place. But even though the reasoning behind the decisions of experts may not be intelligible to the general public, the second-order rule of delegating our energy policy to experts (or of giving their recommendation a significant weight in these decisions) may nevertheless be justified from the point of view of public reason.

[19] See Nguyen (2022) for a compelling defense of this claim.

In defending this second-order decision, we might say: relying on scientific experts to determine our energy policy is the best way to protect our collective interests as well as those of future generations. In making this claim, we would provide publicly accessible criteria that mark certain people as the relevant experts.[20] Even if the reasoning of these experts is not accessible to the general public, the fact *that* these people are experts very well may be.[21] In these circumstances, we would thus justify the rule of delegating the relevant decisions to experts on the basis of reasons that are public and acceptable to all.[22]

In light of this, we can now see how to reconcile the use of opaque algorithms with the requirement of public reason. Just as we can justify delegating certain decisions to experts on the basis of public reasons, so too we can justify delegating certain decisions to algorithms in the same way. For instance, suppose we adopt the following second-order rule: in deciding where to allocate scarce resources, such as the financial aid distributed by the World Bank, we should make decisions based on the recommendations of a machine-learning algorithm. In defending this second-order rule, we can appeal to values that are shared by all reasonable people. After all, we are assuming that the algorithm is on average more accurate or reliable than the human decision-makers it has replaced. The algorithm is also optimizing for a goal whose importance no reasonable person could reject: that of allocating resources to those who most need it.

This means that the requirement of public reason does not, in fact, provide a clear basis for the moral objection to opaque decision-making as such. Although the rules that an opaque algorithm executes—its “expert” reasoning—are not intelligible to the general public, the *second-order* rule of delegating certain decisions to such algorithms (or of giving their recommendation a significant weight in these decisions) may, in principle, be justified from the point of view of public reason. In justifying its decision to rely on algorithms to distribute its resources, the World Bank will point out that the algorithm is, on average, more accurate than human decision-makers at determining which applicants would most benefit from the funds. Even if the reasoning of this algorithm is not accessible to the general public, the fact *that* it is more accurate than any human decision-maker may very well be. In this way, the decision to rely on an opaque algorithm can thus be reconciled with the requirements of public reason.

[20] Anderson 2011.

[21] Of course, whether or not claims about the *expertise* of certain people can be justified on the basis of reasons acceptable to all will depend on some contingent social conditions, including whether citizens have readily accessible evidence for making judgments about who the experts are. On this point, see Anderson 2011.

[22] See Rawls (1996, p. 224) for a statement of the idea that certain scientific truths should be treated as public reasons. For further discussion and defense of this claim, see Jønch-Clausen & Cappel (2016) as well as Badiola (2018).

And yet, as I have urged, it nevertheless seems perfectly reasonable for the person on the receiving end of the algorithm's verdicts to object to being excluded from the program with no explanation. What we still need, then, is a way of understanding this particular moral objection, an objection that persists even in those circumstances in which there are good, publicly accessible reasons for delegating decisions to an opaque algorithm. In other words, we need an account of this moral objection capable of explaining why the ascent to second-order rules and their justification is blocked in the present case. This is the task to which I now turn.

III. THE RIGHT TO A JUSTIFICATION

In this section, I defend a positive account of the basis for a person's objecting to the opacity of a decision-making process in itself. My account centers on the moral right to a *justification*, and has two central parts. The first is a view of the conditions under which individuals have a right to a justification. The second is a view of the content of this right: of what individuals are entitled to when they have a right to a justification.

On my view, the right to a justification is based directly on the other *rights* of the person, rather than on the interests of the person which would be usefully promoted by the possession of an explanation. This feature of the view will help us to see why it can be reasonable to object to opaque algorithms even when their use in decision-making would be in one's expected interest. This same feature of the right to a justification—its basis in the other rights of the person—will also help us explain why the relevant demand for a justification cannot be met simply by pointing to the general accuracy or reliability of the decision-making process in question. In other words, the right to a justification will allow us to avoid the difficulties that arose for the two views examined so far, and thereby help us to see what is objectionable about opaque decision-making as such.

