Croatian Journal of Philosophy
Vol. XXIII, No. 68, 2023
https://doi.org/10.52685/cjp.23.68.5
Received: November 24, 2022
Accepted: July 14, 2023

Unwanted Arbitrariness

STIJN BRUERS
KU Leuven, Leuven, Belgium

I propose a new fundamental principle in ethics: everyone who makes a choice has to avoid unwanted arbitrariness as much as possible. Unwanted arbitrariness is defined as making a choice without following a rule, whereby the consequences of that choice cannot be consistently wanted by at least one person. Other formulations of this anti-arbitrariness principle are given and compared with very similar contractualist principles formulated by Kant, Rawls, Scanlon and Parfit. The structure of arbitrariness allows us to find ways to avoid unwanted arbitrariness. The two most important implications of the anti-arbitrariness principle are discussed: non-dictatorship and non-discrimination.

Keywords: Ethical principles; Kantianism; contractualism; categorical imperative; dictatorship; discrimination.

1. *Introduction*

From Kant (1785) and Bentham (1789) to Scanlon (1998) and Parfit (2011), philosophers have a long tradition of searching for the most fundamental ethical principles. This article fits in that tradition, by defining a new core concept in ethics: unwanted arbitrariness. With this concept, the anti-arbitrariness principle states that everyone who makes a choice has to avoid unwanted arbitrariness as much as possible. This is a new proposal of a fundamental principle in ethics. It is fundamental in three senses. First, the principle offers a necessary (but not necessarily sufficient) condition for an act to be right or a choice to be moral. In other words, a violation of the anti-arbitrariness principle is a sufficient condition for an act to be wrong or immoral. Second, the principle applies to all choices, including for example the choices of moral rules and moral theories. Hence, it is a meta-principle: a principle about principles. Third, unwanted arbitrariness refers to a lack of moral justifications, valid reasons or acceptable rules. When such reasons or rules are absent, we basically leave the realm of morality. The concept of unwanted arbitrariness is so crucial, that it can be said to demarcate morality, to distinguish the moral and immoral from the amoral. This reason-based or rule-based approach to ethics fits in Kantian and contractualist traditions of ethics (Kant 1785; Scanlon 1998; Rawls 2005; Parfit 2011). As such fundamental principles do not tell you what to do or what is moral (i.e. do not make substantive moral claims or judgments), but rather give you a procedure or method to determine what to do, what is right or what is good, such fundamental principles are useful in the field of procedural or formal justice. Instead of offering substantive claims how to solve each case, the principle entails, for example, that one should treat like cases alike. Instead of offering which specific rights a person has, the principle imposes the condition that everyone should have equal rights.

The search for the fundamental principle(s) in ethics is a very ambitious project. It requires giving precise formulations, offering justifications, discussing implications, presenting applications and making comparisons with other proposals of fundamental principles in the moral philosophy literature. That requires a whole book. The main focus of this article is the first step: formulating the anti-arbitrariness principle as precise and unambiguous as possible. The justifications, implications, applications and relations with the existing literature will be briefly sketched, but are mainly left for future work. That means those issues are not yet fully developed in this article. Similarly, whether this anti-arbitrariness principle is a mere reinterpretation or reformulation of Kantian and contractualist theories, or contains substantial differences with such theories, will also be left for future research. Even if it is a mere reformulation of a fundamental ethical principle already proposed in the literature, it could help clarify that proposed principle and more clearly enable us to see certain implications of it.

In this article, I will define the concepts of unwantedness and arbitrariness, give several formulations of the anti-arbitrariness principle, briefly compare them with very similar fundamental ethical principles formulated by Kant (1785), Scanlon (1998), Rawls (2005) and Parfit (2011), use the structure of arbitrariness to look for options how to avoid unwanted arbitrariness, and discuss the two most important consequences of the anti-arbitrariness principle: non-dictatorship and non-discrimination. The unwantedness of a violation of the anti-arbitrariness principle gives us a reason why dictatorship and discrimination are morally wrong.

2. *Definitions of unwantedness and arbitrariness*

Unwantedness for an individual means being incompatible with that individual's largest consistent set of strongest subjective preferences. An individual is a being who has subjective preferences. A subjective preference is a conscious value judgment or evaluation that has a subjective strength (to be distinguished from, e.g. a mere unconscious behavioral disposition). For example, being told a lie is incompatible with a preference for knowing the truth. Two preferences are mutually inconsistent when it is unfeasible or logically impossible to satisfy them both. Consider a reluctant drug addict, who is torn between two preferences: wanting the drugs and wanting to be clean or healthy.[1] This inconsistency in preferences is what makes such drug addiction problematic. When the drug addict values being clean more than the excitement from taking an extra dose of drugs, but still takes the drugs due to the addiction, that behavior can be said to be irrational.

To respect autonomy, an individual can freely choose the method to construct their own strongest consistent set of preferences. One method to construct your individual consistent set is: consider the list of everything you want, ranked according to personal value or strength, with the strongest preferences at the top. Move down the list and delete the items on the list that are incompatible with the higher non-deleted items. The remaining items form a consistent set of your strongest preferences. Everything that is not logically compatible with this remaining list of your strongest preferences is unwanted and cannot be consistently wanted by you. Everything that is compatible can be consistently wanted. Saying that you cannot consistently want something can be interpreted as being equivalent to saying that you can reasonably object against it.

Arbitrariness means selecting an element (or subset) of a set without using a selection rule. A selection rule is a rule that logically determines the selection. It is an if-then statement that consists of a set of conditions with logical operators (conjunctions, disjunctions, negations). For example: "If element X has conditions A and B or not C, then select X." If the question "Why selecting element X instead of element Y?" has no answer that refers to a selection rule (for example if the only answer is "Therefore!"), then selecting X is arbitrary.

