Problems With the Veil of Ignorance, And How We Might Solve Them

Anthony Persson

Supervisor: Jami Weinstein
Examiner: Martin Berzell
# Table of contents

Introduction ............................................................................................................................................. 2

Background ................................................................................................................................................ 2

Rawls’ Veil of Ignorance .......................................................................................................................... 2

Michael Sandel’s Critique .......................................................................................................................... 5

An Experiment about the Veil of Ignorance ............................................................................................ 6

Is the Veil of Ignorance Really Compatible with Egalitarianism? .......................................................... 9

A Feminist Critique of Rawls’ Theory of Justice ...................................................................................... 12

The Problem of Abstraction ..................................................................................................................... 12

Missing the Roots of Injustice Towards Women ...................................................................................... 13

The Veil of Ignorance Favors the Greater Good ...................................................................................... 15

The Veil of Ignorance in Clinical Research ............................................................................................. 17

My Solution ................................................................................................................................................ 18

The Limited Veil of Ignorance .................................................................................................................. 18

The Limits of The Limited Veil of Ignorance ........................................................................................... 18

The Purpose of The Limited Veil of Ignorance ....................................................................................... 20

The Promises of The Limited Veil of Ignorance ..................................................................................... 20

Blind Orchestra Auditions ....................................................................................................................... 21

Conclusion ................................................................................................................................................ 23

Bibliography ............................................................................................................................................... 26
Introduction

Ever since Rawls published *The Theory of Justice* (1971/2009), the theory itself, along with many of its components, have been subject to intense debate. Both the “original position” and the “veil of ignorance” have been argued to be too unrealistic since the restriction on information makes it impossible to apply in the real world. Furthermore, it is argued that this restriction is too extensive, in that it excludes information that is necessary for making a just decision, and thus, that it contravenes fundamental egalitarian principles, making it less just than it should be. However, there still seems to be something appealing about the concept of the veil of ignorance and results from various experiments that have used it have shown results that indicate it has an effect, which implies that the concept might have some use after all.

In this paper, I will argue for a modified version of Rawls’ veil of ignorance that, while limited in its available applications, might both serve a practical use in society to increase justice and have a low risk of obtaining inequalitarian results. I will do this by first summarizing Rawls’ theory of justice with a focus on his veil of ignorance. Second, I will present and discuss the critiques from various philosophers. Third, I will discuss the implications of several experiments that make use of the veil of ignorance, including the issue that these experiments are ‘one-shot,’ or single iteration, experiments. Finally, I will detail my version of the veil of ignorance and explain why it both meets the critiques of the original veil of ignorance and could potentially serve a practical use in society.

Background

**Rawls’ Veil of Ignorance**

For context, I will begin by summarizing Rawls’ *A Theory of Justice* (1971/2009) and explaining the concept of his “original position” with a focus on the veil of ignorance. It is important to note though that the state Rawls describes is meant to be hypothetical, rather than historical like the natural state (Rawls, 1971/2009, p. 23).

Rawls’ *A Theory of Justice* Rawls utilizes the concept of a *veil of ignorance* (henceforth: VoI) in his construction of the *Original Position* (henceforth: OP). One of the major characteristics of
the OP is that no individual in this state knows their – or anyone else’s – status, capabilities, wealth or even gender in the real world, hence the Vol. The point of Rawls’ OP is to allow, or at least make it easier, for people to agree on the principles of justice, which in turn are meant to regulate how society should be built. The reason for such a strict restriction on knowledge is to prevent bias when deciding on principles. One of the first principles that Rawls’ argues would be chosen is the difference principle. The difference principle dictates that all unequal distributions of goods (such as wealth, services, or positions of power) are only permissible if they benefit the least well-off in society (Rawls, 1971/2009, pp. 99-100, 135-136).

In short, individuals in the OP are presumed to know only the necessary facts needed to select fair principles of justice for society. That is not to say that they know what society looks like, only that they have basic understandings of, for example, economic theory, human psychology, and politics. All information that would enable principles to be tailored to one’s own advantage is unknown: “Indeed, the parties are presumed to know whatever general facts affect the choice of the principles of justice. There are no limitations on general information” (Rawls, 1971/2009, p. 136).

While, for Rawls, the Vol is heavily connected to the OP, I think it can be regarded as a separate concept. As I have already explained, the point of Rawls’ Vol in the OP was to even out the differences between individuals to such a degree that the idea that “what is best for everyone else, is also what is best for me” becomes true. And, surely, from the OP, the best option for securing principles that work to individuals’ advantages is to secure principles that are to the advantage of everyone. Another way in which these principles might be considered just, which Rawls’ may have also intended, is that the decision-making process itself would be just since no one could have had an advantage over anyone else and would not be the result of any individual’s understanding of ‘the good’, since that kind of potentially biasing information is not available behind Rawls’ Vol. Instead, the principles would be based on what would be good for people in general. The principles would also have to be chosen without knowledge of one’s generation, i.e., individuals would not know if they would be born in the year 1960 or the year
2050, for instance; as such, conservation of resources, the environment, and capital accumulation are important factors when choosing principles (Rawls, 1971/2009, pp. 135-137).