A. The Right to a Justification: When

Many of our moral obligations are obligations that we owe to a particular person. In these cases, we say that I have a "directed" duty to you. To this duty corresponds a claim, or right, that you have against me. My being obligated to you to perform some action is equivalent to your having a claim against me that I perform this action.[23]

[23] I will be using "claim" or "right" interchangeably (Thomson [1990] speaks of "claim-rights"). Strictly speaking, not all rights correspond to, or covary with, a duty owed to the right-bearer. In principle, one may have a right to perform some action, in the sense of being under no duty not to perform that action, even if others are allowed to interfere with that action (that is, one has no claim to non-interference). See Wenar (2005) for a unified theory

For instance, suppose we are both attending a workshop which is being held in a small town in Eastern Europe. I am familiar with the town, you are not. So, we arrange that I will meet you at the airport and accompany you to your hotel when you arrive. By promising to pick you up, I have put myself under an obligation that is owed to you, specifically. I owe it to you to meet you at the airport when your flight lands. To this directed duty of mine corresponds a claim of yours. You can hold me to this obligation. You have a claim against me that I be there to pick you up when your flight lands.[24]

The right to a justification on which my account centers is a claim-right of just this sort. When you have right to a justification, that right makes a claim on another person, the person who owes you such a justification. In this sense, the right to a justification is just one right among others.

In another sense, however, the right to a justification is a common denominator of *all* rights. This is because the right to a justification is itself a concomitant of any of our other rights, it being a distinctive mark of rights, in general, that they generate a residual claim to a justification whenever they go unmet. Whenever any of our rights is infringed, we thereby have a right to a justification as to why they were infringed.[25]

of rights that includes such "privileges." For the purposes of this article, I will be restricting my attention to those rights that make a claim on another person's conduct.

[24] Thomson 1990, ch. 12.

[25] In moral philosophy, the idea that we owe justifications to other people is sometimes invoked in a foundational capacity (see Scanlon 1998; Forst 2011). Elements of the idea also show up, in passing, in discussions about rights and compensation (see Montague 1988, p. 350), But the most sophisticated account of the duty to justify ourselves is to be found in the philosophy of criminal law, where it has been explored at length by Anthony Duff (2007). Duff argues that we may be under a duty to answer a certain criminal charge, even if we were justified in acting as we did, or if we have an excuse, or indeed even if we are entirely innocent of the accusation. The view I will lay out in what follows is broadly congenial to Duff's own, but there are nevertheless some differences worth mentioning. The most obvious is that Duff's account is (by design) narrower in scope. His is an account of answerability for criminal conduct, whereas I am concerned with the infringement of moral claims more broadly, many of which (e.g., broken promises) we would not want to criminalize. More important, however, are certain *structural* differences between Duff's account and my own that result from our different focal points. For Duff, criminal conduct is a kind of public wrong. This means that we are answerable to members of our polity at large (Duff 2007, p.123 and p.142), rather than any individual in particular, and that we can be answerable for conduct that has not violated anyone's rights (such as the destruction of public goods). By contrast, on the account I will defend, it is central that the duty to justify ourselves is owed to a specific individual, the individual whose moral rights our conduct has infringed. I am grateful to an editor of this journal for bringing Duff's account to my attention.

Suppose that, when you land in the small Eastern European town, I am nowhere to be found. You have to get by on your own, with no cell service and no knowledge of the local language. You had a claim against me that I perform some action—pick you up from the airport—and I didn't. I infringed your claim.[26] As a result, you are, among other things, now entitled to an explanation. Equivalently, we might say that my broken promise to you leaves in its wake a residual duty: I owe it to you to explain why it is that I was nowhere to be found.

Notice, moreover, that I am not exempt from this duty even in the event that my action was, as a matter in fact, irreproachable. Suppose, as it turns out, that the reason I missed your arrival is this: a colleague I was with sprained their ankle on the small town's cobbled streets, and I helped them hobble to the nearest clinic to get an X-ray. Given these circumstances, it was permissible for me to miss your arrival. But this does not exempt me from the obligation to justify my action to you. When I miss your arrival, you are entitled to a justification, whether or not my action was in fact permissible. You are entitled to a justification, precisely because you are not, from your point of view, in a position to determine whether or not my action treated you justifiably.