Combining the above definitions of unwantedness and arbitrariness, we can define unwanted arbitrariness as making a choice without following a selection rule, whereby the consequences of that choice are unwanted (i.e. cannot be consistently wanted) by at least one person. Here, a choice can be defined as a conscious decision. Making a choice means consciously selecting an element from a choice set, the set of eligible options. These eligible options can be feasible actions but also for example preferences, allocations, ideas, moral theories or ethical principles.

[1] I thank an anonymous reviewer for this example.

3. *Formulations of the anti-arbitrariness principle*

The anti-arbitrariness principle states that:

When making a choice, we have to avoid unwanted arbitrariness as much as possible.

To avoid arbitrary exclusion of choices, this principle applies to all possible choices, including very specific actions ("Sit at seat 5 on bus 42 at 1 pm Friday"), to more general choices ("Use public transport"), to justifications ("Take a seat because the seat is empty and you paid for a ticket"), to higher level moral choices ("Choose the action allowed by a contractualist ethic"), to moral theories ("Choose the theory of act-utilitarianism"), to even very basic choices of premises and logical deduction rules used in justifications ("Use deontic logic to determine the validity of an argument"). For practical reasons, we do not have to consider impossible choices ("Avoid unavoidable unwanted arbitrariness").

This anti-arbitrariness principle does not yet say what happens if we don't avoid unwanted arbitrariness. Also, the "as much as possible" hints at the possibility that sometimes unwanted arbitrariness may not be avoidable. Therefore, we can give a more exact formulation of the anti-arbitrariness principle, in a strong and a weak version.

Anti-arbitrariness principle, universal formulation, strong version:

> If you do not avoid avoidable unwanted arbitrariness when making a choice, you are not allowed to make that choice.

The weak version can be derived from this strong version. Suppose unwanted arbitrariness is unavoidable. You have to make a choice that involves unwanted arbitrariness. What about other people making other choices? Are you allowed to determine the choices of others, to impose your choice on them? Are you allowed to choose who may make the choice? Choosing yourself as the dictator who dictates the choices of others, would involve unwanted arbitrariness again. To avoid this new unwanted arbitrariness, you are not allowed to be the dictator. You have to accept the choices made by other people.

Anti-arbitrariness principle, universal formulation, weak version:

> If you cannot avoid unwanted arbitrariness when making a choice, you are allowed to make that choice but other people may make other choices from the same choice set (i.e. you have to tolerate that other people make other choices).

The above formulations are universal, in the sense that everyone and everything must abide by this principle. No arbitrary exceptions are allowed. The principle applies to everyone and everything that is able to make choices based on selection rules. It also applies, for example, to artificial intelligent machines. Of course, when someone cannot make a choice, that is an exception, but not an arbitrary exception because it is justified using an "ought implies can" rule: "If you cannot do something, you have no obligation to do it."

We can give another, personal formulation of the anti-arbitrariness principle:

> For every choice you make, you have to be able to give a justification rule such that you and everyone can consistently want that everyone follows that rule in all possible (including hypothetical) situations (i.e. you and everyone can accept the consequences of a universal compliance by everyone of the justification rule).

This is a personal formulation, because it refers to what you can want. Hence, this formulation applies to everyone who is not only able to make choices, but also able to want something, i.e. someone with personal preferences.

Whereas the first, universal formulation referred to selection rules, this second, personal formulation refers to justification rules. A justification rule is a selection rule that is used in moral reasoning, to justify to other people one's choices. Therefore, a justification rule for (im)permissibility of a choice should be used in a logical deduction. That means a justification rule is basically an if-then statement that consists of a set of conditions: "If conditions C apply, then it is permissible to choose X."

The above second formulation does not yet say what to do when you are not able to formulate a justification rule. Therefore, as with the first, universal formulation, we have to make this second, personal formulation of the anti-arbitrariness principle more precise. And as with the universal formulation, this personal formulation also comes in two versions, of which the weak one can be derived from the strong version.

Anti-arbitrariness principle, personal formulation, strong version:

> If, when making a choice, you cannot give a justification rule of which you would accept universal compliance, then you are not allowed to make that choice nor follow that rule.

Anti-arbitrariness principle, personal formulation, weak version:

> If, when making a choice, you cannot give a justification rule of which everyone would accept universal compliance, then you must accept or tolerate that other people make other choices from the same choice set and follow other justification rules for making those choices.

There are many similarities between the universal and personal formulations of the anti-arbitrariness principle, such that they can be said to be roughly equivalent.

First, there is a correspondence between the selection rule and the justification rule. The universal formulation works with a selection rule to avoid arbitrariness. In the personal formulation, arbitrariness is avoided by the justification rule and by the idea that if you may follow that rule in a specific situation, then everyone may follow that rule in all possible situations. Suppose that the "everyone" and "all possible situations" were no requirements. Replacing them by "some people" and "some situations" would introduce arbitrariness, because arbitrary subsets of the sets of all people and all situations can be chosen.

Second, both formulations look for what can be consistently wanted. The condition "everyone can consistently want that everyone follows that rule in all possible situations" is the opposite of unwanted arbitrariness. Suppose you choose option A arbitrarily and person Y is in a position P in which s/he cannot consistently want that arbitrary choice. If we consider everyone and all possible situations, this includes the situation where person Y chooses A and you are in the same position P that Y had, in which case you cannot consistently want A.

A third similarity between the two formulations, is that they both come in a weak and a strong version. Unwanted arbitrariness may not always be avoidable, because there may always be someone who cannot consistently want a choice that cannot be based on a selection rule. Consider for example the choice of moral theory. There are many, equally consistent theories. Choosing one theory, such as act-utilitarianism, would be arbitrary. And some people may not like that theory. Similarly, it may not be possible to find a justification rule of which everyone can accept universal compliance. The condition that everyone follows the rule in all hypothetical situations, may be too demanding. In these cases, people must tolerate that other people make other choices, for example choose another consistent moral theory (unless for example the act-utilitarians can argue that their chosen theory is not arbitrarily chosen, but chosen by a selection rule).