There is much in Rawls’ theory that appeals to those who feel strongly about justice like me; but there are problems with what has been described above. For instance, the idea that “what is best for everyone, is also what is best for me” might be more idealistic than realistic. Of course, it is possible to imagine a state in which such strict information restrictions apply so that no other choice is left but to appeal to that ideal. However, is it not possible that despite the restrictions on available information, individuals could advocate for selfish principles simply because it might play in their favor? Indeed, it might not even be considered selfish since individuals cannot know if the chosen principles would play in their favor or not. My claim here is that, despite the restriction of information entailed by the Veil of Ignorance in the OP (which in itself, is difficult to imagine\(^1\)), there still seems to be room for some measure of risk-judgement. And, as such, it would still be possible for at least one of the individuals in the OP to take the risk to ‘vote’ for a principle that might play in their favor and disadvantage someone else. For example, it would be possible for an individual in the OP to vote for a principle that dictates that people in leadership positions should receive significantly more resources than their subordinates simply because that individual might end up in a leader position. However, this may be tangential to Rawls’ theory of justice, since it is the concept of the Veil of Ignorance and the principles decided in the OP that are important. In other words, the fact that it is not completely free from corruption does not necessarily mean that it is not ‘good enough.’ While we always should strive for the best of results and the ideal of justice, it is not realistic to expect, or require, that any theory or concept must be completely free from potential misuse to be viable. That it not to say that anything goes since, surely, at least some critique allows a theory to improve.

\(^1\) What I mean is that it is difficult to imagine removing all meaningful information to the degree that we do not know anything about who we are, since we must imagine these things – with necessity – from our own perspective. It is, thus, not possible to enter a totally unbiased state of mind, even hypothetically.
Michael Sandel’s Critique

I will now move on to summarizing the critique of Michael Sandel, which was part of why Rawls decided to revise his *Theory of Justice* and publish a new version that addresses some of these critiques.²

Sandel’s critique centers on the fact that the individuals in the OP seem to have less of a choice in the decision of principles than it appears. Sandel argues that while it may seem that there are many options in terms of which principles to choose, since individuals are said to have complete freedom once the circumstances of the OP have been made clear, the resulting principles are more or less pre-determined because the OP is designed to exclude individual differences. In other words, there is not much choice involved. Sandel also criticizes the assertion that the principles arrived at would be just, since it is not clear whether they would be just based on the procedure being just or because the procedure, by necessity, leads to a specific result that, in itself, is just (Sandel, 1982/1998, p. 131). As he puts it, “The distinction between pure and perfect procedural justice fades, and it becomes unclear whether the procedure ‘translates its fairness to the outcome,’ or whether the fairness of the procedure is given by the fact that it necessarily leads to the right result” (Sandel, 1982/1998, p. 131).

Regarding these circumstances – namely the VoI, Sandel argues that they present problems not only in how justice would be decided, but also on the meaning of ‘an individual’ itself. In other words, once the VoI has been applied with its absolute restriction on information, as Rawls suggests, the distinction between individuals becomes blurred because there is no information distinguishing them from one another. Sandel argues that because of the VoI they are not similarly situated, but identically situated (Sandel, 1982/1998, pp. 30-31). This relates to another objection with Rawls’ VoI, namely that it is too abstract to yield any significant results, since it restricts morally important information that could be deemed necessary to deciding on just principles (Sandel, 1982/1998, pp. 37-38).

---

² As mentioned earlier, this paper is not about this revised edition, but about Rawls’ original edition of *A Theory of Justice* (1971/2009).
Sandel makes a fair point about the issues with Rawls’ OP and VoI. If the ‘individuals’ in the OP are more akin to a single individual, the legitimacy of the principles comes into question, as they would not be the product of a collective agreement, but the decision of a single entity. Furthermore, the seemingly pre-determined choice process further questions this, as it leaves little room for discussion or considerations of different views. However, the concept of restricting biasing information to achieve more just and equal results is not necessarily thwarted by this. Indeed, the issues mentioned above are more or less the result of the VoI applied to the individuals in the OP, but when applied in a concrete context, such as in an experiment, the circumstances dictate what restrictions on information can be feasibly applied, and thus much of the abstraction is removed.

An Experiment about the Veil of Ignorance

In 2010, an experiment was conducted by Hannah Schildberg-Hörisch where they implemented Rawls’ veil of ignorance in a laboratory. The experiment was designed to employ several versions of a Dictator game in order to analyze the effects of risk and social preferences behind a veil of ignorance (more details about this in the next section). The results showed that subjects behind the VoI chose more equal distributions than in front of the VoI and only a minority of participants acted according to maximin principles—choosing the best worst outcome, or maximizing the results for those who will be worst off in the system. Rather than only acting for insurance purposes, many of these participants seemed to have chosen more equal distributions because of impartial social preferences towards equality. Finally, the results implied that “behind the veil of ignorance, maximin preferences are compatible with any degree of risk aversion if impartial social preferences for equality are sufficiently strong” (Schildberg-Hörisch, 2010, p. 1062). This final statement speaks against the utilitarianism argument that the difference principle is unrealistic in that it would only be optimal for infinitely risk averse individuals. I will explain this in more detail later.

