Consequentialism, Decision Theory, and Weak Sufficientarianism

Alexander Christen

Binghamton University

Abstract

This essay was written for an Honor Thesis of the fall of 2019, regarding consequentialist moral philosophy. As such, it does not directly address its application to the law or the legal institutions. However, as a general moral theory, if this argument succeeds it would provide for just laws and an overall goal for law to attain. Judges, for example, could employ this theory as an underlying principle of the law to make ruling in judicial review; how far certain freedoms extend and at what cost require a normative framework to gauge, and this theory could provide it. In this paper, I argue for weak sufficientarism, as a preferable alternative to other consequentialist theories. I first contend that consequentialist decision making should employ self-interested decision making behind the veil of ignorance to find the best outcomes. Then, I propose that classical decision theory fails to model rational self-interest behind the veil of ignorance, when compared to weak sufficientarian decision making. Weak sufficientarian decision making incorporates our primary goal of experiencing a life worth living, to reach this conclusion. Finally, I address possible counter arguments from utilitarians and prioritarians. Following from this theory, certain economic rights should lose legal protection, if it means guaranteeing an adequate life to others. The nature of legal punishment would likely need to be overhauled, insofar as prisons make lives not worth living. New positive rights essential to achieving a sufficiently good life, like access to psychical and mental healthcare may be required by Weak sufficientarianism. Overall, weak sufficientarianism could provide a broad basis for the goals and functionality of the law.

Keywords: Consequentialism, Utilitarianism, Prioritarianism, Sufficientarianism, Distributionism, Well-being

Introduction

Imagine that you can determine how an entire society operates and you are given two options: one society guarantees everyone a lower middle-class existence, while the other gives the majority luxurious lives by exploiting a minority of the population for that end. Assuming the net well-being of either society is equal, can you prefer one society to the other? Most deontological moral theories can account for fairness and distribution, while mainstream consequentialism struggles to yield any preference for distributions. Utilitarianism, as one of the most popular theories, demands the promotion of the highest average well-being, with complete indifference to distribution. Famously, John Rawls (1971) objected to utilitarianism because it fails to consider the "distinction between persons" (p. 27). Though Rawls was correct to emphasize this shortcoming, he then uses it to defend his contractualist model, which, for many, is a leap too far from consequentialism. Many theorists have since tried to abridge this gap and defend other consequentialist variants such as prioritarianism, egalitarianism, sufficientarianism, and person-based consequentialism. However, with the exception of person-based consequentialism, these theories seem to sidestep the objection rather than address it or incorporate it. These defenses seem to placate consequentialists who feel the force of the objection, rather than those who object to consequentialism on this ground. In actuality, this objection should not be any threat to consequentialism at all, when properly developed.

A single value unites almost all consequentialist theories: the goodness of well-being to sentient beings. To appropriately weigh this value, I argue, actually requires judging consequences with rational self-interest from behind a veil of ignorance (*the Value Claim*). Further, I contend that these rational decisions behind a veil of ignorance would prioritize achieving a sufficiently good life, and attempt to maximize welfare as a secondary concern (*the Decision Theory Claim*). Combining these claims, the resultant theory incorporates the separateness of persons without sacrificing consequentialism. In practice, the best outcomes under this theory generally provide the most individuals with the highest, sufficient levels of well-being, with some exceptions. Call this view weak sufficientarianism, which stipulates that attaining sufficiency often dominates concerns of expected utility, but not always. This particular version will also allow for some prioritarian judgement between alternatives that equally address sufficiency, meaning even then utility maximization may not be preferable. Overall this argument relies on two independent claims, meaning that one could reject any consequentialist foundation to the veil of ignorance while agreeing about decision-making behind it. As such, the remainder of this paper will be divided into two main parts, one arguing for each claim. Finally, I will address potential objections to my argument.

Important Asides

Due to the length of this paper, not every detail can be addressed in full. The first major detail is the value and nature of well-being. For this argument, the value of well-being is assumed. Should someone reject this value, this argument will not contest that. Furthermore, this argument will not articulate nor hinge on any specific conception

of well-being. Any of the major conceptions could fit into this theory quite easily. In fact, any conception of well-being that allows quantifying the relative levels of well-being between persons would work. With the assumption that preference hedonism correctly identifies the nature of well-being, this paper will use the term "well-being" interchangeably with "happiness" and "pleasure." Though as stated, this argument does not require any particular theory of well-being.

The second and much more controversial aside concerns personal identity over time, specifically the case of reductionism and the *Extreme Claim*. Derek Parfit (1984) describes reductionism as the view that "a person's existence just consists in the existence of brain and body, and the occurrence of a series of interrelated physical and mental events" (p. 211). He further describes the *Extreme Claim* as "if reductionism is true, we have no reason to be concerned with our own futures" (Parfit, 1984, p. 307). Should someone argue in favor of the *Extreme Claim*, they may wish to deny that any individuals can persist through time. Without individuals, it would make little sense to evaluate consequences in terms of these individuals. Thus, utilitarianism would be the correct moral theory after all, since it does not put any value on individuals.

This conclusion does not necessarily follow from the *Extreme Claim.* In fact, the only change to this theory of consequentialism would be regarding each instant of a mind's existence as an individual rather than individuals that exist over time. Consider the following example:

> At some point in the future, individuals A and B will exist for the shortest segment of time possible. Right now, you can ensure either that A will

> experience 10 hedons while B experiences 0, or that both will experience 5 hedons.

In this situation, no individual will be regarded as existing through time, and thus the effect of the *Extreme Claim* is nullified. It could still be argued that ensuring the more egalitarian outcome would be more valuable than the other outcome. Such an argument will not be made in this paper. Applying distributive consequentialism at the macro-level would then produce some rather counterintuitive results about what someone could even do for their own future good. Reductionists need not accept the *Extreme Claim* either, and can allow for individuals having special concern for their own futures. Rather, they may endorse the *Moderate Claim*, which stipulates that a relation, such as psychological connectedness, may give individuals special concern with their own futures (Parfit, 1984, p. 311). If one accepts a form of the moderate claim, and allows themselves a special concern with their own future, they can recognize that others can be specially concerned with themselves over time. Thus, someone concerned with the well-being of another can account of separate individuals, insofar as they allow for a special concern for one's future. This overall topic cannot nearly be settled nor addressed in full now. The important takeaway remains that reductionism does not necessarily invalidate all forms of distribution-sensitive consequentialism.

