Decision-Theoretic Consequentialism and the Desire-Luck Problem

Sahar Heydari Fard
University of Cincinnati

Biography
Sahar Heydari Fard is a PhD candidate in philosophy at University of Cincinnati. The focus of her work is on moral epistemology and the morality of social movements.

Publication Details

Citation
Decision-Theoretic Consequentialism and the Desire-Luck Problem

Sahar Heydari Fard

Abstract
Jackson (1991) proposes an interpretation of consequentialism, namely, the Decision Theoretic Consequentialism (DTC), which provides a middle ground between internal and external criteria of rightness inspired by decision theory. According to DTC, a right decision either leads to the best outcomes (external element) or springs from right motivations (internal element). He raises an objection to fully external interpretations, like objective consequentialism (OC), which he claims that DTC can resolve. He argues that those interpretations are either too objective, which prevents them from giving guidance for action, or their guidance leads to wrong and blameworthy actions or decisions. I discuss how the emphasis on blameworthiness in DTC constraints its domain to merely the justification of decisions that relies on rationality to provide a justification criterion for moral decisions. I provide examples that support the possibility of rational but immoral decisions that are at odds with DTC’s prescription for right decisions. Moreover, I argue what I call the desire-luck problem for the external element of justification criterion leads to the same objection for DTC that Jackson raised for OC. Therefore, DTC, although successful in response to some objections, fails to provide a prescription for the right decision.

Keywords
Decision Theory, Consequentialism, Moral Luck, Desires, Emotion, Rationality

Introduction
In his 1991 paper, “Decision-Theoretic Consequentialism and the Nearest and Dearest Objection,” Frank Jackson responds to Bernard William’s objection that there is a tension between consequentialism and some of our fundamental intuitions. Jackson’s main project is to show that a proper understanding of consequentialism, namely decision-theoretic consequentialism (DTC), resolves that seeming tension. He argues against the already existing interpretations of consequentialism by proposing a dilemma: they are either too objective, which prevents them from giving guidance for action, or their guidance leads to wrong actions. An example of such an interpretation is Peter Railton’s 1984 proposal of objective consequentialism (OC), in which the only criterion of rightness is fully external, whether in fact the action maximized utility or not.

Inspired by Thomas Nagel’s 1979 problem of moral luck, which I explain shortly, Jackson moves away from the external criterion of rightness toward an internal criterion that is compatible with consequentialism. Jackson argues that “the fact that an action
might have the best results might be obscure for the agent…. Hence, the fact that a course of action would have the best results is not in itself a guide to action, for a guide to action must in some appropriate sense be present to the agent’s mind” (Jackson 1991, 466). Jackson concludes that a theory of right motive should supplement the criterion of rightness. He states, “I thus am agreeing with Thomas Nagel’s claim that morality requires of us not only certain forms of conduct but also the motives required to produce the conduct” (Jackson 1991, 468). In Jackson’s view, either a right motive or maximization of utility can make an action right, but neither are necessary. He provides an account of right motivation as a combination of beliefs and right desires. Beliefs as the internal element of right motivation are the subjective probability function that determines the agent’s degree of credence for the occurrence of different outcomes. And desires as the external element of right motivation ought to conform to the consequentialist ranking of the alternatives.

Jackson’s proposal, although successful in responding to some objections, leads to similar problems that motivated it. I suggest a distinction between the heuristic and justificatory role of decisions in the discussion of the criterion of rightness. According to this distinction, DTC and OC are on the same page when it comes to the heuristic role of decisions and of the decision-making process. They both consider the determination of the right decision process highly contextual and a matter of empirical evidence. Thus, the advantage of DTC is that it provides a prescription for a justified decision, a decision that, regardless of the outcomes, is not blameworthy. A justified decision, according to DTC, is a rational decision, which is defined by decision theory. However, I provide examples in which a rational decision is morally unjustified and therefore blameworthy. The possibility of such cases crumbles the hope of providing a criterion for moral justification based on a decision-theoretic criterion of rationality.