---

3 The difference principle suggested by Rawls has the goal of maximizing the well-being of the least well-off by asserting that any unequal distribution of social or economic goods must be to the advantage of the least-
The *dictator game* is designed with a 50% efficiency loss at transfer, where two players, one recipient and one dictator, are to divide a pie of 12 pieces between themselves. The dictator makes the decision, and the recipient is passive in the process. In this experiment, three versions of the game were used. The first version, which I just described, has two players: one passive (the recipient) and one active (the dictator). The dictator decides how many pieces of the pie will be transferred to the recipient and the dictator keeps the rest. In the VoI version, the rules are the same, except none of the players know which role they will be assigned until after both have decided how many pieces to transfer. Consequently, none of the players know how many pieces they will end up with and are, therefore, unable to secure the largest share for themselves. This version also includes a risk, since everything transferred incurs a 50% efficiency loss, meaning only half of what is sent is received by the recipient. Thus, sending anything but 8 pieces, in accordance with the maximin principle, incurs a risk of the player ending up with less than the other player\(^4\). Lastly, in the risk version there is only one player, but they do not know which role they will have until after they make their decision. In this way, the player must decide between the maximin (8 pieces transferred) to ensure receiving a minimum of 4 pieces or take the risk of transferring any other amount to possibly get more pieces (Schildberg-Hörisch, 2010).

The utilitarian argument against Rawls’ difference principle states that it would only be optimal for infinitely risk averse individuals. This would seem to be true if the result would be similar in both the risk version of the game and the VoI version. However, the results differ significantly between the two versions and, therefore, not only does not support the utilitarian claim but speaks against it (Schildberg-Hörisch, 2010, p. 1063). As Schildberg-Hörisch explains, “Our results imply that the difference principle can also be derived from (sufficiently strong) advantaged members of society. It is similar to the maximin principle in that it tries to secure a more equal distribution of wealth but is different in that its goal goes beyond the simple maximization of the minimum allocation of wealth and goods for people.

\(^4\) For instance, transferring 2 pieces leaves 1 piece to the recipient, and 10 pieces to the dictator, and transferring 10 pieces leaves 5 for the recipient, and 2 for the dictator, hence 8 pieces transferred is the only way to ensure the player receives 4 pieces as the minimum.
impartial social preferences of equality and does not require infinite risk aversion.” (Schildberg-Hörisch, 2010, p. 1063)

Of course, this does not necessarily mean that the utilitarian argument is wrong, but it does present a significant challenge to their claim, even though the results only support Rawls’ difference principle in a limited way. In the Vol version, only 14% of the subjects acted according to maximin principles, i.e., transferred 8 pieces of the pie.

When comparing the results of the standard version and the Vol version, subjects transferred more than double the average in the Vol version than in the standard version: 3.5 compared to 1.6, while in the risk version the average was 3.1. As Schildberg-Hörisch stated: most of the subjects are clearly risk averse, as 68% of the subjects transferred a positive amount despite the large efficiency loss (Schildberg-Hörisch, 2010, p. 1064). Still, 35% of the subjects transferred more in the Vol version compared to the risk version and 44% transferred the same amount, which further indicates the tendency for risk aversion. The remaining 21% transferred a lesser amount in the Vol version compared to the risk version, showing an impartial preference to efficiency (Schildberg-Hörisch, 2010, p. 1064).

This experiment shows that people seem to choose more equal distributions behind a Vol. However, it is not clear what their motivations are. The subjects that chose to transfer more in the Vol version show a preference for impartial equal distributions over their own degree of risk aversion. Though, the average amount transferred is a long way from the maximin with only 14% acting accordingly. As such, the amount transferred, while seemingly in favor of more equal distributions, is not necessarily equal. With the maximin value being 8 transferred pieces, leaving 4 pieces for both, the average value transferred in the Vol version is only 3.5. Though it might be worth keeping in mind that the value includes those who decided to transfer less in the Vol version in favor of efficiency.

Whichever the case, it seems like the Vol leads to more equal distributions and, even though most participants did not act in accordance with the difference principle, the results were still the most equal in the Vol version. As for the utilitarianist claim that the difference principle
requires people to be infinitely risk averse, the fact that 35% chose to transfer more in the VoI version compared to the risk version, shows that an infinite risk aversion is not required for such a stance. Since the risks are very much the same in both the risk version and VoI version of the game, the difference in equal distributions cannot be attributed to risk aversion, rather, the fact that the difference is that other people were involved in the VoI version shows that 35% chose more equal distributions, not because of risk aversion but because other people were involved which, in turn, implies that the participants showed a sufficiently strong concern for equality (Schildberg-Hörisch, 2010, pp. 1064-1065).

Even if the VoI proved to be at least somewhat effective at making the results more equal, it might not have been as effective as many had hoped. This implies that, while a practically applied VoI does not have the effects that Rawls might have hoped, it could still be seen as a (small) victory. However, considering that this is a one-shot experiment, the results might not be as interesting as previously argued. People might act very differently in the real world, where multiple factors play key roles in decision-making. So, it is important to note that, although the results could be subject to some excitement for those hoping for a working VoI, it does not ultimately prove its validity. Furthermore, the VoI has other issues which need to be resolved. For instance, in addition to biasing information, does the VoI also exclude morally relevant information?

Is the Veil of Ignorance Really Compatible with Egalitarianism?

In *Egalitarianism Against the Veil of Ignorance*, John E. Roemer argues that VoI arguments, “when properly done, contravene fundamental egalitarian principles” (2002, p. 167). Roemer argues that the VoI is the wrong tool for deducing the requirements of distributive justice and that egalitarians must cease to use it as such. In his paper, Roemer discusses the concept of the

---

5 This is on my read of Rawls, considering how the intentions were to remove affecting biases to make things more equal through the VoI.
Vol from the perspectives of Harsanyi and Dworkin, but also touches upon Rawls’ version. Since my focus is on Rawls’ version, I will leave out much of his discussion on Dworkin.