The Value Claim

Determining what constitutes good outcomes challenges all forms of consequentialism. Unsurprisingly, far too many authors wrote about this topic to address

them individually. Nevertheless, some of the most prominent positions ought to be articulated and discussed. These positions are classical utilitarianism, prioritarianism, person-based consequentialism, and egalitarianism. None of these theories properly judge the goodness of outcomes in terms of the value of well-being. Instead, the value of well-being requires that we judge outcomes through self-interest behind a veil of ignorance.

Utilitarianism, as the oldest consequentialist theory, remains one of the most influential. Henry Sidgwick provides one of the most popular accounts of utilitarianism. He describes utilitarian theory as "the conduct that which, under any given circumstances, is objectively right, is that which will produce the greatest amount of happiness on the whole" (Sidgwick, 1890, p. 409). Sidgwick (1890) grounds this claim in "egoistic hedonism" and as such also calls his theory "universalistic hedonism" (p. 9). In other words, Sidgwick first argues that an egoistic individual would maximize their well-being. Then, a non-egoist, seeking good for all, would maximize the well-being of everyone, and thus act as a utilitarian. Overall, this argument relies on the idea that individuals should always try to maximize their well-being, and that this intrapersonal good can apply interpersonally, promoting the maximization of well-being.

John S. Mill articulates a slightly different argument for utilitarianism. According to Mill (2011), "the utilitarian doctrine is, that happiness is desirable, and the only thing desirable, as an end" (p. 43). Essentially, Mill argues that well-being constitutes the good. Next, Mill (2011) stipulates that "each person's happiness is a good to that person and the general happiness, therefore, a good to the aggregate of all persons" (p. 43). Thus, Mill

stipulates that the maximized aggregate well-being determines the best outcomes. Sidgwick's argument lends itself to the idea that more total well-being makes outcomes better, and Mill's also stipulates the same. However, Mill's argument can better justify average utilitarianism than Sidgwick's; the general happiness could be stipulated as the one with the highest average utility. Oftentimes, aggregative and average utilitarianism reach the same conclusions, except for in issues like Parfit's (1984) *repugnant conclusion*: the problem of selecting a society with a few happy people or very many just barely happy people (p. 388). Because averaging better avoids this paradox, "utilitarianism" will refer to average utilitarianism for the rest of the paper. This version of utilitarianism still relies on the notion that the interpersonal good requires a form of maximization.

Prioritarianism attempts to sidestep problems present in utilitarianism. Derek Parfit (1997) defines the prioritarian position as "benefiting people matters more the worse off these people are" and clarifies that prioritarianism measures "absolute levels" of well-being rather than relative levels (p. 213-214). This view diverges from utilitarianism by putting a diminishing value on well-being. Gains in well-being do not become less valuable the more an individual has. Instead, the value placed on improving the well-being of others increases when an individual has less well-being than if they had more. At first glance this view can be very appealing but it retains some issues. Consider the following example:

> Two individuals, A and B, could have the following levels of well-being: A has -50 hedons, B has -1 hedon, or A has -53 hedons and B has 1 hedons.

The prioritarian view would demand the first distribution since A's benefit of 2 hedons matters more than B's benefit of 3 hedons. Given the overall terrible situation that A has regardless of the distribution, why should their 2 hedons value more than the 3 hedons of B who has a decent life going already? This example does not disprove prioritarianism, but shows that it may sometimes overvalue those worse off. Further, it shows that valuing the worst-off more without exception may not correctly model the real value of well-being.

Between utilitarianism, prioritarianism, and weak sufficientarianism, the same fundamental assumptions ground each theory. These assumptions are that 1) well-being is the sole good for sentient beings and 2) moral actions are those that produce the best consequences. With the exact same starting assumptions, these theories should all reach the same logical conclusion, yet they do not. This divergence can rationally occur, as Jörg Schroth (2008) points out, because these theories disagree about how the good of well-being relates to, and how it ought to yield, the best consequences. Basically, prioritarianism, utilitarianism, and weak sufficientarianism can start with these same two assumptions and result in different theories because they stipulate different ideals about the goodness of well-being relates to the best consequences. Utilitarianism might derive from the assumption the best consequences are those that have the maximal average well-being, whereas prioritarianism may originate from the assumption that the best consequences are those that maximize the *value* of well-being (because well-being for the worst-off is more valuable than the better-off). Weak sufficientarianism, however,

derives from stipulating that the best consequences are those that are in the best interest of those involved: the best outcomes of well-being concerning all sentient beings.

Opposed to prioritarian or utilitarian conceptions of the good, the one of weak sufficientarianism follows from choosing the best outcomes concerning all sentient beings. This value does not imply maximization in the way that the utilitarian contends. Well-being in the abstract does not matter, rather, the sentient beings do; well-being only has value because it constitutes the sole good for sentient beings. Hence, the only *real* value is sentient beings. Merely aggregating well-being would presume that well-being matters intrinsically, and loses sight of sentient beings. Consequently, mere aggregation cannot constitute the best outcomes. Instead, what constitutes good outcomes must be the outcomes in relation to the individuals involved; essentially, the best outcomes are those that yield the best outcome concerning the collective interests of all or, in other words, yield the best value concerning the lives of all sentient-beings. Prioritarians and utilitarians likely would contend that their theories are in everyone's best interest. Mill and Sidgwick, in particular, define the best consequences as what is best for everyone: the collective good. Yet, they conflate the best interests of all with maximizing total or average well-being, without proper justification. What constitutes the best interest of everyone does not have an immediately clear answer, and this problem divides the consequentialist theories.

If basically all consequentialists argue that the best outcomes are those in the best interest of all sentient creatures, then a way of ensuring that outcomes are in the best interest of sentient creatures must be determined. Clearly, these consequences cannot

literally be in the best interest of every creature involved; certain alternatives will often be better for certain individuals while worse for others and trade-offs will persist between alternatives. Instead, the alternative that is in the best interest of everyone is the alternative that remains the most preferable, while equally valuing the situations of every single sentient being involved: the collective best interest. In order to ensure that every individual gets accounted for equally, one ought to evaluate alternative outcomes from behind a veil of ignorance. The veil of ignorance, Rawls (1971) stipulates, entails that an individual does not know who they are in a given distribution, nor any of their qualities (p. 12). Without this knowledge of themselves, one cannot reasonably favor the situation of any given individual; to do so would be completely arbitrary. If alternative outcomes of individual well-being are evaluated behind the veil of ignorance, any rational adjudicator must equally value the good and bad of all individuals in the alternatives. Thus, selecting alternatives from behind a veil of ignorance equally and fairly concerns the interests of all sentient-beings involved.