A final similarity is that both formulations apply to all possible choices, including the choice of selection and justification rules (in particular the choice of conditions in those rules). That means a selection meta-rule should be given to select the selection rule from the set of all selection rules. Similarly, a justification meta-rule should be given to that justifies the chosen conditions in a justification rule. With the application to all possible choices and the resulting necessary inclusion of such meta-rules (and higher order meta-meta-rules), the anti-arbitrariness principle becomes perhaps the most fundamental principle in ethics.

An example might give some clarification. Consider the situation of taking a seat on the bus. If you choose to take a seat, the rule could be: "If you are white, you may take the seat," or "If you have permission by person X, you may take the seat." But the choice of these conditions is arbitrary (they refer to one skin color or person arbitrarily chosen from the sets of skin colors and people). A better rule would be: "If the seat is empty and you have permission by the people who have a special relationship with the seat, you may take the seat." We have to specify what counts as a special relationship. This can again be done by considering relationships of which everyone can consistently want that they are part of the conditions in the justification rule. Examples of such a special relationship could be "being the owner of the bus" or "having reserved the seat". Having permission could mean "having paid for a ticket" (or generally: "abiding by a system of property rights that does not privilege one person over others").[2]

4. *Connections with other fundamental ethical principles*

The anti-arbitrariness principle is related to other fundamental ethical principles proposed by, e.g. Kant (1785), Scanlon (1998), Rawls (2005) and Parfit (2011). These principles are fundamental, in the sense that they are meta-principles that refer to ethical principles or rules to guide our actions. This section briefly compares the anti-arbitrariness principle with some other proposed principles. Whether the anti-arbitrariness principle is a mere reformulation or contains substantial differences with the other proposals in the literature, is left for future research.

Kant's first formulation of his famous categorical imperative (unconditional obligation), reads: "Act only according to that maxim whereby you can, at the same time, will that it should become a universal law" (Kant 1785). A maxim is a subjective principle of action, i.e. what the agent believes to be the reason for his or her action. A maxim consists of the act (e.g. "lying") and the motivation (e.g. "for a benefit"). When you do an action, find your maxim and imagine a world where everyone (who is able and is in a similar position as you are) follows that maxim. Only if everyone can follow that maxim without contradictions and you can rationally will that everyone follows that maxim, you are allowed to do that action.

This universalizability formulation of the categorical imperative implies for example that when making an action, you cannot make an exception for yourself. You cannot say that you are the only one who may follow your maxim. A universal law does not allow for arbitrary exceptions. This reflects an avoidance of unwanted arbitrariness. We end up with the anti-arbitrariness principle if an act is interpreted more generally as a choice (such that the choice for inaction or allowing something to happen are also considered), a maxim is interpreted as a justification rule and "rationally willing a universal law" is interpreted as "consistently wanting or accepting a universal compliance of the justification rule".

One important difference between Kant's principle and the anti-arbitrariness principle, is that Kant in a sense only considers the most general maxims. Kant claimed that lying is always wrong because a contradiction or irrationality occurs when everyone lies for a benefit. Although Kant not explicitly derived his position of impermissibility of lying from his categorical imperative, such a derivation is only possible by considering only a general maxim such as "lying for a benefit" (for a similar discussion of this point, see e.g. Carson 2010). The anti-

[2] I thank an anonymous reviewer for this point.

arbitrariness principle, in contrast, considers more maxims or justification rules. As a consequence this anti-arbitrariness principle allows for lying in some situations, for example in order to save a life (e.g. when a murder asks you the hiding place of his target victim). I can consistently want that everyone follows the justification rule "if the lie saves the life of an innocent person and has no serious negative side-effects, then you may lie." Kant, if he were to derive his anti-lying conclusion, would only consider a justification rule "if the lie has a benefit, then you may lie", and I cannot consistently want universal compliance of this rule.

Scanlon formulated a contractualist principle of wrongness: "An act is wrong if and only if any principle that permitted it would be one that could reasonably be rejected by people moved to find principles for the general regulation of behavior that others, similarly motivated, could not reasonably reject" (Scanlon 1998: 4). This can also be turned into the anti-arbitrariness principle, when "a principle that permitted the act" is interpreted as the justification rule for a choice, "could reasonably be rejected" is interpreted as "cannot be consistently wanted when universally complied", and "by people" means "by at least one person". Scanlon's theory is reason-based, where a reason must be one "no one could reasonably reject as a basis for informed, unforced, general agreement" (1998: 153). The general agreement contains an anti-arbitrariness condition: an agreement by everyone is required, without arbitrary exceptions.

One important difference between Scanlon's principle and the anti-arbitrariness principle, is that Scanlon only considers a restricted group of people that could reasonably reject a principle, namely those people who are moved to find principles. This reflects a contractualist position, as only those people are able to mutually agree to a "contract", i.e. a set of principles for the general regulation of behavior. In contrast, as the definition of unwantedness refers to someone's subjective preferences, the anti-arbitrariness principle includes everyone who has preferences. That includes, e.g. young children and non-human animals.

As Scanlon, Rawls (2005) also proposed a contractualist principle which is characterized by its reason-giving nature, where a reason must be one others can "reasonably be expected to reasonably endorse" (2005: 450). Such an endorsed reason could be reinterpreted in terms of the unwanted arbitrariness principle, where the reason refers to a selection rule that justifies the selection of an element such that this selection is not arbitrary, and the endorsement of the reason refers to the selection rule not being unwanted by anyone.

Parfit made an important attempt to unify the Kantian and contractualist moral theories with a third theory, rule consequentialism, by suggesting that their fundamental principles could be interpreted in a converging way. This "Triple Theory" is summarized as (Parfit 2011: 412):

> An act is wrong if and only if, or just when, such acts are disallowed by some principle that is
> 1. one of the principles whose being universal laws would make things go best
> 2. one of the only principles whose being universal laws everyone could rationally will, and
> 3. a principle that no one could reasonably reject.