On Rawls’ suggestion that the difference principle would be derived from behind the Vol, Roemer argues that the only sensible reason why a soul in the OP would make such a decision is if they were infinitely risk averse. However, as discussed earlier, Roemer’s claim is at the very least challenged by the fact that the difference principle did not seem to require infinite risk aversion in the experiment.

Roemer claims that despite the goal of the Vol, it does not ensure that the outcome becomes equal or just. Since it excludes all biasing information, it also excludes some vital information for making just decisions. For example, it excludes the experience of disabled individuals. Would it not be possible that the conclusion reached behind the Vol could be something like: since disabled individuals cannot contribute as much, they will get a lesser allocation of wealth compared to non-disabled people, but there will also be reduced expectations on them? This shows that even if the Vol forces impartiality, it does not necessarily lead to egalitarian conclusions (Roemer, 2002, p. 182).

However, I would argue that the Vol does not necessarily lead to the consequences Roemer implies – leading to individuals with a disability receiving less transferable resources. While the Vol cannot consider the experience of everyone, it can account for ability by having the decision-maker(s) not know if they will live with a disability or not. As such, it would be possible for a decision process, using the Vol, in which the decision-maker(s) would have to take ability into account. However, this does not guarantee that individuals with a disability will receive the same amount of transferable resources as individuals without a disability since even though the condition is accounted for, the experience is not. When deciding overarching principles as in the OP with a Vol, experience seems vital. Thus, I would argue that the problem is not with the Vol itself but with the situation in that it is applied. For instance, on the decision of whether

---

6 He mostly discusses Dworkin’s “thin” Vol and, since it is not closely related to Rawls’ concept, I have decided to exclude it.
individuals with a disability should get priority to jobs that are not affected by the disability, a VoI perspective could prove useful, as the decision-maker(s) would have to consider what they would want if they were to have a disability.

However, I agree with Roemer that there is a problem that needs to be resolved if the VoI is to be used in practice. The exclusion of morally relevant information is quite serious in terms of justice and equality since it makes the ideal of both justice and equality impossible to achieve. By excluding morally relevant information such as the experience of disabled people, it becomes impossible to take their experience into account and thus, the use of a VoI could never do justice to everyone that it excludes. Considering the aim of maximizing justice and equality, it would then make sense to conclude that using the VoI does not achieve this.

However, while I agree that the VoI should not be used in this case, or at least not Rawls’ version of it, I do think that the VoI can still be a viable tool for maximizing justice and equality. The main problem here, I think, is that the situation is altogether much too abstract for the VoI to be applied without similar issues to what Sandel criticized. The restriction on information also removes information vital for just principles, and if we were to have these principles as a foundational pillar for a just society, then they need to account for the experience of those less well-off, which the VoI seems unable to do. I would argue then that the VoI should not be applied for situations such as deciding the principles of society, these situations are too abstract and allow for too much information to be restricted.

To summarize: I agree with Roemer that the VoI seems insufficient as a tool for distributive justice (at least at such a scale), and so I agree with his main point: that the VoI leaps the risk of excluding vital and morally relevant information for just decisions. However, I argue that despite the VoI not guaranteeing that individuals with a disability will receive similar allocations of wealth, it can prove beneficial in decisions regarding compensation for a disability, such as priority to certain jobs. Lastly, while I agree that the VoI might not be useful when deciding upon allocation of transferable resources in the disability case, since it excludes information that is vital for a just decision, I still argue that the VoI can be a valuable tool in other situations.
I will go into more detail as to what situations specifically later in the essay. Next, I will present and discuss some of the feminists’ critique on Rawls’ Theory of Justice.

A Feminist Critique of Rawls’ Theory of Justice

In A Feminist Critique on Rawls Theory of Justice (1986), Mari J. Matsuda argues that abstracting theories from how they work makes them implausible in real life situations, or even in the real world in general. She claims further that this not only challenges the foundation on which Rawls theory rests, but also shows that there is not even a foundation in the theory (Matsuda, 1986, p. 621). Rather, the abstraction implies that the theory rests upon ideology and is not anchored to human experience or the lives it is meant to regulate. In combination with Conor Arsenault’s critique in A Feminist Critique on the Limits of Rawls (2016), this shows that the principles that Rawls suggests would not affect justice in the family, which would lead to continued injustice in the public sphere, most of all for women. One of the reasons for this is that Rawls’ assumptions do not take into account women’s history of being oppressed and facing injustice within the private sphere. This is linked to the notion of Rawls’ principles being too abstract and not anchored to reality. The result is a theory that might work theoretically but does not when applied practically.

The Problem of Abstraction

Since these principles, conditions, and limitations are created within a hypothetical space, not directly derived from reality, they would be difficult, if not impossible, to implement. While the OP is meant to be a tool for justice, it seems difficult to use it; and even if we could, it would provide a skewed result of justice because of it failing to take women’s experiences into account.