As a general matter, then, our rights include within them a right to a justification if and when they are infringed. Those whose claims are infringed, even permissibly, have a right to know on what grounds. If there are any rights at all, there is also a right to a justification.[27]

We now have a view as to the conditions under which a person has a right to a justification.[28] To understand how this right provides a moral basis for objecting to

[26] I am using the notion of an infringement in its neutral sense. That is, as I am using the term, the infringement of a claim may be either permissible or wrong all things considered (cf. Thomson 1990, p. 122).

[27] This way of formulating the account is most congenial to a view of claim-rights as non-absolute, that is, as allowing that there are cases in which claims may *permissibly* be infringed. Indeed, the very notion of a residual duty, such as the duty to justify oneself, is often cited in support of non-absolutism (see Montague 1988; Kamm 1996, p. 312). The idea here is that the presence of a right, permissibly infringed, is precisely what explains the residual duties which such an infringement brings in its wake. However, the account I have provided here could, in principle, be transposed within an absolutist framework. After all, absolutists agree that there are such residual duties (Shafer-Landau 1995). They simply explain the existence of these residual duties in a different way, for instance on the basis of the interests which *would* normally have generated a right, in the absence of competing considerations (Wallace 2019, pp. 174-175).

[28] I have focused here on a *sufficient* condition for the right to a justification: the infringement of another right. I do not think that this is the only condition that may activate the right to a justification, but I will focus on it here for ease of exposition.

opaque decisions, however, we need one more thing: an account of *what* a person is entitled to receiving when they have a right to a justification.

B. The Right to a Justification: What

Suppose I have infringed some legitimate claim of yours. I now owe you a justification. What, exactly, is it that I am obligated to provide?

The content of this obligation must be at least broadly continuous with its basis. I have a duty to justify myself to you as a result of having acted in a way that infringed one of your claims. It stands to reason that the justification I provide you should be *responsive* to that state of affairs. That is, it should answer to the particular conduct of mine that created a need for it in the first place.

In particular, an adequate justification should show that, even though your claims were not met, they nevertheless were given an appropriate weight in one's deliberation. It must show that your claims were, as we might put it, respected if not fulfilled.[29] This means that the relevant moral justification must meet three basic conditions.

First, it must be *intelligible* to the person to whom it is provided. An adequate justification is one that shows, to the person concerned, that their claims were taken seriously even if those claims were not met. For a justification to achieve this, it must, at the very minimum, be intelligible to the person to whom it is offered.

Second, the relevant justification must provide the *particular* reasons that justified one's conduct, that is, the reasons that made it permissible to infringe this person's rights. These are the only reasons which are apt to show to the person that their claims, in particular, were taken seriously. They are the only reasons that are responsive to the specific state of affairs that called for such reasons in the first place.

This second aspect of the duty to justify ourselves can be further sharpened by way of contrast. Indeed, the idea that a moral justification must be responsive to the particular infringement that prompted it rules out, as inadequate, any attempt to justify ourselves to others by appeal to the *general* reliability of one's conduct or reasoning. Such facts may provide some evidence that I have acted permissibly in any given case, but they are not responsive to the specific situation that called for a justification. My track-record may show that *I* am generally conscientious, but it does not settle whether my conduct respected *your* claims.

[29] I borrow this distinction between "respecting" and "fulfilling" an obligation from David Owens (2012, pp. 90-91).

Consider again the case in which I left you stranded at your arrival in the small Eastern European town. Running into me at the workshop later that day, you might reasonably ask me what happened, and how come I didn't show up. Suppose I were to answer by saying something along the lines of the following: "I know, I'm sorry. But rest assured, I'm a *very* conscientious person. Generally speaking, I break my promises only when I have good reasons to do so. In fact, if you ask around, you'll find that I have a stellar track-record when it comes to doing the right thing."

Notice that this response is inadequate even if it is true. It is inadequate simply because it is unresponsive to what is at issue. What is at issue is not whether I am generally morally conscientious, or whether I have a good track-record of breaking promises only when it is permissible, or anything like that about me. Rather, what is at issue is whether in my conduct I have accorded sufficient importance to *your* rights—whether I have acted permissibly towards you. Because it is your claim that was infringed, after all, you are within your rights to demand an explanation that shows whether that claim was permissibly infringed or not.