DTC distinguishes itself from fully internal views about right motivation by including an external criterion of rightness for desires. However, this inclusion makes a new problem that I call the desire-luck problem. According to this problem, it seems justified to morally assess people’s desires when moral assessment is appropriate for matters over which the agent has control, but I discuss studies that show the significant influence of outside factors on our desires in the process of decision making. Therefore, the rightness of a motivation and thereby justification of the related decision can be dependent upon factors over which the agent does not have any control. Thus, DTC is facing a familiar dilemma. If it does not provide any prescription for how to achieve the right desire, its criterion for the rightness of motivation is questioned by the desire-luck problem. But
if DTC provides a prescription for achieving the right desire, then there might be cases in which following those prescriptions leads to unjustified decisions in another context.

**Decision-Theoretic Consequentialism**

Nagel's moral luck problem inspires Jackson's project. The problem of moral luck concerns the contradiction between two moral intuitions. The intuition that “moral assessment is only appropriate for matters over which the agent has control” does not seem compatible with the intuition that “it is sometimes justified to assign praise or blame to things over which agents do not have full control” (Jackson 1991). Jackson reformulates this problem and introduces a new one that I call the prescription problem. If I understand Jackson correctly, the idea is that a moral theory cannot prescribe something that it considers blameworthy. In other words, assuming that wrongness and blameworthiness have to be co-extensional, it is contradictory to prescribe something that might turn out to be wrong. Agents do not always have all the information about what turns out to be right. Thus, a moral theory cannot blame an agent for making a decision that leads to a wrong outcome when the agent did not know what exactly the outcome would be. This is how Jackson explains the problem:

> When we act, we must perforce use what is available to us at the time, not what may be available to us in the future or what is available to someone else, and least of all not what is available to a God-like being who knows everything about what would, will, and did happen. (Jackson 1991, 472)

The prescription problem emphasizes the lack of direct relationship between the morally right action and the morally right decision. We cannot be certain about the outcome of our decision, but this does not imply that we are not responsible for making the right decision. Thus, there should be a criterion of right decision that guarantees that, regardless of the outcome, making a decision based on that criterion does not make the agent blameworthy. Considering this problem, the fact that OC defines rightness only in terms of the external outcome leads to a dilemma. If OC does not provide any prescription for a right decision, then the blame or praise worthiness of a decision seems arbitrary and a matter of moral luck. But, if OC prescribes the right decision,

---

1. This is not an obvious assumption and not everyone agree with it. But it seems necessary to make this assumption to understand Jackson’s argument.
then there might be cases in which those prescriptions lead to worse outcomes that are blameworthy:

Suppose that consequentialism says nothing about the mind of the agent at all. It says merely that right action is action with property $\varphi$, for some consequentialist treatment of $\varphi$ which pertains solely to what in fact would happen and not at all to what the agent thinks. In that case, consequentialism, as Williams puts it, “has to vanish from making any distinctive mark in the world,” by which, I take it, he is, at least in part, making the point we made earlier that consequentialism must say something about right decision. On the other hand, suppose that consequentialism is expressed as a doctrine about how to go about making the morally right decision, as a variety of subjective consequentialism in Railton’s terms, and suppose in particular that it says to think along $\varphi$ lines. What then if thinking along $\varphi$ lines is discovered to have bad consequences in certain situations? (Jackson 1991, 470)

Jackson proposes a new account of right action that is neither fully objective nor fully subjective. In this account, neither outcome nor motivation is necessary to determine the rightness of an action, though they both are sufficient conditions. He states that “What is true is that doing an act for the right reason is sufficient but not necessary for it being what ought to be done in the sense we are insisting is central in ethics” (Jackson 1991, f.n. 20). If one makes a justified decision, a decision based on right motivation, regardless of the outcome, then one has done the right thing. Also, if one makes a decision that leads to the best outcome but was based on evil motivation, then the decision is still right. In a footnote, Jackson states that “on my view, consequentialism does not imply that a morally good intention is essential to a morally good act, at least if morally good act here means what an agent ought to do. It is possible to the right thing for the wrong reason. For an act which maximizes expected moral utility, and it might be that which prompts the agent to action” (Jackson 1991, f.n. 20).