Once behind the veil of ignorance, a self-interested adjudicator would then pick their most preferred alternative, which would constitute the alternative that is in the best interest concerning everyone. The veil of ignorance ensures that they could be anyone in a distribution; thus in attempting to secure their own best interest, they must treat everyone's interests as their own, since they will live as any one of them. Further, since it removes complete knowledge of the self, every adjudicator would behave identically behind the veil of ignorance, have an identical self-interest, and in turn, decide the same way. Regardless of who evaluates, the same outcome will be preferred. Thus, an

adjudicator behind the veil of ignorance would both decide how everyone else decides, and equally concern themselves with the good of every single individual. Because of these constraints, an adjudicator's self-interest necessarily aligns with the group-interest once behind the veil of ignorance. Thus, the best consequence, as the alternative which is best concerning the collect, would be selected from a self-interested decision maker behind the veil of ignorance.

If a consequentialist concedes these points, then they agree with *the Value Claim*: that to appropriately weigh the value of well-being requires judging consequences with individual self-interest behind a veil of ignorance. We value sentient beings, and in turn we value their well-being, because it is their sole good. Since we value everyone equally, we want to establish the best consequences concerning all others. Only by evaluating outcomes behind a veil of ignorance can the interests of everyone weigh equally. Finally, as self-interested adjudicator, that treats everyone's interest as their own due to the veil of ignorance, they would necessarily select the outcome that is in the best interest of all individuals involved. Properly valuing the well-being of sentient creatures requires this conclusion. Coupling *the Value Claim* with *the Decision Theory Claim*, argued next, leads to the establishment of weak sufficientarianism. But without this next claim, utilitarians and prioritarians might still be able to contend that their conception remains the proper consequentialist conception.

The Decision Theory Claim

Employing a self-interested evaluation of outcomes behind a veil of ignorance certainly does not immediately guarantee weak sufficientarianism. A decision rule must

determine the best outcome for an individual. Furthermore, no prima facie decision rule exists; standard expected utility theory, the maximin rule, and other decision theories remain co-equal options, at least initially. In this section, I argue that individual self-interest behind the veil of ignorance would entail risk-aversion due to the way we value well-being. In short, the fundamental goal of sentient beings is to experience well-being, not necessarily attempt to maximize it. In other words, in many situations, a sentient being would act rationally when opting for a sufficiently good outcome, rather than engaging in risk for a maximized expected well-being.

The veil of ignorance used in this calculation will differ significantly from the way initially employed by Rawls. Notably, the use of the veil of ignorance operates as a means to determine the best consequence rather than a contractual agreement about the basic principles of society; Rawls (1971) writes his goal is to determine what "free and rational persons" would agree to with "the initial position of equality" regarding the "basic structure of society" (p. 11). In weak sufficientarianism, agreement plays no role in selecting outcomes nor is it about establishing any moral, let alone, economic and political principles. Instead, the sole use of self-interest behind the veil of ignorance is a heuristic for selecting better outcomes on well-being. This difference is minor, since self-interest and agreement behind the veil of ignorance operate indistinguishably.

Overall, the essential features of the veil of ignorance largely remain the same as Rawls originally stipulates. He described that the veil would ensure that no one "knows [their] place in society, [their] class position or social status, nor … [their] natural assets or abilities" (Rawls, 1971, p. 12). All these principles essentially remove bias from

decision makers behind the veil; they could end up as any individual in the outcome. Still, Rawls (2001) stipulates a significant difference about the veil of ignorance that a consequentialist theory would not allow; it removes knowledge about probabilities (p. 98). He limits knowledge of probabilities since he comes from the Kantian tradition: all individuals must be treated as an end. If probabilities are known, individuals might be willing to sacrifice the good of the one for the many. In place of this absence of probability, this argument will assume "equiprobability" behind the veil of ignorance as articulated by John Harsanyi (1975): "assume that society consists of *n* individuals… [and an individual] would have the same probability, 1/*n*, of taking the place of the best-off individual, or the second-best-off individual," or the worst-off individual (p. 598). Notably, this requires that the adjudicators have complete information of the possible alternatives to work properly. The equiprobability assumption allows for unbiased decision-making, without potentially overvaluing any single individual as could happen behind the original veil of ignorance.

Even with the changes made to the veil of ignorance, the maximin principle may viably still work as the decision theory. Rawls argues that because individuals could be anyone in society, they would adopt the maximin principle, which stipulates that that one should select the outcome that has the best outcome for the worst-off: the best worst outcome (1971, p. 152-153). Further he contends that individuals must live in the society they choose (Rawls, 1971, p. 137). Combining these facts, Rawls argues that a rational individual would adopt that maximin rule, to guarantee the best minimum. In other words, to avoid the prospect of living out a worse life, individuals would be wholly risk

averse. Insofar as Rawls can claim that it is always irrational to risk having a worse life, the maximin remains a strong contender as a decision rule.

However, risking a worse life does not always constitute an irrational decision, but arguably a fully rational one. John Harsanyi wrote powerful criticisms of the maximin rule as a decision rule. He first gives the example of a person deciding to leave their poor job for a much more fulfilling job in another city, which is a plane ride away. Because the plane ride carries a small but non-zero risk of dying in a plane crash, the maximin rule determines that one should remain working at their poor quality job, since that outcome is better than dying in a fiery explosion (Harsanyi, 1975, p. 595). Given that the maximin rule would yield the same result no matter the unlikelihood of the worst outcome, it seems to severely overvalue the worst-off consequences of situations.

Additionally, this same problem with the maximin can be applied further to interpersonal conflicts. Harsanyi considers that two individuals are dying from pneumonia, one otherwise healthy person, and another terminally ill person who will die in a few months regardless, and you can save only one with medication; he claims that one should save the healthy person, whereas employing the maximin rule would entail saving the terminally ill person (Harsanyi, 1975, p. 596). Harsanyi falsely claims the decision of the maximin rule, however, since the worst-off outcome of immediate death remains constant between the two distributions. Nevertheless, the example can be edited slightly to work properly if we assume that the healthy person could live longer with pneumonia than the terminally ill one. In this case, the prospect of immediate death for the terminally ill patient would lead to the maximin principle opting to spare them, rather

than the healthy one. Once again the maximin principle seems to overvalue the worst-off. Given that intuitively the worst-off cannot matter that much more than others, the maximin principle has serious short-comings.