Finally, the reasons one provides must also align with one's operative reasons.[30] For instance, if the reason why I didn't pick you up from the airport was, in fact, that I wanted to mingle with the panel of famous scholars, I may not cite as my justification that the roads were icy that day. If I do, you may reasonably reject my justification, precisely on the grounds that it was not the reason why I didn't pick you up.[31]

In sum, the right to a justification entitles its bearer to the particular reasons in virtue of which another's conduct towards them was justified or permissible. These considerations must also be intelligible, and align with the agent's operative reasons. Only when all three conditions are met can a justification serve its purpose: to show its addressee that their rights were taken seriously, if not satisfied.

This account of the content of the right to a justification is continuous with the account I have given of its basis. We have a right to a justification when a person's action or decision infringes another of our rights. What such a justification must do is explain why this particular infringement was permissible. It must provide the reasons in virtue of which one was justified in acting against another person's rights, and in light of which that person can thus see that their rights were nevertheless taken seriously.

[30] See Scanlon (1998, pp. 18-20) for this notion.

[31] Duff 2007, p. 281; Gardner 2011, p. 87.

C. The Right to a Justification and Opaque Decision-Making

Having provided an account of the right to a justification, I will now argue that this right provides a moral basis for those who would object to decisions made on the recommendation of an opaque algorithm.

Consider, once more, the applicants to the World Bank's financial aid program. Many of those who apply to this program make a legitimate moral claim on the World Bank: they have a right to the necessary means of meeting their most basic needs.[32] Given that resources are limited, not everyone can receive financial aid. In these circumstances, some of these legitimate claims will have to go unmet. The decision to exclude some individuals from the program may very well be justified. The claims of these individuals might, in principle, be outweighed by the competing claims of other applicants whose material needs are greater. Nevertheless, since those who are excluded from the program have a legitimate claim to financial aid, they have, as a result, a right to a justification if that claim goes unmet. They have a right to know what, if anything, justified their exclusion from the program.

The right to a justification provides a strong reason against relying on opaque algorithms to make these decisions, for the use of these algorithms guarantees, ahead of time, that no such justification will be available. As a result of outsourcing its decisions to an opaque algorithm, the World Bank doesn't know why certain applicants were excluded rather than others. Since it does not know why these applicants were excluded from the program, the World Bank cannot provide any such reasons to those excluded. A fortiori, it cannot provide reasons capable of justifying the relevant decisions. Those whose claims go unmet can thus object to the World Bank's reliance on an opaque decision-making system. They can object to the World Bank's relying on a decision-making process that precludes, ahead of time, the provision of that to which these individuals have a right: a justification for their exclusion from the program.[33]

[32] The selection process was made in two stages (Human Rights Watch 2023, p. 2). In the first stage of the process, agents of the Jordanian government determined whether an applicant household was eligible, which required living under the official poverty line. In the second, the algorithm ranked the applicant households deemed eligible. Every applicant household that the algorithm excluded from the program was thus under the official poverty line.

[33] Of course, the identity of those to whom a justification is owed, because they have been excluded from the program, will not be known in advance. What is known in advance is that many individuals will have a right to a justification, which the World Bank will not be in a position to provide, or, in other words, that the World Bank will fall short of its obligation towards many (as of yet unidentified) individuals. This seems to me sufficient to generate a reason against using an opaque algorithm, at the time at which that decision would be made.

Against this, it might be argued that the World Bank *can* justify their decision to those excluded from the program, despite relying on an opaque algorithm. For the ranking provided by the algorithm is itself directly morally relevant. The financial aid ought to go to those who need it most, and input into that decision is precisely what the algorithm provides, when it ranks different households in terms of the severity of the poverty they face. Thus, it might seem that the World Bank is in a position to provide an adequate moral justification after all. To each of the households excluded from the program, it can say: limited resources allowed us to extend aid to a limited number (*n)* of households, and your household was not among the *n*-poorest (as determined by our algorithm).[34]

This justification would not, I think, be adequate as it stands. The problem is that the consideration we are imagining being put forward as a justification—that a household is not among the *n*-poorest—is itself what we might call a "summative" judgment: a judgment about the balance of various morally significant reasons. It thus falls short of providing the particular *reasons* that justified or made it permissible to exclude this or that household from the program. For it already embodies a conclusion about what is justified by the balance of such reasons.