Jackson proposes a theory of right motivation that is based on a combination of right desires and beliefs. Desires are idealized to and defined by a value function that is determined by the objective ranking of the possible outcomes based on their objective utility. According to Jackson, desires “rank the states of affairs in terms of how much the person would like the state of affairs to happen” (Jackson 1991, 464). Beliefs are the subjective probability function, or the degree of credence, that the agent assigns to each
outcome. This function indicates what an agent in fact believes, instead of what an agent ought to believe. Hence, in Jackson’s account, a morally right action is an action that either has the best outcome or is chosen by the right motivation that was also partially determined by the best outcome. This partial determination is due to the contribution of desires to the criterion of right motivation.

Encountering motives in the account of rightness enables Jackson to solve the dilemma; it solves the prescription problem without making consequentialism self-defeating. Jackson argues that the “decision-theoretic account of consequentialism disarms the second horn of the dilemma by answering that in such situations the agent ought not to think along ϕ lines, for the agent’s beliefs will then include that thinking along ϕ lines, in such situations has low expected moral utility” (Jackson 1991, 470). What an agent ought to do is to have desires that rank the alternatives in accordance with the consequentialist value function. Then, the subjective probability function, which is the idealized and “quantitative guise” of the agent’s beliefs, multiplied by the value function, tells the agent what the right decision is. In this process, the agent is maximizing expected utility without consciously aiming for it.

Decision-theoretic consequentialism disarms the second horn of the dilemma by rejecting commitment to the view that maximizing expected moral utility is the right motive for action. The consequentialist value function to which the agent’s desires should conform does not assign any additional value to what maximized expected utility since that would be “double counting” (Jackson 1991, 471). The subjective probability function, which indicates the agent’s beliefs, does not have anything to do with maximizing expected utility either. So, maximizing expected utility cannot be the motive for action. Thus, consequentialism prescribes the right motive for action that guarantees the right decision based on the objective outcomes, regardless of whether in fact the action has the best outcomes or not, and without demanding that the agents have the motivation of maximizing utility or expected utility.

Jackson’s proposal responds to the objection that Michael Stocker (1976) raises for modern ethical views. According to Stocker, considering motivation in the account of rightness of action implies that “a morally good intention is an intention to do the act for the sake of its goodness or rightness” (Jackson 1991, 469). However, according to DTC, “What ought to move a person to action according to consequentialism are desires which may be represented as ranking states of affairs in the consequentialist way, but maximizing expected utility is not a factor in this ranking” (Jackson 1991, 471).

In sum, the rule for action in DTC is to maximize expected moral utility instead of moral utility. The shift from utility to expected utility enables this theory to talk about
right actions in terms of right motivations, which leads to a criterion for right decisions. An action with right motivations can be right even if it does not have the best outcomes. This move obviates any need for commitment to any mental process as long as either the action maximizes utility or the decision is justified. In Jackson’s term, DTC has “built into its very account of right action, a doctrine about right motivation” that “is not committed to any particular view about the mental process that an agent ought to go through in deciding what to do” (Jackson 1991, 468).

An action can be consequentially motivated without any need for consequentialist deliberation. Even when the action does not maximize utility, the agent’s decision is right if it is justifiable. But the justification does not require any mental process. In Jackson’s terms, “sometimes you ought not to go through any mental process at all” (Jackson 1991, 472). For example, Jill has to decide between drug A and B, and she spontaneously chooses drug A without even thinking about it consciously. But it turns out that drug B was the better choice. In this case, Jill’s decision is justified since she knew that it is more likely for drug A to improve the symptoms than drug B. Jackson supports this justification by arguing that “spontaneous action is not action without belief, it is action without conscious reviewing of belief” (Jackson 1991, 472). In sum, if, without any conscious consequential deliberation, Jill’s desires are aligned with the consequential value function, and she applies her beliefs to them rationally, then she is not blameworthy. To determine the justification for her action, however, it does not matter whether she actually applied the beliefs rationally or whether she knew what she desired. What matters is whether it is possible to describe her decision from a third-person perspective and to attribute the right desires to the application of her beliefs in that passive description.