The second and third conditions represent Kantianism and Scanlonian/Rawlsian contractualism. The first condition refers to rule consequentialism (which says that everyone following the obligatory rules or principles generates the best consequences). Again, its reference to principles being universal laws reflects an anti-arbitrariness condition, but the words "making things go best" require more translation work to arrive at the anti-arbitrariness principle. Perhaps what makes things go best is a kind of preference satisfaction, such that a bridge can be built with the notion of unwantedness.

Expressed in a shorter "Kantian contractualist" formula, Parfit (2011: 342) claims: "Everyone ought to follow the principles whose universal acceptance everyone could rationally will." This unified formula turns into the anti-arbitrariness principle, by translating "Follow the principles", "universal acceptance" and "everyone could rationally will" into respectively "give justification rules", "everyone follows those rules in all possible situations" and "everyone can consistently want." This suggests that the anti-arbitrariness principle is like Parfit's Triple Theory, a kind of unification of Kantian, contractualist and rule consequentialist fundamental ethical principles.

5. *The structure of arbitrariness*

We can study unwanted arbitrariness by the most simple but sufficiently general structure: a choice set containing two elements {X,Y}. One could choose both elements, in which case there is no arbitrary selection of elements (there is only one way to select both elements). Or one could choose one element, either X or Y. This allows room for arbitrariness: if X is chosen, one could ask for the selection rule why X instead of Y is chosen. Finally, one could choose none of the elements, in which case there is no arbitrariness possible. All the options can be grouped together in the power set of all subsets: {{X,Y},{X}, {Y}, {} }. This power set has a hierarchy with several levels:

- Top level (no arbitrariness possible): {X,Y} (the full set of all elements)
- Intermediate level (arbitrariness possible): {X} or {Y} (the subsets of individual elements)
- Bottom level (no arbitrariness possible): {} (the empty set)

Only at the intermediate level is arbitrariness possible. This arbitrariness can be called first-order or horizontal arbitrariness, because there is another, meta-level arbitrariness possible, namely the choice of the level. We can consider the set of levels: {Top level, Intermediate level, Bottom level}. If one chooses the top level without following a selection rule, that choice is arbitrary. This second-order arbitrariness can be called vertical arbitrariness. One could use a selection rule, such as "choose the level that does not allow for horizontal arbitrariness and contains at least one element", that uniquely selects the top level. Now the choice for the top level is no longer arbitrary (i.e. no vertical nor horizontal arbitrariness), but the choice of the selection rule can be arbitrary, because one could equally choose a selection rule such as "choose the level that does not allow for horizontal arbitrariness and contains no elements" (which selects the bottom level) or "choose the highest level where horizontal arbitrariness is possible" (which selects the intermediate level). Hence, there is a third-order arbitrariness. Avoiding this arbitrariness requires a fourth level, where a fourth-order arbitrariness occurs. This indicates that there will always be some arbitrariness: there will always be some level n with an n-order arbitrariness. It is impossible to avoid all arbitrariness.

6. *How to avoid unwanted arbitrariness?*

Horizontal arbitrariness involves choosing an element from a choice set. One way to avoid unwanted horizontal arbitrariness is by choosing the full set of choices (the top level) or choosing the empty set (the bottom level). However, it may not always be possible to choose the full or the empty set, because of some logical inconsistency. It may also be less desirable to choose the top or the bottom level. This undesirability happens in a general sense when at least someone cannot consistently want the full set or the empty set, or it happens in a more strict sense of "preference dominance" (similar to "Pareto dominance"): when those who cannot consistently want the intermediate level also cannot want the top or bottom level, and at least one person who can consistently want the intermediate level cannot consistently want the top or bottom level (in this case the top or bottom level is preference dominated by the intermediate level). We can categorize the situations where choosing the intermediate level is unavoidable or more desirable.

The full set and empty set are impossible: these situations often involve a choice set {do X, don't do X}. Of course, choosing both or choosing neither, is impossible.

The full set is impossible, the empty set undesirable (i.e. not wanted by at least someone): consider a choice between moral theories {moral theory X, moral theory Y}. Moral theories, such as a utilitarian welfare ethic and a deontological rights ethic, are based on universal principles. We may have a choice between {maximize total welfare, minimize the use of people against their will as merely a means to someone else's ends}. Respecting both principles of both utilitarian and deontological theories is logically impossible: there are cases when maximizing welfare involves using people as a means against their will. Choosing none of the principles and moral theories is not impossible, but it is undesirable, because it is likely that at least someone cannot consistently want an anything goes situation without guiding ethical principles.

The full set is undesirable, the empty set impossible: suppose that helping both persons X and Y is impossible, and one faces a choice between {don't help X, don't help Y}. It is possible to choose both, but if both people want to be helped, this is less desirable than choosing either one of the options.

The full set and the empty set are undesirable. An instructive example is the choice of road traffic laws, such as the choice set: {make driving left permissible, make driving right permissible}. Choosing none of the options implies a prohibition of driving, and there are people who want to drive. Choosing both options results in more unwanted traffic accidents. Another example is: {eliminate starvation by feeding hungry people, eliminate starvation by killing hungry people}. Hungry people cannot consistently want the empty set, because that means not eliminating starvation. And they do not want the full set either, as that involves killing hungry people.

If choosing the intermediate level is unavoidable or more desirable, we might face horizontal arbitrariness, unless we are able to use a selection rule that selects one of the elements at the intermediate level. We can look for a rule "If a set of conditions C are satisfied, then choose X instead of Y." Now the challenge becomes choosing a proper set C of selection rule conditions that everyone can consistently want (otherwise, the choice of the selection rule itself generates unwanted arbitrariness). If such conditions cannot be found, then we have truly unavoidable unwanted arbitrariness.