As such one could say that Rawls’ theory of justice, with its included OP and even the VoI, has failed as an instrument of justice; and, when abstraction fails, it seems intuitive to instead work towards concretion. Matsuda proposes that we should work in front of the veil rather than behind it when trying to create a better world. As Matsuda states it: “To argue at the level of abstraction proves nothing and clouds our vision. What we really need to do is to move forward
through Rawls' veil of ignorance, losing knowledge of existing abstractions. We need to return to concrete realities, to look at our world, rethink possibilities, and fight it out on this side of the veil, however indelicate that may be” (1986, p. 624)

What does this mean for the future of the Vol? The suggested answer would be to abandon it, leave behind the unnecessary abstraction and instead try to work from the reality of individuals to figure out ways to improve people’s lives. However, I think there is still value in a Vol applied to concrete situations. As stated earlier, in a concrete situation the information restrictions are limited to the circumstances of what can feasibly be restricted and should also lower the risk of vital information being restricted. Matsuda acknowledges that while abstraction is not ideal, we should not get rid of it altogether as it still has value:

This is not to suggest that theory and abstraction are without value. The suggestion made here is a more modest one. Theory has value, as long as we remember that real people create theory and that real people live their lives in worlds affected by theory. Half of those people are women, and their experiences can teach us something about justice. (1986, p. 629)

Clearly the Vol does not have (or at least does not seem to have) the desired results when used in large scale as Rawls does in his A Theory of Justice (1971/2009). My suggested solution would be to limit its use to small-scale contexts instead of trying to derive overarching principles of justice with it. This will be discussed further below.

**Missing the Roots of Injustice Towards Women**

Arsenault builds on Matsuda’s inquiry, arguing that because Rawls’ principles (including the Vol) exclude the private sphere, the result would not maximize justice for all (Arsenault, 2016, ss. 10-11). Despite the justice of principles in the public sphere, she continues, people are raised in the private sphere where injustice is likely to initially develop because of tradition and a history of women being subject to injustice in many different forms (which are not so easily removed in practice). The OP and the Vol are not enough to ensure a just foundation on which to build justice in practice, because we simply cannot remove all bias, values and such, and
some people will still see women as subordinate. Even if we were to remove the theoretical possibility of an unjust and biased basis for the judgement of women, in practice it would persist since it is impossible to regulate all aspects of life with overarching principles. A CEO of a company, for instance, might be regulated from making discriminatory decisions in hiring people based on the company’s principles, but there is still no way to regulate the CEO’s behavior or (unconscious) biases. While job candidates might be protected by the law, in practice they would not be protected from unjust or demeaning behavior from people who have values or views that enable such behavior as normal. In short, this means that the limits on Rawls’ principles makes them ineffective at promoting actual justice and equality for women.

The fact that the Vol overlooks the roots of injustice towards women makes it ineffective to prevent it. It is simply very difficult to imagine what it is like to be a woman for a man, or for anyone at all to imagine being someone whose experiences, based on who they are, differ from theirs.

If there was no such thing as a status quo or if people’s minds could spontaneously change for the better after years of thinking in a certain way, it could be argued that the original position being historically neutral could be justified. It seems, though, that people cannot simply forget about history, and that change, especially attitudinal change, normally takes a lot of time. That being said, although it makes sense that Rawls’ original position is historically neutral because by definition it is indifferent to empirical circumstances, the reality of injustices towards women is far too apparent for the original position to be historically neutral and therefore neglect the fact that many people still view women as subordinate. Choosing principles of justice from the original position, then, although great for “public justice,” does nothing to address the reality of “private anarchy,” that is, wherein there is no law
and the naturally stronger will rule, and the naturally stronger, as was previously argued, is usually men. (Arsenault, 2016, s. 10)\(^7\)

For the VoI this means that it would likely not be effective at increasing justice for women in practice when used on such a large scale, as we simply cannot remove all bias through a hypothetical VoI that removes all relevant individual knowledge. But, as previously mentioned, there might still be some use for the VoI. Before I address that, I will first discuss another Vol-based experiment.

**The Veil of Ignorance Favors the Greater Good**

In *Veil-of-Ignorance Reasoning Favors the Greater Good* (2019), Karen Huang, Joshua D. Greene and Max Bazerman make the claim that VoI reasoning promotes the greater good. They came to this conclusion through the results of seven experiments they conducted in which the participants were faced with various ethical dilemmas. One example is the classic Trolley-Footbridge case where the decision-maker (the participant) must decide whether to throw a bystander over the railing of a footbridge to stop an incoming trolley heading for five people tied to the track, or not throw over the bystander and leave the five to their death. In all the experiments there was a VoI condition and a control condition. In the VoI condition, the experiment’s first stage uses a VoI version of the dilemma in which the participants must make the decision under the premise that they do not know which of the people involved they will end up being (the bystander, or one of the five), and in the second stage the experiment uses the standard version of the dilemma in which the participants play the role of decision-maker, and will not be directly affected like the bystander or the five people on the track. This is done to measure the difference in participants who made a more utilitarian decision after the VoI condition, compared to the participants in the control condition who only answered to the standard version of the dilemma. Afterwards, the participants were asked to rate how morally acceptable their decision was. In all experiments both the proportion of those favoring the utilitarian response and the moral acceptability of that response, were the highest in the VoI condition.

---

condition of the experiment (Huang, Greene, & Bazerman, 2019). Hence the implied conclusion is that the Vol promotes the greater good.