Along with his critique of the maximin decision rule, Harsanyi argues that classical decision theory would be employed behind the veil of ignorance. Classical decision theory purports that individuals rationally should seek to maximize their expected utility; under uncertainty, one would value a payoff by multiplying its probability of occuring with its payoff (Harsanyi, 1975). Importantly, this decision theory assumes absolute indifference towards risk. In the job switching and pneumonia examples that Harsanyi presented, this indifference towards risk seems to avoid the problems of the maximin rule: switch jobs and treat the healthy person respectively. Ultimately, Harsanyi (1975) argues that one needs classical decision rules to avoid the irrational implications of the maximin principle, and consequently, claims that veil of ignorance does not support the contractualist position, but rather the utilitarian one (p. 598). One's expected utility is identical to the average utility of a society. Thus, if using classical decision theory, one decides identically to a classical utilitarian. Overall, the use of the veil of ignorance does not require immediately rejecting utilitarianism. Still, Harsanyi's argument contains a hole; he acknowledges that the maximin rule does contain "approximate validity" in cases where the classical model yields similar results (Harsanyi, 1975, p. 597-598, p. 606). This concession may have been more significant than Harsanyi realized, since he never provides an in depth defense of why one should not deviate to the maximin rule when it has similar, but not completely identical, results.

This potential gap in classical decision making provides the main counter-arguments for Rawls, Lara Buchak, and myself. In defense of the maximin rule, Rawls stipulates that when three conditions obtain, rational agents would opt for the maximin rule. These conditions are: probabilities are unknown (first condition), a satisfactory guaranteeable level exists (second condition), and this guaranteeable level is significantly better than the worst outcomes of alternatives (third condition) (Rawls, 2001, p. 98). Because of the equiprobability assumption stipulated earlier, consequentialist theories cannot allow the first condition to obtain. Nevertheless, even with only the second and third conditions, Rawls could make a strong case. Considering that individuals could guarantee a good life, Rawls would contend, with a single chance to decide on an outcome behind the veil of ignorance, an adjudicator would not risk it, and employ the maximin rule. A simple case could be the following:

> Two distributions of well-being could be selected for 10 individuals: either a single person could experience 100 hedons and the other nine individuals experience -1, or all ten could experience 9 hedons. Assume that 5 hedons provides a satisfactory life.

In this case, both the second and third conditions obtain for the latter distribution, meaning the maximin principle would select it. In contrast, the former maximizes average utility, so utilitarians prefer it. Arguably, the latter distribution is superior since it has a similar average utility but avoids the repugnant outcomes for the worst-off. This shows some support for the maximin rule. Yet, in circumstances where the Rawlsian conditions do not obtain, it is unclear which decision theory should be selected.

As a sort of median theory between Harsanyi and Rawls, Lara Buchak provides a concept of weak prioritarianism that often aligns with the maximin rule, but can provide

answers in all cases. She assumes equiprobability, as Harsanyi does, but criticizes the notion that rational decision making requires indifference about risk; she writes "there are other ways to aggregate and no reason to privilege a linear risk function…[and] there is no special reason to be globally neutral" (Buchak, 2017, p. 618). Buchak pinpoints the fact that classical rationality seeks only to maximize utility and often assumes infinitely repeated interactions. In other words, classical decision-making needs repeated interactions to ground weighting risk as equal to its probability. From here, Buchak (2017) contends that behind the veil of ignorance, agents have a moral imperative to adopt the most reasonable risk-averse utility function (p. 631). In place of neutral risk function, she proposes "Risk Principle," which states that one should assume that others have the most "risk-avoidant attitude within reason" (Buchak, 2017, p. 632). Notably when using the risk principle, individuals behind the veil of ignorance would thus be forced to choose with risk-aversion, since they do not know the risk functions of themselves or others. Overall, since the maximin principle often aligns with reasonable risk-aversion when the Rawlsian conditions obtain, the two would agree. On the other hand, when the Rawlsian conditions do not obtain, weak prioritarianism can provide an answer. Finally, it would avoid the issues of Harsanyi's job change and pneumonia examples because the maximin principle would be unreasonably risk-avoidant.

For Buchak, small gains for the worst-off do not unilaterally override larger gains for the better-off. Her theory does come with a drawback: she must stipulate a new moral imperative to avoid risk (the risk principle) (Buchak, 2017, p. 631-632). For consequentialists, the sole moral imperative would be to create the best outcomes. Thus,

consequentialists cannot appeal to risk-aversion, at least, not in the way that Buchak does. In light of this, I will argue that self-interest behind the veil of ignorance will broadly be risk-avoidant, which results in weak sufficientarianism being the proper consequentialist theory.

While classical decision theory assumes that individuals seek to maximize their expected utility, individuals actually aim to experience a sufficient level of well-being, at least primarily. This claim does not mean that individuals become indifferent once they attain a certain level of well-being, nor does it mean that well-being has diminishing marginal returns. Although individuals will always prefer more well-being to less well-being, it does not follow they should risk experiencing low levels of well-being or negative levels just for a chance of maximal well-being. In other words, individuals rationally prefer guaranteeing a *sufficient* level of well-being to risking a less than sufficient level for maximization. The sheer importance of living a sufficiently good life comes into effect when considering large periods of time: months, to years, to entire lifetimes. Behind the veil of ignorance, individuals make decisions that determine the net value of their whole lives. Only as a secondary concern would individuals try to achieve maximal well-being. These primary and secondary goals ultimately determine the reality of self-interest behind the veil of ignorance. Thus, due to the primary goal, decisions behind the veil would be sufficientarian. Yet the secondary goal, which the primary goal does not unilaterally dominate, makes the sufficientarian decision weak: weak sufficientarianism. Combining these concerns, sentient beings ultimately aim to guarantee, or maximize the likelihood of attaining, lives worth living. With Buchak's

(2017) observation that no special reason makes sentient beings risk neutral, these goals can explain how and why self-interested individuals would weigh payoffs under uncertainty, without simply multiplying it by its probability. Essentially, weak sufficientarianism, like Buchak's weak prioritarianism, allows probabilities to be weighed non-neutrally, as opposed to classical decision theory.