By way of analogy, consider another variation on our promise example. Suppose that, when my colleague asks why I never showed up, I reply: "something unforeseen came up, which was of greater moral importance." Although a step in the right direction, this justification would be inadequate on its own. That something of greater moral importance came up is itself a judgment about the balance of moral reasons. What I owe my colleague is not this summative judgment, but rather the particular reason that made it the case that the moral balance tilted in this way (in this case, the fact that a different colleague sprained their ankle and needed help getting to the hospital).

Although this is less immediately obvious, the consideration cited by the World Bank is a summative judgment in just the same way. The notion of "poverty" that the World Bank employs is a multidimensional concept meant to serve as a measure of human welfare, not merely a measure of income.[35] Factors such as dwelling characteristics, health, economic opportunities, and access to services are all aspects of poverty in this morally nuanced sense. None of these aspects are reducible to one another. A household's higher-than-average income might provide a reason against

[34] I am grateful to David Gray Grant and an anonymous reviewer for pressing this objection.

[35] Human Rights Watch 2023, pp. 14, 141. This multidimensional concept of poverty is explained in World Bank (2025). This notion takes explicit inspiration from Amartya Sen's capabilities approach to welfare, which Sen (1992, p. 5) characterizes in terms of a person's having the capability to achieve functionings that "he or she has reason to value."

its receiving aid, but its remote rural location might count in its favor, given the lack of services and economic opportunities in that area. The conclusion that a household is not among the *n*-poorest thus corresponds to a summative judgment about the balance of these and other morally significant reasons. Even if the algorithm is correct in its overall determination that a household is not among the *n*-poorest, citing this fact would thus not be adequate as a justification. What the excluded household is owed is not this summative judgment, but rather the reasons that justified their exclusion from the program (by way of making the relevant summative judgment true). These reasons are precisely what is unavailable, when decisions such as these are outsourced to an opaque algorithm.

I thus take it that the right to a justification provides a strong presumption against the use of an opaque algorithm in this case. Notice, moreover, that the right to a justification grounds a moral objection to the use of opaque algorithms that is essentially non-instrumental in nature. The person whose legitimate claims have gone unmet is owed a justification as a matter of respect. They have a right to see, or at least to be in a position to appreciate, that their claims were taken seriously even though those claims were not satisfied. Such a justification is owed to someone out of respect for that individual and their rights, regardless of whether its possession will be advantageous to them in the pursuit of their other interests.

This account, unlike the self-advocacy view, can thus explain why one may object to an opaque decision process even when that process is more efficient at satisfying people's interests, including one's own. Even if the World Bank reinvests into the financial aid program some of the expenses it saves by outsourcing the decision to an algorithm, many legitimate claims to aid will still go unmet. Those whose claims go unmet will still have a right to a justification as to why they were excluded from the program, and they may still reasonably object to the World Bank's failure to provide such a justification.[36]

Nor can this justificatory burden be met simply by pointing to the general accuracy or reliability of the relevant decision-making system. As we have seen, the right to a justification entitles its bearer to the particular reasons in virtue of which the

[36] The right to a justification also differs from the self-advocacy view in terms of its content. As an instrumental account, the self-advocacy view supports the provision of (explanatory) information that is *useful* to its recipient. For such purposes of useful guidance, a simplified explanatory model may often suffice, rather than the actual reasons on which the decision was based (see Vredenburgh 2022, p. 225). In this sense, Vredenburgh's account and my own might be seen as complementary aspects of a broader requirement to explain decisions, one which requires different *kinds* of explanations depending on the basis of that right in a given case.

infringement of their claims was permissible. The duties to which this right correspond thus differ, in this crucial respect, from the requirements of public reason. The person whose claims have gone unmet is owed the reasons that justified their exclusion from the program within this decision-process, not reasons in favor of the prior decision to rely on this decision-making process rather than another.

In short, the right to a justification captures what is objectionable about opaque decision-making in itself. Those who apply to the World Bank's program have a legitimate claim to financial aid. Those whose claims go unmet have a right to a justification. No such justification will be available if the relevant decisions are made by an opaque algorithm. The right to a justification thus provides a basis for objecting to the use of such opaque systems.