One of the possible objections about the conclusion of this outcome, is the explanation that the participants were simply anchoring their previous utilitarian choices and, thus, that it is not actually a matter of moral reasoning. However, Huang, Greene and Bazerman hypothesize that “participants are engaging in a specific kind of moral reasoning, mirroring the reasoning of Rawls and Harsanyi, whereby participants perceive a connection between what is morally defensible and what they would want if they did not know whom they would be among those affected by the decision” (Huang, Greene, & Bazerman, 2019, p. 4).

I will not go into detail about all these experiments, but instead focus on the first one regarding the classic trolley-footbridge case. As stated already, the conclusion reached in that study does not necessarily mean that the effects of Vol reasoning is desirable, only that it seems to promote utilitarian decision-making and can be used as an argument in situations where decisions involving difficult trade-offs need to be justified (Huang, Greene, & Bazerman, 2019, p. 6). Though, the question of whether that is enough to make the decision right remains debatable. While it might seem very plausible and rational from a perspective of someone trying to act “for the greater good” or from a utilitarian perspective, a feminist might add that taking this perspective means neglecting some very important factors of reality. Namely, who would the people involved be? Would the participants still push if it were their brother or sister on the footbridge? Perhaps they would not even hesitate if it were their whole family on the tracks. These details are immensely important in real life decision-making and, thus, a Vol perspective is not always favorable, nor will it always aid in decision-making. That is not to say that it has no value. At the very least, a Vol perspective allows for more impartial reasoning and can be useful in situations where we want to avoid bias, like when selecting a candidate for a job, or when distributing resources.

It is important to note though that the results of these experiments and, therefore, their implications, are derived from isolated, controlled situations; as such they do not necessarily have a direct connection to practical applications. Though, this is a problem with experiments in
general. Nevertheless, I would argue that the results and their implications (namely, that a VoI perspective usually leads to more utilitarian/egalitarian decisions) could be useful for practical applications. But, as I pointed out with respect to how a feminist might respond to the trolley-case, we need to be mindful of how we apply the VoI and when. The natural question then is: when is the VoI applicable? My short answer to that question is that it is determined by what type of situation and the scale of the decision. As previously mentioned, abstraction, overarching principles and large-scale decisions seems problematic for the VoI, but in concrete situations, such as the experiments mentioned, the VoI works as expected – decreasing bias in decision-making. I will discuss this topic further below.

The Veil of Ignorance in Clinical Research

In The Ethics of Clinical Research, David Wendler explores the concept of the VoI in clinical research to establish an obligation to participate in non-beneficial research. In order to do this, he modifies Rawls’ VoI to incorporate the lack of knowledge on generational status to remove the bias that might come from knowing if you are alive at the time in which the decision is being made and, therefore whether the individual would have already reaped the benefits that come from clinical research (Wendler, 2017).

Wendler argues that behind this version of the VoI, it would seem reasonable for people to participate in clinical research, even non-beneficial research, if the overall benefits outweigh the burdens (2017). However, there are several problems with this approach. Even if it would seem reasonable, and especially if we try to adhere to the maximin principle, the ethical implications are not always desirable and might very well go against our current moral understanding. As Wendler puts it: “almost all research ethicists would regard as unethical a study which intentionally infects a few subjects with the HIV virus, even when the study offers the potential to identify a cure for AIDS” (Wendler, 2017, ss. 6. Contract Theory, paragraph 10).

The issue here is that individuals behind a VoI, who are only concerned about their own self-interest, might in this situation allow the infecting of study participants to potentially save millions in the future. The dilemma here is that there are options of either not going forward with the procedure, resulting in many more deaths over time, or proceeding with it
and intentionally infecting a few to potentially save millions over time. With a Vol perspective like the one in the classic trolley case, one can clearly see that the implied “correct answer” would be to infect a few to save millions. Yet, in our reality, this would be deemed highly immoral.

The conclusion derived from this is very much the same as the one in “Is the Veil of Ignorance Really Compatible with Egalitarianism?”, namely that the Vol also has the consequence of limiting vital information, and that decisions made with such reasoning does not guarantee that they will be in line with what we deem just or moral. Since I have already discussed this, I will not do so now, and instead elaborate on it in below.

My Solution

I have now addressed the key critiques of Rawls’ OP and Vol in addition to discussing the Vol as a concept. In summary, it seems that the Vol has some vital flaws that either need to be fixed or circumvented somehow to give the concept some practical use, such that the benefits outweigh the consequences. As such, I will now begin to explain my suggestion for how the Vol can be modified to enable practical use, which I argue either corrects or circumvents most of the problems, though also with some strict limitations and likely with some problems of its own.

The Limited Veil of Ignorance

The Limits of The Limited Veil of Ignorance

First, the limits of this new Vol will be restricted to only concretely applicable situations, roughly conceived. What I mean by that is that the situation at hand must be concrete enough for the Vol to manifest itself through actual change, meaning through a change of procedure. In a job application, this could mean that names are replaced with a number instead, and any information that can be deemed as biasing is removed for the one who reviews it. In contrast, the Vol is not meant to be applied to elements that are more theoretical than practical, like principles on how to treat people. We can quite easily see, even in a hypothetical situation, that
applying the Vol in such a situation could result in principles like: “treat everyone the same” or similar. And we have already concluded that one of the critical faults of the Vol is that it not only excludes irrelevant information for equality and justice, but also information that is relevant for equality and justice. The reason I say that this new limited Vol is roughly restricted to practically applicable situations, is because I can imagine that there could be situations where it is applicable without any grave consequences, and where the situation is more theoretical than practical.