Objections

The main advantage of DTC is the justification of a decision. The prescription problem can be interpreted in two different ways. One interpretation has to do with the way the criterion of rightness guides the agent to in fact achieve the desired outcome. The main concern in this interpretation is to prescribe a decision process that guarantees the best outcome; I call this interpretation the heuristic concern. The other interpretation has to do with how the criterion of rightness can prevent the agent from being blamed. The main concern of this interpretation is the justification of the prescribed action; I call this interpretation the justificatory concern. The justificatory concern is the key to connecting Jackson’s prescription problem to the problem of moral luck. In what follows, I argue that DTC and OC both leave the heuristic concern of the prescription problem
to be resolved by empirical evidence. The main difference between DTC and OC is that DTC attempts to solve the justificatory element of the prescriptive problem while OC does not. However, I argue that DTC is not successful in its attempt since it raises other problems that I discuss shortly.

In his paper, “Alienation Consequentialism and the Demands of Morality,” Peter Railton addresses Jackson’s prescription problem in its heuristic sense. He considers the objection that the “lack of any direct link between objective consequences and a particular mode of decision making leaves the view too vague to provide adequate guidance in practice” (Railton 1984, 116). Railton’s solution for this problem, which leaves Jackson unsatisfied, is that “objective consequentialism sets a definite and distinctive criterion of right action, and it becomes an empirical question… which modes of decision making should be employed and when” (Railton 1984, 116). Railton’s response, similar to DTC, leaves the decision process a matter of empirical question. OC and DTC agree that the decision process that in fact maximizes utility or expected utility is highly context dependent and should be a matter of empirical evidence. If lack of deliberation is the method that achieves the desired outcomes, neither OC nor DTC demands conscious thinking.

Rationality, in its narrow sense, is the main idea behind the criterion for justification of a decision for DTC. DTC ends up with a set of objective requirements that if an agent satisfies, she will not be blameworthy. These requirements are mainly derived from the idea that the agent’s decision should be rational. Thus, regardless of how a person in fact came to a decision, as long as it is possible to interpret her decision process in terms of maximizing expected utility and as rational, the decision is justified and not blameworthy. As discussed in the previous section, DTC is not committed to any specific mental process; indeed, it does not require any sort of mental deliberation to make an action justified. This lack of commitment to any mental process is possible because the justification criterion is not based on what the agent in fact does but rather on an after-the-fact description of the action.

The rationality criterion proposed by DTC does not accommodate the justificatory concern. In other words, it is possible to act in accordance with the criterion for right actions in DTC and still be morally blameworthy. In example 1, I provide a case in which intuitively the agent seems blameworthy, and morally unjustified, while her decision seems rationally justified.

Example 1: Jill is about to leave work when she figures her expensive watch has been stolen. She knows that the watch was on her desk all day and that no one entered or left her office after she got there. Six people are in the room, and she does not know
any of them personally. However, one of the people in the room is African American. She knows that African Americans commit 60% of the total crimes in her country. Thus, if we take Jackson’s criterion seriously, regardless of whether she is right or not, she is morally justified to believe that there is a high chance that the African American person in the room is to blame. However, the rational justification of this conclusion does not make it morally justified. There is a strong intuition² that she is in fact blameworthy if she thinks the person who belongs to a group that commits more crime in society has more likelihood of being the one who stole her watch.

Example 1 is one of many possible examples that show that rationality does not provide moral justification for action. Jill did not act in accordance with her conclusion, so the outcomes are irrelevant to her blameworthiness. She made a rational decision, and she has the right desires that want to find the criminal. Thus, her motivation is right. But she is still intuitively blameworthy. This problem can be traced back to the narrow definition of rationality in decision theory. The formal model of rationality that is introduced by decision theory is notorious for over-simplification and/or over-rationalization of human behavior (Sobel 1994).

Various modifications of the prisoner’s dilemma are the standard counterexamples against the narrow definition of rationality. This famous example in game theory is used to show that a rational decision is not the best decision in terms of maximizing overall utility (Joyce 2007; Hitchcock 2016). The prisoner’s dilemma is a standard case with usually two participants who each need to decide between two alternatives. However, the decision of the other participant partially determines the outcome of the participant who is deciding. If the agent acts in accordance with the decision-theoretic criterion of rationality, the outcome will be worse overall. A rational decision, in this case, is to “play safe” and not assume that the other agent will collaborate. However, this decision guarantees a worse outcome. The “irrational” decision is to assume the collaboration of the other agent, and it achieves the best outcome overall. In sum, if best consequences for everyone is concerned, in the prisoners’ dilemma it seems justified for each individual to not act “rationally.”