It bears stressing that this moral objection need not always be decisive. In introducing the right to a justification, we saw that rights may sometimes be permissibly infringed. Indeed, the right to a justification corresponds to the residual obligation which such permissible infringements leave in their wake. But the right to a justification itself is just one right among others. Thus, there may, in principle, be cases in which this right too may permissibly be infringed. In these cases, we would have to justify the infringement of that right, that is, the failure to provide any justification to those entitled to one. Perhaps, in some cases, the benefits of using an opaque algorithm are sufficiently great, or the costs of all alternatives sufficiently high, that the right thing to do, all things considered, is for decision-makers to outsource the decision to an opaque system.[37] We should leave room for such cases. The right to a justification is not meant to provide an absolute prohibition on opaque decision-making. Rather, it is an account of what is objectionable about opaque decision-making in itself.

IV. RIGHTS AND HIGH STAKES

My primary aim in this article has been to understand the *basis* on which a person may reasonably demand an explanation. Insofar as it underlies this demand, however, the right to a justification will, to a significant extent, also serve to define its *scope*. I close

[37] One consideration that will *not* in itself be sufficient to justify the use of an opaque system (and so the infringement of a person's right to a justification) is the fact that it is in a person's *interest* that such a system be used. In general, one cannot justify infringing another person's rights simply on the grounds that it was in that person's interests. (If the pack of cigarettes belongs to you, I cannot steal it from you simply because smoking is bad for your health. If I promise to give you tickets to the local theatre performance, I cannot renege simply because I come to realize you would be better off not going, given your other commitments. And so on.)

by providing a brief sketch of the account's scope, using this as an opportunity to circle back to the role of experts in political life, now seen through the lens of the right to a justification.

In the first instance, it seems to me that the right to a justification can provide useful guidance for distinguishing those contexts in which the use of opaque systems is morally problematic, from those in which it is benign. The familiar exhortation is to avoid using opaque systems in so-called "high-stakes" situations.[38] The right to a justification can help us make this exhortation more precise.

For instance, consider the oft-cited example of using an opaque algorithm to grant or deny bail to defendants.[39] This is a paradigmatic instance of a high-stakes case. Whether one is granted or denied bail has an obvious and immediate impact on one's life. The account I have put forward here, however, can help us say something more. After all, it is not only in a person's *interest* to be granted bail while they await trial; they also have a presumptive *claim* against being imprisoned until proven guilty. To deny a person bail is to restrict one of their most basic liberty rights: freedom of movement. Any such decision must thus be justified to the person whose freedom it restricts, on the basis of the particular reasons that justify this infringement. The use of an opaque algorithm precludes the provision of such reasons. That is why it is objectionable.

By contrast, consider the case of college admissions. College admissions are also high-stakes. Whether and where one goes to college may be life-transforming. The Harvard College Admissions Office is making "high-stakes" decisions in any familiar sense of that term. Yet it does not seem to me that the Admissions Office owes an explanation to the many applicants it does not admit in any given year. If the Admissions Office had reason to think that an opaque algorithm would do a better job of identifying academic potential, together perhaps with creating cohorts whose members would learn from each other, then I do not think it would be objectionable, in itself, if the Admissions Office decided to rely on such an algorithm. Of course, we would still want to be sure that the algorithm really is reliable, that it treats all applicants fairly, and so on.[40] But there would be nothing objectionable about the opacity of the decision-process as

[38] Rudin 2019: Coyle and Weller 2020, p. 1433; Vredenburgh 2022, p. 226. For an argument that many views (including some of the views discussed earlier on in this article) do not have a plausible scope, see Fritz (2025).

[39] Morin-Martel 2024.

[40] Since the algorithm is opaque, our confidence that it is fair would have to be based in its meeting certain statistical criteria of fairness (see Hedden [2021] for a critical examination of such criteria).

such—about the Admissions Office's inability to provide applicants with an explanation for its decisions.

The right to a justification can help us see why. Admission to an Ivy League school is not something to which one has a moral right. Thus, the decision not to admit someone leaves no justificatory burden in its wake. There are thus "high-stakes" cases in which the use of an opaque algorithm is not necessarily objectionable.