Naturally, appealing to sufficiency in decision making leads to two questions: what constitutes sufficiency, and why is it special? As per the first question, the very idea of sufficiency may not seem strange and hard to defend. After all, it would be strange to imagine someone who experiences a level of well-being just below purported sufficiency, receiving their favorite chocolate bar and then thinking "Ah, wow my life is suddenly much more valuable!" Rather, it seems fair to assume that every marginal benefit remains just as significant as any other marginal benefit, regardless of one's current level of well-being. However, at least one sufficient level exists: a level of well-being greater than zero. Essentially, experiencing a positive level of well-being means that one prefers existence to non-existence: one would rather live than not. The initial claim that the fundamental goal of all sentient beings is to experience well-being appeals to the idea that everyone desires a life worth living, before all else. Since most consequentialists agree that well-being constitutes the sole good for sentient beings, a life worth living merely amounts to experiencing a net positive level of well-being. Arguably, multiple sufficient levels could exist, each pertaining to a special concern. However, this argument only seeks to establish this single one and relies on no others, but does not necessarily disparage the existence of other sufficient levels.

Next, one might reasonably ask: what matters so significantly about this level and does it really warrant such a special concern? As entities that exist merely through sentient experiences, individuals care fundamentally about their own well-being because it determines the value of their existence to themselves. Individuals want to experience well-being because without it, their existence cannot be worth experiencing at all–a positive level of well-being justifies their existence. What matters so significantly about this level is that surpassing it gives sentient beings reason to prefer having experience to none at all. When making decisions behind the veil of ignorance, these individuals know that they will live out the life of an individual in the situation; they must endure in the alternative. Since they make this judgment knowing they will have to exist, they have warranted concern of ensuring that their existence holds a positive net worth to them. If they do not treat this level of sufficiency with this special concern, then they run the risk of living out lives they would rather not have, all things considered. Considering the earlier example of someone just below sufficiency receiving a small benefit, if this small benefit really makes their life on the whole worth living, then it really did make all the difference to them.

Finally, establishing this type of sufficiency appeals to the primary goal of attaining a sufficiently good life, as defined as a life worth living. A utilitarian, for example, might retort that to even hold this primary goal would demonstrate irrationality, and thus weak sufficientarianism would fail to model self-interested rational decision making. This line of argument would fail, since this goal answers the most basic question for rational sentient beings. Us humans, as sentient beings capable of complex rational

thought, serve as a strong example. We are painfully aware of our own existence and also the pains and sufferings involved in it. Why bother existing? Why not just commit suicide, and end one's existence? In accepting this primary goal, we ensure a reason to exist: a life that benefits us more than it hurts us. It gives us reason to deal with suffering, since that suffering in the end would be worth it. In maximizing expected utility, one cannot help but notice that they risk suffering, for no justifying reason, if that alternative does not assure sufficiency. In primarily seeking sufficiency, we always have a reason to exist and always have a reason to endure suffering. This reason grounds our rationality, because nothing else can give us a reason to exist as rational creatures at all.

Even after accepting the existence of a sufficient level, one would still have reason to employ weak prioritarian decision rather than defaulting to utility maximization, when operating beyond concerns of sufficiency. Essentially, even if everyone attains a sufficient level of well-being between alternatives, an adjudicator still has reason to select with risk-aversion rather than with risk neutrality. Classical decision theory uses the mathematically optimal route for obtaining the most well-being, and a self-interested individual would always aim to obtain more well-being. Thus, one could argue that, assuming guaranteed sufficiency, individuals would choose the alternative with the highest average utility, because it would maximize their well-being without the risk of negative life. However, this argument would not model decision making behind the veil of ignorance. Behind the veil, an adjudicator makes a single decision that regards their entire lifetime, while classical decision theory assumes they can repeat the

interaction. Repeated interaction allows for risk neutrality, since enough interactions would produce the payoff with mathematical certainty.

Unfortunately for the classical decision theories, repeated interaction does not exist for decisions behind the veil of ignorance, individuals would not have reason to choose with risk neutrality. Instead, in attempting to experience a life worth living, they would have good reason to prioritize the worst-off; by prioritizing the worst-off, individuals would increase the likelihood they experience the best life that they could. In a sense, they would seek to maximize the probability of living a better life, and not seek to maximize the average level of well-being at the risk of experiencing less. Overall, the primary concern of attaining a sufficient well-being, not only allows risk-aversion toward net negative lives, but also allows risk-aversion toward less good lives. If sentient beings seek to establish sufficiency since it justifies existence itself and enduring suffering, the larger net gains justify it more strongly. Therefore, one has reason to try to ensure having the most justifiable existence risking having a less justifiable one. Further, in many situations, showing a concern for sufficiently good outcome and having special concern for bad outcomes, better models decision making than classical decision theory.

For example, the Allais Paradox shows a large gap in decision theory that weak sufficientarianism would avoid. Maurice Allais came up with the following set of two gambles:

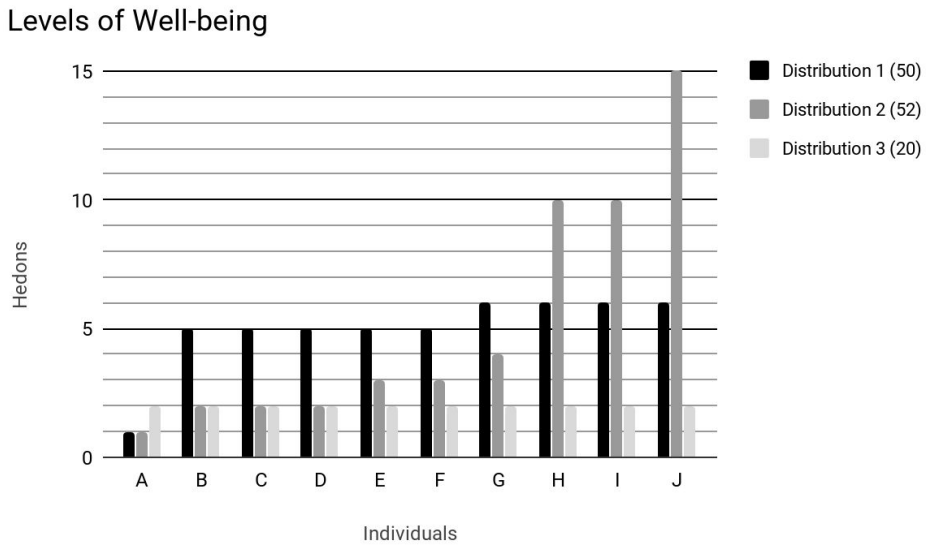
Gamble A					Gamble B			
Ticket	1	2-11	12-100		Ticket	1	2-11	12-100
L_1	$0	$5	$0		L_3	$0	$5	$1