² This intuition is due to her action falling into the category of discrimination, which “is prohibited by six of the core international human rights documents. The vast majority of the world’s states have constitutional or statutory provisions outlawing discrimination (Osin and Porat 2005). And most philosophical, political, and legal discussions of discrimination proceed on the premise that discrimination is morally wrong and, in a wide range of cases, ought to be legally prohibited” (Altman, 2016).
DTC has an inevitable problem that I call the desire-luck problem. DTC pushes the moral-luck—and thereby prescription—problem one step back to the desires, which leads to problems similar to those that motivated its proposal. The formulation of the desire-luck problem is as follows. Our moral assessment seems appropriate for matters over which an agent has control. We do not have full control over our desires, but according to DTC, we are justified in assessing people morally based on their desires. Jackson argues that the desire that the agent ought to have is the one that conforms to the consequentialist value function. However, there is no prescription for how agents should acquire such desires, and the psychological studies that I discuss shortly suggest that it is not always possible to have full control over our desires.

The desire-luck problem causes a dilemma for DTC in the same way that standard consequentialism was subject to a dilemma. If DTC does not prescribe how we can get to the right desires, it needs to deal with cases in which the desires are affected unbeknown to the agent. Instances of implicit bias and situational bias, like the bystander effect, show how vulnerable are our desires to the effect of things that we do not have much control over. In fact, there might be no way for us to realize that those affects exist without professional help. On the other hand, if DTC assumes control over desires and prescribes a method, this could be self-defeating. It may turn out that such a method in fact leads to desires that will not conform to the consequentialist value function in other contexts.

We do not have full control over what affects our desires, so the fact that they conform to the consequentialist value function may be due to pure luck. Many studies suggest that our ranking of alternatives is affected by things that we are not aware of and that we do not have full access to how they affect our judgment. There are studies that suggest that our desire to help a person, or our judgment about how serious her situation is, might be affected by how much of a hurry we are in. Studies about the bystander effect also suggest that whether we rank helping a victim high among the alternatives is significantly affected by whether other people are willing to help the victim or not (Asch 1951; Carlson et al. 1988; Rodin 1969). The common feature of these studies is that their participants show significantly less desire to help in certain mental states or in a situational context compared to normal contexts and that they are unaware of this difference. However, none of these uncontrollable effects on desires prevents us from blaming someone who doesn’t help a victim because she is late to work. Therefore, lack of access to what changes our desire makes having the right set of desires in a particular situation a matter of luck.

DTC can be self-defeating if it provides guidance for how to adopt the right desires. A person might do everything that maximizes expected utility, but in doing so,
unbeknown to her, her desires could be affected to the point that they do not conform to the consequentialist value function anymore. In what follows, I provide an example to make this point. Then I explain why DTC needs to talk about desires in an intuitive sense and why defining desires in terms of emotions seems like a natural option. I use an account of learning for emotions that seems more compatible with decision theory, and finally, I discuss why the provided example leads to a contradiction for DTC. The following example is a case in which desires are adjusted to maximize expected utility in one context, but the permanent change in desires makes the agent blameworthy in other contexts.

Example 2: Jalisa is a good nurse, but in order to do her job and stay sane, she has, over time, lost her sensitivity to people’s pain. She does not prioritize someone’s pain over answering a phone call or over something that she can do to help, for instance, stitching someone’s wound and not be emotionally distressed by the patient’s pain like a normal person. This manner is in fact motivated by all professionals in the community since it helps them avoid the effects of emotional distress in their decision-making process. However, her insensitivity to pain that came from years of working as a good nurse hurts her new partner’s feelings. Jalisa’s reaction to her partner’s pain is far from what it needs to be.