By the same token, there will be high stakes decisions in political life in which reliance on (opaque) expert reasoning need not be objectionable, either. For example, suppose my country's central bank sets a benchmark interest rate of 3%. This decision is highly consequential, affecting a large number of individuals. But it does not infringe anyone's moral rights. I do not have a moral claim that my country's central bank set the interest rate in any particular way. Thus, even though the decision to set the benchmark interest at 3% carries high stakes, it does not generate a right to a justification.

For similar reasons, I think, we should have no qualms about allowing expert reasoning to drive policy-making in other important areas of public policy. We should want experts to inform decisions about which vaccines to offer to the general public, what road infrastructures to prioritize, or how to efficiently transition to renewable energy sources. Although this expert reasoning will, of necessity, be opaque to most, the decisions such reasoning supports need not infringe anyone's moral rights. If those decisions do not infringe anyone's moral rights, they will leave no justificatory burden in their wake—no justificatory burden, at any rate, of the sort I have argued for here.

V. CONCLUSION

In this article, I have developed an account of the moral objection to opaque decision-making, one which is based on the right to a justification. As I have argued, such a right to a justification is part of the very basic fabric of moral rights. The resulting account is thus one which is independently well-motivated. It also, I argued, marks an improvement over two prominent views. Unlike the self-advocacy account, the right to a justification emphasizes the *non-instrumental* importance of providing an explanation. Unlike the public reason account, the right to a justification entitles its bearer to the *particular* reasons that made it permissible to infringe that person's claims. Because of this, the right to a justification provides a basis for the moral objection to opaque decision-making that is not easily obviated by reference either to the efficiency or overall reliability of opaque algorithms.

In addition, I have suggested that the right to a justification provides useful guidance for determining whether, in any given case, the opacity of a decision-making process is, in itself, an objectionable feature of that process. This is because the right to a justification arises in a principled way: as a result of the infringement of any of our other rights. When considering the permissibility of deploying opaque algorithms, that is what we should attend to. Contrary to the familiar exhortation, the important question is not whether the stakes are high, but whether rights are at stake.

ACKNOWLEDGMENTS

For their written comments, I am grateful to Jenna L.A. Donohue, Megan Entwistle, David Gray Grant, and Milo Phillips-Brown, as well as Jeff Howard and three anonymous referees for this journal. For helpful feedback on earlier versions of this article, thanks also to participants at: the Philosophy, Politics & Economics Society's 8th Annual Meeting; the UC-Santa Cruz Philosophy Colloquium; and the Mich Seminar of the University of Louvain.

COMPETING INTERESTS

The author declares that he has no competing interests.

REFERENCES

Anderson, Elizabeth. 2011. Democracy, public policy, and lay assessment of scientific testimony. *Episteme*, 8: 144-164. https://doi.org/10.3366/epi.2011.0013

Badiola, Cristóbal Bellolio. 2018. Science as public reason: a re-statement. *Res Publica*, 24: 415-432. https://doi.org/10.1007/s11158-018-09410-3

Binns, Ruben. 2018. Algorithmic accountability and public reason. *Philosophy & Technology*, 31: 543-556. https://doi.org/10.1007/s13347-017-0263-5

Burrell, Jenna. 2016. How the machine 'thinks': understanding opacity in machine learning algorithms. *Big Data & Society*, 3: 1-12. https://doi.org/10.1177/2053951715622512

Coyle, Diane and Weller, Adrian. 2020. "Explaining" machine learning reveals policy challenges. *Science*, 268: 1433-1434. https://doi.org/10.1126/science.aba9647

Duff, Anthony. 2007. *Answering for Crime: Responsibility and Liability in the Criminal Law*. Oxford: Hart Publishing.

Fleisher, William. 2022. Understanding, idealization, and explainable AI. *Episteme*, 19: 534-560. https://doi.org/10.1017/epi.2022.39

Forst, Rainer. 2011. *The Right to Justification: Elements of a Constructivist Theory of Justice*. New York: Columbia University Press.

Fritz, James. 2025. On the scope of the right to explanation. *AI and Ethics*, 5: 2735-2747. https://doi.org/10.1007/s43681-024-00586-4

Gardner, John. 2011. Relations of responsibility. Pp 87-102 in *Crime, Punishment, and Responsibility: the Jurisprudence of Anthony Duff*, ed. R. Cruft, M. Kramer and M. Reiff.