One starting point for the selection rule could be: "If choosing X can be consistently wanted by most people, then choose X." It is already possible that everyone can consistently want this condition C that represents the majority criterion. If there remain some people who can reasonably object against this majority criterion, then they can propose another criterion (i.e. another set of conditions for the selection rule). Now we face the choice of selecting an element from the set {majority criterion, another criterion}. Choosing both elements (the full set) is impossible, choosing the empty set undesirable. To avoid horizontal arbitrariness, we need another, higher level selection rule that selects either the majority criterion or the other criterion. This process can continue to even higher levels. We can go on as far as is feasible, to minimize unwanted arbitrariness. But the further we go, the more important the choice of a higher level selection rule becomes, the more depends on it, and the harder it becomes to have reasonable objections against the choice. The preferred higher level selection rule becomes so fundamental, that one is likely to have a strong preference for it. It is for example already difficult to have a stronger preference for another criterion than the majority criterion. That means the majority criterion selection rule is likely consistent with someone's largest consistent set of that person's strongest subjective preferences.

With the above line of reasoning, we can apply the anti-arbitrariness principle to itself. The choice set involves the two options {avoid unwanted arbitrariness as much as possible, don't avoid unwanted arbitrariness as much as possible}. Choosing both or none of the options is impossible. So we are stuck at the intermediate level, where we can arbitrarily pick one of the two options. But picking the second option (not avoiding unwanted arbitrariness) immediately becomes extremely unwanted. Allowing avoidable unwanted arbitrariness has so many ramifications, that it is likely in contradiction with anyone's largest consistent set of strongest subjective preferences. So you cannot consistently want the arbitrary choice for the second option.

To see this in more detail, suppose that you disagree with the anti-arbitrariness principle. You say that avoidable unwanted arbitrariness is permissible. But then you cannot give reasonable counterarguments when I allow unwanted arbitrariness in my moral choices. I may follow arbitrary principles that you cannot consistently want. When I impose my choices on you, you are not able to complain. You are not able to give justified arguments against the imposition of my choices, because you acknowledged that unwanted arbitrariness is allowed, and hence that it is permissible to arbitrarily ignore or violate someone else's largest consistent set of strongest preferences.

If you permit unwanted arbitrariness, I can say to you that your moral values and judgments are not valid. And if you complain and say that your ethical theory is valid, then I can reply that if you are allowed to arbitrarily exclude other moral views and make an ad hoc exception for your own moral rules, then so am I. So I may even make the exception that everyone's moral views should be respected, except yours. All your objections can easily be bounced back by saying: "If you are allowed to arbitrarily do that, then so am I, and so is everyone. What would make you so special that you are allowed to arbitrarily exclude others but I am not? You should not arbitrarily pick yourself from the set of all individuals and say that you are the only one who may do that thing." In summary: rejecting the anti-arbitrariness principle while avoiding irrationality, is extremely difficult, if not impossible. The above discussion applies to the cases where the top and bottom levels are impossible or undesirable. There are two other interesting categories to consider.

The full set is possible and not clearly undesirable, the empty set is undesirable or impossible. A prime example is the choice set {I decide, you decide}, or {I have a right to vote, you have a right to vote}. Someone has to decide, and at least someone wants to vote, so the bottom level is impossible or undesirable. But choosing the intermediate level and arbitrarily choosing one of the options results in a kind of dictatorship where one person can decide or vote.

The full set is impossible or undesirable, the empty set is possible and not clearly undesirable. Here we deal with choice sets such as {harm person A, harm person B} or {privilege A over B, privilege B over A}. It is undesirable to harm both A and B and it is not possible to privilege A over B and B over A at the same time, so the top level is undesirable or impossible. But choosing the intermediate level and arbitrarily choosing one of the options results in a kind of discrimination where one person is harmed or disadvantaged.

As the anti-arbitrariness principle deals with choices and rules, we are confronted with two important questions. Who decides or chooses the choices and rules? And who is affected by the choices and rules? These two questions relate to the dual problems of dictatorship and discrimination. The next two sections discuss how the anti-arbitrariness principle implies the non-dictatorship and non-discrimination principles.

7. Implication 1: Non-dictatorship

The non-dictatorship principle says that no-one should have the unconditional power to always unilaterally make decisions that negatively affect some other people. A vote is a power (or right) to influence a decision (the outcome of a decision process) made by a group, such that the outcome is more in accordance with one's personal preferences. In a dictatorship, there is at least one individual whose vote is excluded from the decision process and who does not want this exclusion. A dictatorship clearly violates the anti-arbitrariness principle, because the choice for the dictator is arbitrary (as the dictator's power is unconditional, no rule was followed to grant that power), and unwanted (when there are affected people who do not want the decisions made by the dictator).

Suppose person X wants to make choice A, but person Y cannot consistently want the consequences of that choice, and hence prefers choice B. Instead of the principle might makes right, which is a dictatorship of the most powerful, those people can look for other methods to decide who gets to decide. One such alternative method is generating justifications by giving arguments. Instead of the strongest person winning, now the strongest reason, justification or argument wins. The principle that the best argument wins, is also arbitrary, just like the principle that might makes right, but it is less likely to be unwanted.

Person X can simply claim: "I, person X, decides." This is the moral rule: "If the person is X, then that person may choose." Person Y does not want that, and counters: "No, person Y decides." The justification rule proposed by person X refers to X, and that choice should be justified as well. So person X can claim the meta-rule: "Person X decides who decides." But here again, person Y can complain, and the meta-rule arbitrarily refers to person X again. This discussion can go on to infinity. For practical relevance, the anti-arbitrariness principle should state that an infinite regression of justification rules is not allowed.