L_2	$1	$1	$0		L_4	$1	$1	$1

(Buchak, 2009, p. 18)

In gamble A, in most cases people would choose L_1 whereas in gamble B most people would select L_4 (Buchak, 2009, p. 18). Under classical decision theory, a self-interested individual must prefer gambles L_1 and L_3 since they have higher levels of expected utility. However, expected utility only represents what one would earn if a gamble was repeated ad infinitum, yet the Allais paradox requires a one time gamble. In this case, classical decision making seems to assume that individuals are acting irrationally, in order to solve the observed choices. That assumption is a hard bullet to bite, because it relies on the strong assumption that individuals must weigh all probabilities equally. On the other hand, weak sufficientarianism can better explain the difference. Gamble A has a large risk regardless with a marginal difference in probability, where no guarantee exists, no sufficient outcome. With such a high prospect of an undesirable outcome, make the good outcomes better. But with Gamble B, one can guarantee a sufficiently good outcome. In that case the same marginal probability becomes much more potent. The risk of receiving nothing does not outweigh the sufficiently good and guaranteed outcome. Classical decision theorists might argue that diminishing marginal utility of money explains the paradox. However, this explanation fails, since then L_2 would be preferable to L_1.

Beyond paradoxes, weak sufficientarianism has a strong intuitive appeal in its selection of distributions, what both classical decision theory and the maximin principle lack. Consider the following hypothetical distribution:



If these distributions were offered behind the veil of ignorance, employing the maximin rule would result in distribution 3, which seems quite irrational. Akin to Harsanyi's objections the likelihood of being individual A is low and their utility is only slightly worse-off. Thus, the maximin rule would clearly overvalue the worst-off in this instance. Because distribution 2 has the highest utility both total and average, a classical utilitarian would select that distribution. Yet distribution two comes with some large arguable drawbacks, first, 60% of the individuals would be significantly better off under distribution 1 than under distribution 2, and that all the best-off in distribution 2 would remain the best-off in distribution. For the utilitarian, however, these considerations do not matter, only distribution 2's expected utility (average), higher by 0.2 hedons, matters. In contrast to both other theories, a weak sufficientarian individual would likely choose distribution 1 behind the veil of ignorance. Knowing they must live out the life one of these beings, a self-interested individual would have the best odds of living a quite good life in distribution 1; distribution 2 has too much risk, and distribution 3 sacrifices too

much to avoid it. Although assuming that 2 hedons provides a sufficiently good life could arguably make both distribution 1 and 3 more appealing more appealing to the weak sufficientarianism, neither would become the best bet. Guaranteeing a sufficient life in distribution 3 may not be worth merely risking a slightly insufficient a large chance of a more than doubly good life. Likewise, merely having the number of individuals at or above sufficiency in distribution 2 would not outweigh the near guarantee of having a life at more than double sufficiency. Notably, prioritarianism would also pick distribution 1, since weak sufficientarianism and prioritarianism always overlap when concerns of sufficiency do not arise.

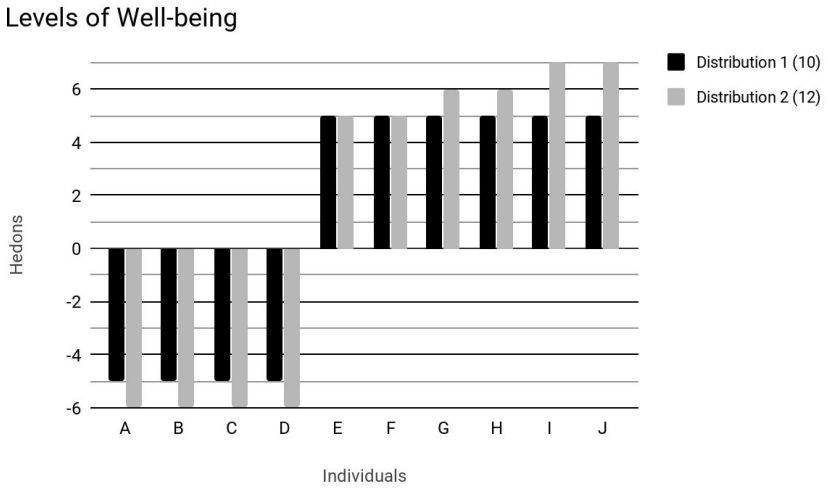
Before discussing how it differs from prioritarianism however, it would help to clarify when and why weak sufficientarianism becomes risk averse. Recall the second and third Rawlsian conditions: a satisfactory guaranteeable level exists, and this guaranteeable level is significantly better than the worst outcomes of alternatives (Rawls, 2001, p. 98). These conditions can be reapplied to weak sufficientarian thinking. The second condition essentially stipulates the existence of a guaranteeable sufficient outcome. The third condition essentially stipulates that the worst outcomes of other distributions are considerably insufficient or less sufficient. Further, contrary to Rawls, these principles can be thought of as sliding scales than conditions with a threshold to meet. So the second condition would obtain in full effect when everyone can enjoy a sufficient level of well-being, in near full effect if 90% of people could attain it, or almost no effect if only 20% of people could attain it. Likewise, the third condition obtains to a higher degree as the quality of the guaranteeable level increases in relation to alternatives,

and less as it decreases. So if both the second and third conditions obtain the maximum degree, self-interested individuals would tolerate absolutely no negative risk, meaning they would not allow risks that would jeopardize them falling below this level.

Furthermore, alternative distributions that maintain the same guaranteeable level, but unequal benefits beyond it would face similar risk-aversion, but possibly tend toward more risk neutrality. For example, in two distributions where everyone attains the guaranteeable level, but either 80% of them experience a slightly higher level or only 40% of them experience much higher level (with a slightly higher average overall), one could still easily select the former, but the latter is not entirely unreasonable. As weak sufficientarianism allows for prioritarian decision making in situation this one, one would likely still remain risk-averse. Contrastingly, in situations where a select few would fall below the guarantable level, for an especially large benefit for the rest, that option might be preferable to ensuring the guaranteeable level; sufficiency concerns do not universally dominate decision making. In short, as the Rawlsian conditions obtain, an individual should act with risk-aversion in regards to falling below sufficiency, though a range of reasonable choices could exist.