The reason I propose strict restrictions is to limit the harm it can do. I do not mean though that the Vol is necessarily harmful, only that the results of its use can sometimes lead to undesirable results as we have seen in earlier sections. The thought is that if we restrict its possible applications, we avoid the problem of the Vol excluding necessary information in questions about morality. The problem is not completely avoided though, of course, as it is still possible to exclude relevant information for just decision-making, however, that might also depend on how it is used, rather than what it is. When the limited Vol is used properly, it will not be used for cases that decide the ruling principles of humanity, rather for situations like auditions, job applications, or any situation where the primary focus is equality in a decision.

This limited Vol does not limit what kind of information the Vol can restrict but adjusts to the circumstances of the situation, meaning it can only restrict information that can practically be restricted. For instance, in a job application biasing information can be things such as name, age, place of residence, gender or looks. Applying the limited Vol on that kind of situation would mean removing access to that information and only keeping the information that is relevant to the job itself. As in Rawls’ Vol, any information necessary to make a decision is not affected, but the risk of excluding morally relevant information is circumvented by being irrelevant for the situation. If morally relevant information, to the situation, were to be excluded, then the Vol should not be applied to that situation, or if possible, the restrictions adjusted to allow such information while still limiting access to biasing information.

To clarify and summarize the limits of this new limited Vol: the limits are that its use is restricted to situations in which it can manifest through concrete change or change of
procedure. Note also that this restriction is more like a blurry line rather than a strict one, as there might applications that fall outside said restrictions on application. Also, note that there is no limit on what kind of information can be restricted, but that it adjusts to the circumstances of the situation. In general, this means that it removes information that might lead to a decision based on irrelevant, untrue assumptions or unwanted, biased information.

The Purpose of The Limited Veil of Ignorance

As already mentioned, this new VoI might still have the problem of excluding information that is relevant for justice. Because of this, the new VoI does not ensure justice. But, if used correctly, it should ensure increased equality. Another reason why it does not ensure justice, and why achieving justice is not part of its purpose, is because justice can be very different for different people, and for different situations. As such, it would be difficult to tailor a VoI, or perhaps any principle at all for that matter, to encompass every situation in a way that makes the final outcome just. However, it should ensure increased levels of equality by removing unwanted bias in decision-making. And by equality I mean the common-ground basis on which all human beings should be judged. But as I said, the limited VoI does not ensure justice also because while the results may be based upon equal terms, it cannot differentiate just from unjust. It is the flaw of a rigid system that is left for the decision-maker, and the people overall, to address.

The Promises of The Limited Veil of Ignorance

In short, the new limited VoI promises limited results, but it also results in fewer problems. The main issue is that it does not ensure justice, as it cannot differentiate relevant from irrelevant information, and the solution of “leaving it to the people” might not be ideal, as it may lead to it being misused and lead to unjust or unequal procedures. However, I would argue that most tools can be misused. Hammers can be used both to build cities, but also to cause harm. Even abstract tools such as the grading system are subject to the risk of misuse. Should the tool not be used just because there is a risk of misuse? I think it should, and that measures can be taken to limit the risk of misuse, such as institutionalizing or regulating its usage. It is important to note that this new limited VoI is only meant to be a tool for justice and equality, with justice
being the desired outcome through the equality it is meant to promote. Since, as mentioned earlier, it cannot differentiate just from unjust, it is important to be mindful of when it should be used and whether the result brings a desired outcome or whether something else needs to be done to obtain the desired result. This will be made clearer in the discussion about “Blind Orchestra Auditions” below.

When the limited Vol is used properly, it might not change the world, but if it would work as I hope, it would at least aid in making life for people just a bit more just and aid in providing more equal terms. Because, by limiting the amount of biasing information, decisions about who gets a job, who is accepted to study, and who gets a promotion will be based on relevant skills rather than heritage, contacts, or other biasing assumptions. While I have already mentioned that this new limited Vol does not ensure justice, but rather equality, it is my hope that by ensuring more equality, more justice will follow.

**Blind Orchestra Auditions**

Blind auditions began to be implemented around the 1950s, which meant that auditions for orchestras used a screen to conceal the identity and gender of the musician to the jury toward the aim of preventing biased decisions. It was not until the 1970s and 1980s that it saw more widespread use and more orchestras used the method in recruiting musicians. Between the 1970s, when blind auditions started to get more widespread, and 1993, the percentage of women in the highest-ranked orchestras increased from 6% to 21% (Goldin & Rouse, 2000). According to a study conducted by Cecilia Rouse and Claudia Goldin, a large part of that increase can be attributed to the blind auditions decreasing gender-biased hiring and increasing the likelihood of women being selected by about 30% (Goldin & Rouse, 2000).

The interpretations and results of this study are disputed though, and some argue that the implementation of blind auditions might not have had as big an impact as first thought. Jonatan Pallesen claims that the overall results of the study are not significant enough to claim that blind auditions are the primary reason for the large increase of women within orchestras and, further, that blind auditions might have even been disadvantageous for women in some cases. (Pallesen, 2019).
I agree that the study seems to have several issues in terms of interpretation of the results, as some numbers seem to suggest a decrease in number of women being admitted from blind auditions compared to non-blind ones (Claudia & Cecilia, 2000, s. 13 (727)). But, according to the authors, a possible explanation for this is that, as blind auditions started to get more widespread, it made more women audition overall and, as a result, lowered the average skill of female auditioners (Claudia & Cecilia, 2000, s. 12 (726)). That interpretation though seems to suggest that the non-blind auditions were actually advantageous for women in some cases.