The non-dictatorship principle can also be applied to moral theories. These theories are logical systems of ethical principles that represent moral intuitions or values. There are different moral theories, such as a deontological rights ethic, a consequentialist utilitarian welfare ethic, a libertarian ethic or pluralist ethics that combine several ethical principles. But which theory should we choose? The anti-arbitrariness principle sets strong constraints on a moral theory. The theory should be coherent in the sense that it should be constructed following some rules, such as:

1) One should not arbitrarily limit the ethical principles to an arbitrary group of objects, beings or individuals.
2) One should not arbitrarily give weaker (less strongly felt) moral intuitions stronger priority. One should not arbitrarily change or exclude basic moral judgments.
3) One should not arbitrarily allow inconsistencies and gaps in the ethical system.
4) One should not arbitrarily introduce ambiguous or vague principles that one can interpret and apply arbitrarily in concrete situations.
5) One should not arbitrarily add artificial, complex, ad hoc constructions and exceptions to save the moral theory from counterintuitive implications.

These construction rules for a coherent theory can be consistently wanted. If, for example, I allow inconsistencies, gaps, ambiguities or arbitrary exceptions in my theory, then I have to accept that your moral theory also contains such things. With such an incoherent theory, you can easily justify choices that I cannot consistently want. An incoherent theory always contains avoidable unwanted arbitrariness that should be rejected.

To avoid dictatorship, everyone is allowed to construct a coherent moral theory that best fits one's moral intuitions and values. Incoherent theories are impermissible. But there are many possible coherent moral theories. We do not have a rule that determines which of those coherent theories is the best. If we are against unwanted arbitrariness, we have to recognize that every equally coherent moral theory is equally valid. I cannot say that my coherent theory, based on my moral intuitions, is better than yours if both our theories are equally coherent. I prefer my theory, but I cannot impose my theory upon you, because what would make me so special that I would be allowed to do that? And the same goes for you and everyone else. It would be an unwanted kind of arbitrariness if I claim that my moral theory is special without good reason.

So picking one of the coherent moral theories always involves unavoidable arbitrariness. The non-dictatorship principle says that we should democratically choose which moral theory to apply. And if you follow a coherent moral theory without being able to give a justification rule that selects that theory, you should tolerate that other people follow other coherent moral theories. We should be tolerant towards all other coherent ethical systems, no matter how much they go against our own moral intuitions.

A choice for an incoherent system, on the other hand, does not have to be condoned, because you can give a justification rule "If the theory is incoherent, then it is impermissible to choose it," and everyone can consistently want that everyone follows this rule. If you choose to follow an incoherent theory, I am allowed to reject that theory and impose my theory on you, and you are not able to complain. You are not able to give reasonable or justified counterarguments against the imposition of my ethical principles, because by following your incoherent theory, you are acknowledging that unwanted arbitrariness and hence arbitrary exclusion are allowed. That means it is also permissible to arbitrarily exclude your moral theory and ignore your moral views and ethical principles. You can only give a valid complaint or argument if you accept the anti-arbitrariness principle. Without that principle, any critique becomes invalid and complaints become impossible.

As the ethical systems of, e.g. racists, rapists or religious fundamentalists contain inconsistencies, avoidable arbitrariness, unscientific beliefs and vague principles, they can easily be rejected. If your ethical system is more coherent than theirs, then you can rightfully say that your ethical system is better than theirs and then you may oppose their incoherent systems.

The prohibition of incoherent theories allows us to avoid an extreme form of moral relativism that says that all moral theories, including incoherent ones, are equally valid. This extreme relativism implies that everything would be permissible, and we cannot consistently want that. The non-dictatorial claim that coherent moral theories are equally valid is a kind of weak moral relativism, which is a consequence of the anti-arbitrariness principle.

How do we deal with that plentitude of coherent ethical systems that are equally valid? Everyone (who is able to do so) constructs their own coherent ethical systems, and we can aim for a consensus or democratic compromise between everyone's system by using a democratic procedure. In a democracy, everyone has one vote, and everyone's vote is equally important, because we cannot say that one vote (one coherent theory) is better than someone else's. But those who cannot provide a coherent moral theory that does not contain unwanted arbitrariness, lose their vote. In other words: in this moral democracy it is not allowed to vote for parties who have incoherent moral theories, such as racist parties. Those parties cannot participate in elections.

Note that the coherence of moral theories imposes very strong constraints on the construction of moral theories. We can expect that the resulting theories that people construct, if they follow the anti-arbitrariness principle carefully, are not extremely divergent from each other. This strong selection and convergence of moral theories makes a democratic choice of theory more feasible.

So there are two reasons why our moral theories should not contain unwanted arbitrariness. First, if it contains such arbitrariness, someone else is allowed to arbitrarily reject our theory and we are not able to complain. Second, the avoidance of unwanted arbitrariness puts strong constraints on the possible moral theories, which makes a democratic consensus between the resulting coherent moral theories more feasible.

8. *Implication 2: Non-discrimination*

Discrimination can be defined in different ways, suitable for different contexts (see e.g. Altman 2016). One could for example define discrimination merely as a different treatment of two individuals (or groups of people), but then we must distinguish permissible versus impermissible discrimination and define the latter. The following definition of arbitrary discrimination is suitable to derive the non-discrimination principle from the anti-arbitrariness principle.

Arbitrary discrimination of individual (or group) A relative to B by discriminator C is a systematically different treatment of A and B, whereby

1) B is given more advantages by C than A,
1) C believes A has a lower moral status than B (e.g. A has less intrinsic value or weaker rights than B) in the sense that C would not tolerate swapping positions (treating A as B and B as A), and
3) there is no justification or the justification of the difference in treatment refers to morally irrelevant criteria (properties that are not acceptable motives to treat A and B differently in the concerned situation), whereas A and B both meet the same morally relevant criteria to treat and value them more equally.

The first two conditions reflect unwantedness. The discriminated person A does not want the disadvantage, but also the person C who discriminates does not want swapping positions of A and B. The third condition reflects arbitrariness, i.e. the lack of a justifying rule. Discrimination is based on arbitrariness, and this arbitrariness is avoidable and unwanted, because the discriminated people do not want their negative treatment, their arbitrary exclusion from the moral community.