Turning to the differences between prioritarianism and weak sufficientarianism, a few differences set the theories apart in practice and in theory. In prioritarianism, the value of a change in well-being fluctuates based on the absolute level of well-being of the recipient, though Buchak's weak prioritarianism would not always operate in this way. In this line, prioritarianism on the whole will give more value and disvalue to positive and negative changes to the worst-off level of well-being. In weak sufficientarianism, this

large value does not always exist. In particular, if the worst off are in a sufficiently terrible situation regardless of distribution, weak sufficientarianism would likely value their changes in well-being equally to those with a sufficiently high level of well-being. Likewise, in situations where everyone exists at a sufficiently high level of well-being, weak sufficientarianism would be more likely to err on the side of maximized utility than prioritarianism. For example, consider the following distributions:



In this example individuals A, B, C, and D, are going to have poor lives regardless. However, as a blanket rule, prioritarianism must give worst-off, and consequently it must regard distribution 1 as better. However, weak sufficientarianism, in the virtue of 40% of outcomes being undesirable with only slight variation, can defer to making good outcomes as best as possible, selecting distribution 2. In this example, it may seem like prioritarians could claim the worst-off do not outweigh the claims of the rest, but some configuration must allow for a situation like this one to obtain. Beyond some intuitive appeal this example may not suggest that weak sufficientarianism is unilaterally better than prioritarianism. However, if the goal of guaranteeing the best sufficient life does

actually matter to sentient beings and gives them cause for risk-aversion, it seems less clear why this goal would allow for risk-aversion for those unable to attain sufficient outcomes. Rather, it seems more likely that a weak sufficientarian would weigh those considerably below sufficiency equally to those above it.

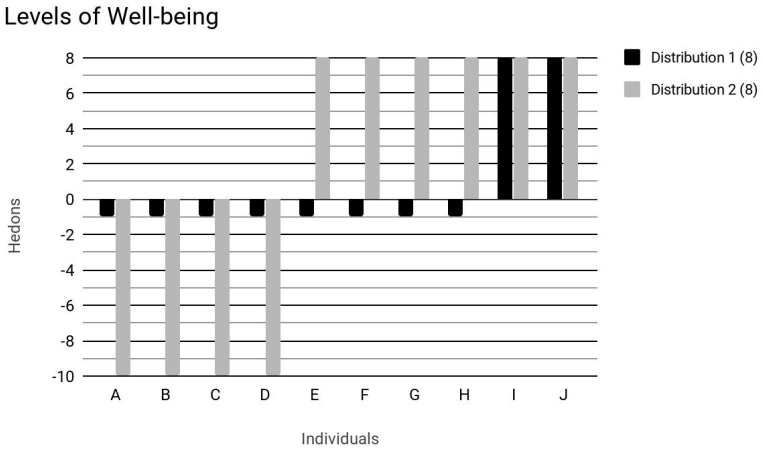
Ultimately, weak sufficientarianism seems more compelling as a theory than prioritarianism, due to its justification of why certain outcomes are better, especially in response to utilitarianism. Parfit's prioritarianism, as a famous example, require just valuing improvements to the worst-off more than others; priority comes from absolute level (Parfit, 1997, p. 214). Parfit cannot really give a better reason for this priority than it just feels like these claims to well-being matter more. Arguably, this feeling derives from the distaste towards certain distributions and jumping onto that reasoning for the explanation. Weak sufficientarianism, on the other hand, relies on the assumption that the fundamental goal of sentient beings is to experience well-being rather always maximize expected utility. Consequently, it can often value improvements to the worst-off over the better off. Buchak (2017), similarly, requires a strange assumption to make her weak prioritarianism: that imposing risk is generally unacceptable in moral terms (p. 631-632). She can impose this consistently since she does not come from a consequentialist angle. For consequentialists, morality derives from the consequences alone. Thus, a consequentialist cannot impose additional moral constraints, because the entire purpose of the veil of ignorance is to discover the contents of morality. Although Buchak's constraints cannot fit, a modified version of the rationality constraints given in this paper

could be used to justify a weak prioritarian theory. This idea, would be one of a few objections that seem to surmount against weak sufficientarianism.

Objections and Answers

In the case of fundamental aim of sentient beings, a prioritarian could feasibly contest that they aim to guarantee the best life possible, even if those lives remain net negative. In this case, one would be employing the case for risk-aversion given in this paper, without appealing to any special level of well-being that counts as sufficiency. In fact, if the case for sufficiency was debunked, this argument would still offer some reason to engage in risk-aversion. Given the difficulties surrounding justifying a sufficient level, why not abandon it entirely? After all, one can easily concede that it matters more to make one's life worth living, from being just barely not worth living, than giving a well-off person an equal benefit, with a prioritarian outlook. Prioritarians would only differing in contending that it matters more to reduce suffering for really poor off person, than give an equal benefit to someone just below sufficiency. A lot of the issues with utilitarianism would no longer persist, and the idea of sufficiency would not need defense.

However without the concern of sufficiency—the concern for a life worth living—and only prioritarian concerns, it must preferable to live in a world where the vast majority of individuals experience a net-negative level of well-being than in a world where the majority of individuals experience a net-positive level . Consider the following distribution of well-being:



In this case, the overall utility remains identical, but the distributions vary differently. In Distribution 1, 80% of the individuals experience -1 hedon and 20% experience 8 hedons, whereas in Distribution 2, 60% of the individuals experience 8 hedons and 40% experience -10 hedons. According to prioritarians, Distribution 1 must be better than Distribution 2, regardless that 80% of the population would prefer not to live their lives at all. In both cases, individuals I and J remain at the same level. The benefit of 9 hedons to individuals A, B, C, and D *must* be more valuable than the same 9 hedons of individuals E, F, G, and H. Prioritarism, by its very definition, demands this perspective. It seems rather dystopian that Prioritarianism could demand that nearly everyone experiences lives they do not want to live for the improvement of a minority that would also prefer not to exist at all. When aiming to guarantee experiencing a life worth living, such a situation could never arise. Although sizable minority would have worse lives, at least their additional suffering exists for the net good of the rest.