However, the main point is not whether blind auditions achieved what the equality intended, or whether they were as effective as some suggest. The number of women in orchestras did, in fact, increase significantly with the implementation of blind auditions. Thus, even if blind auditions were not always as advantageous to women as expected, they had enough of an effect to increase the number of women being admitted and decrease gender-biased admission for both women and men. My claim here is that blind auditions are, thus, a good example of how a limited VeI can be implemented. The situation is concrete enough to implement changes to a procedure that could limit the risk of biased decision-making, and implementing the change seems to have very little risk of leading to more unequal or unjust results than without it. What is clear is that implementing any changes of this kind requires being mindful of the precise aims we wish to achieve.

Anthony Tammasini of The New York Times suggests that we remove blind auditions if we want to increase diversity. According to him, the changes do not occur quickly enough to reflect the changes in society (Tommasini, 2021). He has a point, given that blind auditions do not necessarily lead to diverse orchestras that represent the actual cultures and peoples in society; because, even if blind auditions prevent bias, they do not compensate for already unequal chances of admission caused by economic differences, access to music training, and so on. Furthermore, he suggests that as the standard of skill among musicians has increased through the years, it often results in many auditioners being more or less equally good, but if the auditioners are more likely to be of a specific ethnicity rather than another, then it is that group that will be more statistically reflected in the final result of people admitted (Tommasini, 2021).
The question is, should we remove blind auditions to knowingly make biased decisions that would increase diversity? It is in dilemmas like these that equality and justice clash, since it would be unjust towards some people to remove the protection from bias but, at the same time, it might be more just towards other people to compensate for their unequal chances of being admitted as a consequence of the circumstances of their economic class, making it more difficult for them to even get to the audition. In the end, this is something that is not easily solved, considering both sides of the dilemma have a point, and using one solution that increases equality in some aspects, might decrease it in some other aspects. However, a limited Vol in a situation like this does what it is supposed to do—decreases bias—but the question remains whether the solution it provides is a desirable one. Nonetheless, I think the implementation of a limited Vol in more situations could prove beneficial to decrease bias in areas where decision-making needs equal grounds. But it does not solve all the problems of equality and justice.

Conclusion

I have discussed and analyzed some of the critiques towards Rawls’ veil of ignorance in addition to discussing the Vol on its own. Overall, Rawls’ version of the Vol seems to have many issues that makes it unlikely to be used in practical settings and ineffective in achieving what it intends. My suggestion of a limited Vol seems to avoid most of these problems:

1. The problem that the OP and Rawls’ Vol are too abstract and difficult to implement in the real world is circumvented by the fact that the limited Vol adjusts to the circumstances of what information can feasibly be restricted to limit the risk of bias. In addition, the limited Vol requires that the situation is concrete enough to even allow for practical application. Regarding how to implement it, the guidelines I have provided should be sufficient to implement it in applicable situations. To summarize the guidelines:
   a. Restrict information that can be deemed as biasing in the decision-process.
b. If morally relevant information, or other vital information risk being restricted, then the restriction on information should be adjusted to allow for the morally or vital information to be accessed, while still restricting biasing information if possible. If not, then the limited Vol should not be applied to the situation.

c. The application of the limited Vol is limited to practically applicable situations in which a change of procedure is possible. It is not to be used in determining overarching principles of equality or justice.

2. The problem that Rawls’ Vol results in inegalitarian outcomes is somewhat circumvented by the limited Vol having less influence overall, but also by the conditions stated above. If used according to the above guidelines, there should be little risk of inegalitarian results because (1) the information would be regarded as morally relevant, and (2) it is not to be used for overarching principles which leaps the risk of achieving inegalitarian results. The main point of the limited Vol is to only allow information which is relevant for the decision-process. The problem that it cannot differentiate between just and unjust is because of it being a tool for increased equality, with the desired outcome being increased justice, but it cannot ensure justice because of the complexity of what justice is, as demonstrated the section about “Blind Orchestra Auditions”. Considering the results of the experiments and cases mentioned in this paper where a Vol was implemented, the limited Vol seems unlikely to cause more injustice. But the conditions of the limited Vol itself cannot see the difference between unjust and just. That is, with necessity to the circumstances needed for the limited Vol to be applicable in a correct manner (as described above), the matter of what outcome means increased justice is left to the decision-maker(s) and the public. Although, when used correctly it should either be avoided entirely or irrelevant for the tool itself.

3. The feminist critique that involved Rawls not taking into account the history of women’s oppression is something that the limited Vol could correct by creating more equal grounds on which decisions are based.
Overall, the limited VoI should be able to provide more equal grounds for decision-making in several situations in society. It is not the solution in inequality nor injustice, but in theory it should increase equality, and in turn justice. The experiments mentioned in this essay indicate that the VoI have the effect of increasing egalitarian results, though these are one-shot experiments and just because it works in an experiment, does not mean it will when applied in society. However, these indicated results combined with the limited VoI, though abstract in its construction, could potentially serve a practical use as a tool to increase equality by decreasing bias in decision-making. And, I argue, would be worth continued testing to see if it would have the intended effect.
Bibliography