The anti-arbitrariness principle specifies what counts as morally irrelevant criteria. A criterion or property is morally irrelevant in a specific context (such as political elections or job opportunities), when it is arbitrary (in the sense that there is no non-circular rule that selects the property out of a multitude of similar kinds of properties), or it has a high risk of introducing arbitrariness. The latter happens with, for example, ambiguous properties, properties that are inherently impossible to detect, define or delimit, or non-empirical properties for which there are no objective or scientific criteria and methods—not even in principle—to clearly see whether the property is present. With such properties, there is the risk that one arbitrarily assigns the property to individuals as one pleases. Consider a non-natural property such as a soul, and the claim that only beings that have a soul have rights. The danger is that one can arbitrarily assign a soul to some preferred entities or persons.

With the anti-arbitrariness principle we can derive which properties are morally irrelevant in which contexts and hence result in discrimination in those contexts. Some properties that are irrelevant in, for example, the context of political voting are: physical characteristics and appearances (e.g. skin color, behavior, gender), genetic properties (e.g. race, ethnicity, genetic kinship), supernatural properties (e.g. having a soul), preferences (e.g. sexual, political), and belonging to an arbitrary group.

As a concrete and important example of the non-discrimination principle, consider the choice of moral community: the subset of all entities in the universe that have moral status (in the sense of, e.g. having moral rights). Consider only living beings. According to the biological classification, we can classify living beings in a vertical taxonomic hierarchy, with the taxonomic rank "life" at the top, followed by ranks such as "domains" (e.g. eukaryotes), "classes" (e.g. mammals), "orders" (e.g. primates), and finally the taxonomic rank "populations" (races, subspecies) at the bottom. A white supremacist first chooses the lowest level in this hierarchy (the populations or ethnic groups), and then picks a subset at this level (the ethnic group of whites). Similarly, a speciesist first selects the level of the species, and then selects a specific species (e.g. *Homo sapiens*) as the moral community. If no selection rules were followed, these two choices involve respectively vertical and horizontal arbitrariness. We can first ask the non-trivial question: "Why choosing a species and not, e.g. a biological order or a phylum?" And at the level of the species, we can ask: "Why choosing *Homo sapiens* (humans) and not, e.g. *Sus scrofa* (pigs)?" One could answer: "Because most humans have the capacity for moral thought", but it is possible that this answer also applies to some levels up or down in the hierarchy. If, for example, there are less than 14 billion primates alive, containing more than 7 billion humans with the capacity for moral thought, then the majority of primates have this capacity. Hence, one could equally well first select the level of orders and then the order of primates. By selecting a biological group as a moral community, it is not easy to avoid arbitrariness.

The definition of discrimination means you can avoid discrimination in three ways: either treating A and B equally, tolerating swapping their positions or justifying the preferential treatment using non-arbitrary criteria.

If you tolerate swapping the positions of A and B, you give them equal moral value. This implies that some kinds of partiality are not (yet) discriminatory. Consider a burning house dilemma where you can either save Alice or Bob from the flames. Suppose you want to save Bob first because he is your child, whereas Alice is a child from another country, with another skin color. Non-discrimination does not imply that you should flip a coin and give each child an equal 50% survival probability. You are not a racist or sexist (at least not necessarily) if you want to save Bob, as long as you do not condemn someone else who wants to save Alice. If you criticize someone who saved Alice, and you do so by using arbitrary criteria such as skin color or gender, then you discriminate and then it becomes racism or sexism. It permissible for you to show partiality to Bob for reason r (for example, because you feel attachment to Bob) if you tolerate others failing to show partiality to Bob for reason r.

Considering the above, we can formulate the following ethical principle of tolerated partiality: when helping others, you are allowed to be partial in favor of one individual or group (e.g. your own child), as long as you tolerate someone else's choice to help the other party (e.g. another child). In this sense, saving your child is not inconsistent with the claim that all children have an equal moral value. Two children can have different personal values for you, but they inherit an equal moral value when a tolerated symmetry (swapping their positions) is satisfied. Having a stronger empathic connection for one individual or having a stronger inclination to save one individual instead of the other, and acting on those feelings, is not necessarily discrimination.

This principle of tolerated partiality can be derived from the unwanted arbitrariness principle: everyone should tolerate your preference for saving the people you hold dear, even if your selection of those people is arbitrary (e.g. from my perspective), because everyone can consistently want to be able to save the people they hold dear.

What if you do not tolerate swapping the positions of Alice and Bob? Suppose Bob is your child and Alice is the name of my car. You would not tolerate me saving the car. The definition of arbitrary discrimination implies that to avoid discrimination, there must be a valid reason or justification, based on non-arbitrary criteria, why one entity (the child) is more important or valuable than the other (the car). In this example you can easily give a valid reason: the child has preferences to be rescued, to keep on living and to avoid the pain from the flames, whereas the car does not care at all about being burned or rescued.

Similarly, suppose you give a piece of chocolate to Bob, a child, instead of Alice, a dog. You have a non-arbitrary justification: chocolate is unhealthy for dogs. Being able to safely eat chocolate is a non-arbitrary criterion, because both the dog and the child prefer safe food. Non-discrimination does not say that we must treat everyone the same and give everyone the same food.

However, some reasons are invalid in cases when you do not tolerate swapping positions. For example, the reason to save Bob instead of Alice because Bob belongs to a certain social group or believes in a certain God. Those invalid reasons refer to arbitrary criteria, such as skin color, religious beliefs or group membership. A white supremacist might help Bob instead of Alice (and does not tolerate someone saving Alice instead of Bob) based on their skin colors, but what does skin color have to do with a preference for being helped? Skin color is but one bodily characteristic, and it is arbitrary to claim that this particular characteristic relates to subjective preferences.