For another objection, a utilitarian may argue that the relative disvalue that weak sufficientarians give exceptionally good, but improbable, outcomes relies on a confusion. Recall this previous graph or this example:

> Two distributions of well-being could be selected for 10 individuals: either a single person could experience 100 hedons and the other nine individuals experience -1, or all ten could experience 9 hedons. Assume that 5 hedons provides a satisfactory life.

In this example, as could be applied to other examples given, the utilitarian might argue that it seems unclear how anyone could experience a level of happiness over ten times as great as someone already with the decent life. Perhaps our mere inability to imagine this happiness leads to unjustifiably discounting the goodness of the outcome compared to the misfortune of others. Or maybe no conceivable life could reach such a good level and thus it means nothing to discuss such an example. Indeed these objections carry weight; it seems quite likely that there is a limit to how much well-being an individual could even experience. Still, experiencing ten times more well-being must not be thought necessarily in quality, but rather duration. Consider the following example:

> A company of 36 workers must decide how to allocate vacation days. The company can give every worker 10 paid vacation days each, or due to less overlap of vacation or just higher efficiency, can give 10 workers 40 paid vacation days.

In this situation, the overall amount of paid vacation days would either be 360 days or 400 days. Assuming that all of these workers enjoy their vacation days equally, a classical decision theorist would have to select the distribution where only a third of the workers get any vacation, while the rest get none. Further, assume that all workers get weekends and holidays off anyway and those without vacation days do not get stressed

out or exhausted from work. In this case, a utilitarian must opt for the second distribution, whereas a weak sufficientarian or prioritarian must opt for the first distribution. The distinguishing factor remains caring about the sheer number versus the distribution of well-being. Diminishing marginal returns seems unlikely to explain the utilitarian failure here: why would having one day off result in more happiness the first day than the second? Even so, one could imagine instead that during these days individuals hook up to Robert Nozick's (1974) experience machine and experience maximal well-being all day during vacation. Overall, the severe differences in level of well-being need not be accounted for in terms of quality, but also quantity.

The final potential objection could come in multiple forms, but all relate to the quantity of people in an outcome. One such variant of the conclusion would be Parfit's (1984) *repugnant conclusion*, where he questions whether it would be better to have a world with only a small number incredibly happy people, or a world with many more people who experience lives just barely worth living (p. 388). An objector might question, if a special concern for guaranteeing lives worth living exists, then one should maximize the number of people with net-positive lives. This objection would fail on the basic account that, between choosing between these alternatives, the smaller population of very happy people would universally be better for anyone who enters that society. One might then question: why not kill all the less happy people? In that case, one would have to factor in the fact that you would deprive these people of their lives and worth living. The crux of the matter depends on the fact that people who exist warrant moral concern

while potential people do not. One can question that assumption, but it cannot be tackled at any length in this paper.

Conclusion

Starting with the well-being of sentient beings, the path to derivative moral truths remains muddled. The intuitions of maximizing the good conflicts with intuitions that demand fairness and distribution. Nevertheless, it seems widely apparent that the real goal of consequentialist ethics is to create the best circumstances concerning every single sentient being possible, but weighing interpersonal good divides theories. The use of Rawls's veil of ignorance allows for an easier way to ensure the consequences reflect the common good, but raises difficult questions of rational self interest. Ultimately, the contents of rational judgement hinge on the goals an actor seeks. For sentient beings, this goal must be to have a life worth living. Though prioritarians and utilitarians still have avenues to contest some of this claim, at a minimum, they need to take justifying the distribution of well-being seriously and how it relates to the beings that they value. The utilitarian needs a better reason for risk neutrality, and the prioritarian needs to show why less well-being can universally entitle someone to more. Finally, they need to show why sentient creatures must abandon the idea of attaining a sufficient level of well-being.

As creatures capable of qualitative experience, the nature of sentience determines the value of our existence itself. In choosing to have experience and remain in existence, we need to be able to justify existence to ourselves. We realize that we pursue well-being for its good to us, and that it gives us reason to exist. If the well-being we experience fails to make up for the sufferings, it seems like we should not bother living at all. Thus, when

making self-interested choices that affect our lives over significantly long terms, we cannot avoid the pressing concern of achieving a life that justifies itself; we cannot avoid seeking sufficiency, and the best level of it. This interest compels us sentient beings to make risk averse choices behind the veil of ignorance, in accord of guaranteeing a satisfactory level of well-being, and not risking the loss of these good enough levels for a chance of something better. We would not endure suffering if it was not worth it in the end, and we should not make others bear that burden without exceptionally good reasons.

References

Buchak, L. (2009). Risk Aversion and Rationality. (n.p.) Retrieved from http://web.mit.edu/philosophy/colloquia/buchak.pdf

Buchak, L. (2017). Taking Risks behind the Veil of Ignorance. *Ethics*, *127*(3), 610–644. Retrieved from https://search-ebscohost-com.proxy.binghamton.edu/login.aspx?direct=true&db=pif&AN=EP121940436&site=ehost-live

Harsanyi, J. (1975). Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory. *The American Political Science Review, 69*(2), 594-606. doi:10.2307/1959090

Mill, J. S., & Ebooks Corporation. (2011). *Utilitarianism*. Luton: Andrews UK. Retrieved

from https://search-ebscohost-com.proxy.binghamton.edu/login.aspx?direct=true&db=nlebk&AN=390854&site=ehost-live

Nozick, R. (1974). *Anarchy, State, and Utopia*. New York: Basic Books.

Parfit, D. (1984). *Reasons and Persons*. Oxford: Clarendon.

Parfit, D. (1997). Equality and Priority. *Ratio*, *10*(3), 202. https://doi-org.proxy.binghamton.edu/10.1111/1467-9329.00041

Rawls, J. (1971). *A Theory of Justice : Original Edition* (Vol. Original edition). Cambridge, Mass: Harvard University Press. Retrieved from https://search-ebscohost-com.proxy.binghamton.edu/login.aspx?direct=true&db=nlebk&AN=282760&site=ehost-live

Rawls, J., & Kelly, E. (2001). *Justice as Fairness: a Restatement*. Cambridge, Mass: Harvard University Press.

Jörg Schroth (2008) Distributive Justice and Welfarism in Utilitarianism. *Inquiry*, 51:2, 123-146, DOI: 10.1080/00201740801956812

Sidgwick, H. (1890). *The Methods of Ethics*. London: MacMillan and Co. Retrieved from http://hdl.handle.net/2027/uc1.%24b44192