Jackson’s project can be summarized in three major moves, but for them to be valid, formal desires must be connected to an intuitive understanding of desires. His notion of desire is simply to formal and abstract to be what intuitively we expect desires to be. But, the plausible option for this connection is to use emotions to make the connection. In his first move, Jackson argues for the importance of motives for a moral theory and for justification of a decision. In the second move, he uses the intuition that motivation is composed of emotions and feelings. Finally, in his third move, he uses decision theory, which describes decisions in terms of a subjective element that he calls beliefs and an objective element that he calls desires. However, to transition from the second to third move, Jackson needs to show that his definitions of desires and beliefs are close to what intuitively composes motives. The formal notion of desires, common in linguistic and decision-theoretic descriptions of desire-like mental states, is usually understood as having a close relationship with emotions. The common view about the role of emotions among philosophers is that “Emotions make certain features of situations or arguments more prominent, giving them a weight in our experience that they would have lacked in the absence of emotion” (de Sousa 2014). This view about the role of emotions has important similarities with the role of the preference function in decision theory that
Jackson defines as desire. Moreover, other considerations about the nature and the role of emotions and desires makes an appeal to understanding desires in terms of emotions. 3

Understanding desires in terms of emotions enables us to talk about how an agent can have control over what she desires. Emotions are learned by association with “paradigm scenarios.” A type of a situation and a set of characteristics of expected responses to the situation are two elements of a “paradigm scenario” (de Sousa 2014). A complex and controversial mix of biological and cultural factors determines what the expected responses are in each type of situation (de Sousa 2014). The process of learning, however, happens through time by associating the proper responses to each paradigm scenario. For some more fundamental emotions, the associations are “drawn first from our daily life as small children and later reinforced by the stories, art, and culture to which we are exposed. Later still, they are supplemented and refined by literature and other art forms capable of expanding the range of one’s imagination of ways to live” (de Sousa 2014).

Jalisa is blameworthy in the context of her relationship, but her blameworthiness is a natural consequence of following the prescriptions in another context. In Jalisa’s example, the paradigm scenario is a type in which someone is in pain and needs help. In the context of the hospital that she works in, it is expected that she not feel any emotional distress and not let the patient’s pain change her desires or her preference function. But in the context of her new relationship, Jalisa needs to be sensitive to her partner’s pain. The right desire in this context for Jalisa is to prioritize her partner’s pain. But when Jalisa successfully learns to have the appropriate desire in one paradigm scenario, it is not possible for her to immediately have the right desire in a novel context, namely in her relationship, that triggered the same paradigm. Jalisa’s motivation and therefore her decision in response to her partner is wrong and blameworthy since she does not have the right desire or the right preference function. Still, her blameworthiness for her

3. The following is de Sousa’s explanation for why this account of emotions makes it appealing to understand desires in the way that Jackson talks about in terms of emotions: “This account does not identify emotions with judgments or desires, but it does explain why cognitivist theorists have been tempted to make this identification. Emotions set the agenda for beliefs and desires: one might say that they ask the questions that judgment answers with beliefs and evaluate the prospects that may or may not arouse desire. As every committee chairperson knows, questions have much to do with the determination of answers: the rest can be left up to the facts. In this way emotions could be said to be judgments, in the sense that they are what we see the world ‘in terms of.’ But they need not consist in articulated propositions. Much the same reasons motivate their assimilation to desire. As long as we presuppose some basic or preexisting desires, the directive power of ‘motivation’ belongs to what controls attention, salience, and inference strategies preferred” (de Sousa, 2014).
decision and her desire is due to the DTC prescription for right motivation and to her normal response to learning the right desire in a paradigm scenario.

In sum, Jackson uses the DTC to solve the prescription problem as the main advantage of this theory over OC. However, a normative theory needs to address our moral intuitions. I chose an example of racial discrimination to motivate the intuition that moral reasoning might not be easily captured by a general claim about how we should/we do make decisions. The details about the context matters in a way that is not easily captured by decision theory. Nurses are the paradigm example of a morally good person with consequentially praiseworthy contributions. I used an example of a nurse to show that right desires cannot make up for our expectations of right motivation. If being a nurse is a good, and being a better nurse is even better, it is praiseworthy that Jill wants to be a better nurse. But DTC considers a nurse blameworthy in some contexts although it prescribes being a good nurse.

References