Grant, David Gray et al. 2025. What we owe to decision-subjects: beyond transparency and explanation in automated decision-making. *Philosophical Studies*, 182: 55-85. https://doi.org/10.1007/s11098-023-02013-6

Hedden, Brian. 2021. Statistical criteria of fairness. *Philosophy & Public Affairs*, 49: 209-231. https://doi.org/10.1111/papa.12189

Hillebrecht, Michael et al. 2023. The dynamics of poverty targeting. *Journal of Development Economics*, 161: 1-9. https://doi.org/10.1016/j.jdeveco.2022.103033

Human Rights Watch. 2023. *Automated Neglect. How the World Bank's Push to Allocate Cash Assistance Using Algorithms Threatens Rights*. https://www.hrw.org/sites/default/files/media_2023/11/thr_jordan0623%20web.pdf

Jacovi, Alon and Yoav Goldberg. 2020. Towards faithfully interpretable NLP systems: how should we define and evaluate faithfulness? *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 4198-4205. https://doi.org/10.18653/v1/2020.acl-main.386

Jønch-Clausen, Karin and Klemens Cappel. 2016. Scientific facts and methods in public reason. *Res Publica*, 22: 117-133.

Kamm, Frances. 1996. *Morality, Mortality Volume II: Rights, Duties, and Status*. Oxford: Oxford University Press. https://doi.org/10.1093/0195144023.001.0001

Maclure, Jocelyn. 2021. AI, explainability and public reason: the argument from the limitations of the human mind. *Minds and Machines*, 31: 421-438. https://doi.org/10.1007/s11023-021-09570-x

Montague, Phillip. 1988. When rights are permissibly infringed. *Philosophical Studies*, 53: 347-366. https://doi.org/10.1007/BF00353511

Morin-Martel, Alexis. 2024. Machine learning in bail decisions and judges' trustworthiness. *AI and Society*, 39: 2033-2044. https://doi.org/10.1007/s00146-023-01673-6

Nguyen, C. Thi. 2022. Transparency is surveillance. *Philosophy and Phenomenological Research*, 105: 331-361. https://doi.org/10.1111/phpr.12823

Owens, David. 2012. *Shaping the Normative Landscape*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199691500.001.0001

Quong, Jonathan. 2011. *Liberalism without Perfection*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199594870.001.0001

Rawls, John. 1996. *Political Liberalism*. New York: Columbia University Press.

Raz, Joseph. 1986. *The Morality of Freedom*. Oxford: Clarendon Press. https://doi.org/10.1093/0198248075.001.0001

Ribeiro, Marco Tulio et al. 2016. 'Why should I trust you?': Explaining the predictions of any classifier. *Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining*, 1135-1144. https://doi.org/10.48550/arXiv.1602.04938

Rudin, Cynthia. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1: 206-215. https://doi.org/10.1093/biostatistics/kxy002

Scanlon, Thomas. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Selbst, Andrew D. and Solon Barocas. 2018. The intuitive appeal of explainable machines. *Fordham Law Review*, 87: 1085-1139. https://doi.org/10.2139/ssrn.3126971

Sen, Amartya. 1992. *Inequality Re-Examined*. Oxford: Oxford University Press. https://doi.org/10.1093/0198289286.001.0001

Shafer-Landau, Russ. 1995. Specifying absolute rights. *Arizona Law Review*, 37: 209-225. https://journals.librarypublishing.arizona.edu/arizlrev/article/id/8119/

Thomson, Judith Jarvis. 1990. *The Realm of Rights*. Cambridge, MA: Harvard University Press.

Vredenburgh, Kate. 2022. The right to explanation. *Journal of Political Philosophy*, 30: 209-229. https://doi.org/10.1111/jopp.12262

Wallace, R. Jay. 2019. *The Moral Nexus*. Princeton, NJ: Princeton University Press.

Wenar, Leif. 2005. The nature of rights. *Philosophy & Public Affairs*, 33: 223-252. https://doi.org/10.1111/j.1088-4963.2005.00032.x

World Bank. 2025. Brief: multidimensional poverty measure. https://www.worldbank.org/en/topic/poverty/brief/multidimensional-poverty-measure.