In summary, when swapping positions is not tolerated, the reason should not be arbitrary. When swapping positions is tolerated, the arbitrariness of the reason is not problematic. 'Being your child' may be an arbitrary reason to save your child, because what does that have to do with a preference for being helped? So if you use this as your reason, then you have to tolerate swapping positions (i.e. someone else saving another child).

To avoid discrimination, we have to expand the moral circle (Singer 2011). This expansion visualizes the traditional approach in a rights-based ethic. One traditionally starts with the list of rights and then asks the question: what are the entities in the world that should get these rights? Then we see an expanding circle: from the individual to the family to the tribe to the ethnic group to the species, ending up with the Universal Declaration of Human Rights. But selecting some entities or persons is arbitrary, and the consequences of this selection cannot be consistently wanted by individuals who are not selected.

The anti-arbitrariness principle suggests a reverse approach: to avoid arbitrary exclusions, we first start with the condition that everyone and everything gets rights. Then we ask the question: what are the basic rights that should be granted to all entities in the world?

Of course we cannot grant all possible rights to all entities, because that results in contradictions. Hence, the choice of rights might involve unavoidable arbitrariness. To avoid unwanted arbitrariness, we can look for the rights that are least unwanted or that can be selected following some rule.

Consider the right not to be killed. This right is trivially satisfied for non-living things, but if all living things get this right, we are no longer allowed to kill and eat plants. We can restrict this right to the right not to be killed against one's will. The addition of "against one's will" is possible, because everyone can consistently want such addition. Assuming plants do not have a consciousness and hence no will, this right is trivially satisfied for plants: even if we eat them, we do not kill them against their will and hence do not violate their right. We can easily grant plants this right.

But this right can still be unwanted: there are situations where we can save many people, only by accidentally or unintentionally killing one person against his will. When that person is a rightholder who has the right not to be killed against his will, the presence of that person imposes a cost on others: the other people can no longer be saved. They lose the freedom to be saved. The rightholder becomes an obstacle: it would have been better for the other people if that one person was absent or did not exist.

As argued by Walen (2014), there is however another right that does not impose costs on others: the right not to be used as a means against one's will. One is used as a means for someone else's ends if one's existence and presence is necessary to achieve the ends. If everyone has this right instead of the right not to be killed, it is still allowed to save people by accidentally killing someone. Bringing into existence a person who has that right is not costly or harmful for others, because other people would not have been better-off if the person were absent. Consider the case of an unwanted pregnancy: abortion violates the right of the embryo not to be killed, but not the right not to be used as merely a means. Performing an abortion, the embryo is not used as a means, because the woman could still achieve her end (i.e. not being pregnant) if the embryo did not exist. In contrast, when the pregnancy is unwanted, one could say that the mother is used as a means against her will: her existence is necessary for the embryo to live. The embryo uses the body of the mother against her will.[3]

This right not to be used as means against one's will reflects a Kantian mere means principle (see Kant 1785 and Parfit 2011): if "use" generalizes to "treat" and "against one's will" translates into "merely", the no-mere-means right says that we should not treat someone as merely a means for someone else's ends.

Now we can formulate a selection rule to select this no-mere-means right: choose the right that refers to the person's will and does not impose costs on others (in the sense that others would not be better-off and cannot be made better-off if people who have the right were absent). The absence of costs means that it is difficult to complain against granting people this no-mere-means right. With the selection rule and the difficulty to complain, choosing this no-mere-means right is likely to avoid unwanted arbitrariness. Everything gets this right, but the right is only non-trivial for individuals who have a will (which consists of subjective preferences). This includes children and animals. As a practical result, this right imposes a duty of veganism. If animals have negative experiences when their bodies are used for food, their no-mere-means right is violated. And if only humans and some preferred non-human animals get this no-mere-means right, we are guilty of discrimination.

9. *Conclusion*

The anti-arbitrariness principle states that everyone who makes a choice has to avoid unwanted arbitrariness as much as possible. This principle strongly relates to Kantian, Scanlonian and Parfitian cate-

[3] Note that in this sense, using someone as a means does not have to be intentional.

gorical imperatives (Kant 1785; Scanlon 1998; Parfit 2011). Its most important implications are non-dictatorship and non-discrimination.

I will leave this discussion with some open questions for further research. Could the anti-arbitrariness principle be too strong in the sense that it prohibits too many ethical principles and choices that we deem to be valid and permissible? Could we find some kinds of arbitrariness that can still be justified, even if someone cannot consistently want them? Are there other fundamental ethical principles, conflicting with the anti-arbitrariness principle, that everyone can consistently want? If yes, can those other principles be justified? And when people have different coherent moral theories but cannot find a democratic consensus, how can we select the best moral theory? The latter moves us to the area of “normative uncertainty” (MacAskill 2014).

References

Altman, A. 2016. “Discrimination.” In Edward N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/archives/win2016/entries/discrimination/.

Bentham, J. 1789. *An Introduction to the Principles of Morals and Legislation*. London: T. Payne & Son.

Carson, T. L. 2010. “Kant and the Absolute Prohibition Against Lying.” In T. L. Carson. *Lying and Deception: Theory and Practice*. Oxford: Oxford University Press, 67–88.

Kant, I. [1785] 1993. *Grounding for the Metaphysics of Morals: Third Edition*. Translated by James W. Ellington. Indiana: Hackett.

MacAskill, W. 2014. *Normative Uncertainty*. PhD thesis. University of Oxford. https://ora.ox.ac.uk/objects/uuid:8a8b60af-47cd-4abc-9d29 400136c89c0f/files/m0e6d06ceaf493f85c33c6faee369d19b

Parfit, D. 2011. *On What Matters*. Oxford: Oxford University Press.

Rawls, J. 2005. *Political Liberalism*. New York: Columbia University Press.

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

Singer, P. 2011. *The Expanding Circle: Ethics, Evolution, and Moral Progress*. Princeton: Princeton University Press.

Walen, A. 2014. “Transcending the Means Principle.” *Law and Philosophy* 33 (4): 427–464.