Decision Theory and Virtue Ethics: the benefits of formalization

Tony Skoviak

Department of Philosophy

Central European University

In partial fulfillment of the requirements for the degree of Masters of

Arts

Supervisor: Simon Rippon

Budapest, Hungary

2016

Abstract

The application of rational decision theory to virtue ethics has received little attention from philosophers, despite the popularity of rational decision theory among ethicists in general. In this thesis I will perform an analysis of the core components of virtue ethics—virtue and practical wisdom—using the tools of decision theory and Bayesians norms of rationality. From this analysis, I conclude that, (1) if we believe virtue, phronesis, and Eudaimonia are useful concepts for discussing ethics, then we have good reason to believe a Bayesian decision theory will benefit our ethical theorizing; and (2) we have reason to think there is no fundamental incompatibility between virtue ethics, decision theory, and Bayesian norms.

CONTENTS

Abbreviations and Symbolsii
List of Rationality Requirementsiii
Introduction 1
Chapter one: Formalization, Bayesian decision theory, and virtue ethics
1.1 Formalization
1.2 Bayesian decision theory
1.3 Virtue Ethics
Chapter two: How does decision theory contribute to ethics simpliciter, and how could it contribute to virtue ethics?
2.1 How does decision theory contribute to ethics simpliciter?
2.2 How can Bayesian decision theory contribute to, or enhance, theories of virtue ethics?
Conceptual clarification of phronesis and virtue
On Ramsey, useful habits, and intellectual virtue
Credence and Subjective Bayesianism
Chapter three: Is it possible to have a bayesian decision theory representing virtue?
Rationality requirements and virtue
Degrees of desire and virtue
Conclusion
Bibliography

Abbreviations and Symbols

cr(), represents a credence function.

 $cr_b(x) = 0.6$, (i.e. 60%) would be read "Agent b has a credence of 0.6 in x."

 $cr(x \mid y)$ represents the credence that 'x' is true when supposing 'y' is true. This is often called a conditional credence, opposed to an *un*conditional credence such as cr(x) or cr(y).¹

pr(), represents a probability function.

pr(x) = 0.5, indicates that x is assigned a probability of 50%.

u(), represents a utility function.

 $u_b(x)=10$ would read "Agent b assigns 10 units of value to x." Unspecified units of value are sometimes called 'utils'.

EU() represents expected (weighted) utility.

EU(x)=u(x)*cr(x)

Preference symbols:

~ for indifference 'x ~ y' means that the agent does not prefer 'x' to 'y', nor 'y' to 'x'

> for strict preference 'x > y' means that the agent prefers 'x' to 'y'

 \gtrsim for weak preference 'x \gtrsim y' means that the agent either prefers 'x' to 'y' or is indifferent between 'x' and 'y'

¹ One might wonder if a conditional credence is simply an unconditional credence in a conditional proposition. I will not address the issue here, but for a further discussion see Alan Hajek, "What Conditional Probability Could Not Be," *Synthese* 137, no. 3 (2003).

List of Rationality Requirements

x, y, z = variable propositions
X = the set of propositions in our language
T = logical truth/tautology
cr(x)=y means that an agent's credence in x is y
u(x)=y means that an agent assigns a value of y to x

On Credence Distributions

Kolmogorov's Probability Axioms:

Non-negativity $(\forall x \in X)(cr(x) \ge 0)$ A rational agent will have a credence of at least 0 in all propositions in their credence distribution.

Normalization $(\forall x \in X)(x \text{ is a tautology} \Rightarrow cr(x) = 1)$ A rational agent will have a credence of 1 in all tautologies in their credence distribution.

Finite Additivity $(\forall x, y \in X)((x \lor y) \Rightarrow [cr(x \lor y) = cr(x)+cr(y)])$

A rational agent's credence in the disjunction of any pair of mutually exclusive propositions will be equal to the sum of their credence in each disjunct.

e.g. cr(it will rain tomorrow or it will not rain tomorrow) = cr(it will rain tomorrow) + cr(it will not rain tomorrow)

Other Axioms

Ratio Formula $(\forall x, y \in X)(cr(y) > 0 \Rightarrow [cr(x | y) = cr(x \& y)/cr(y)])$ A rational agent's credence in proposition 'x', when supposing proposition 'y' is true, will be equal to the quotient of their credence in 'x & y' and their credence in 'y'.

Conditionalization $[(cr(x|y) = p \text{ at } t_1) \& (cr(y) = 1 \text{ at } t_2)] \Rightarrow (cr(x) = p \text{ at } t_2)$ If a rational agent has a credence of p in proposition 'x' when supposing proposition 'y' at time 1, and learns 'y' at time 2, then they will have a credence of p in 'x' at time 2. e.g. If cr(it will rain tonight given that there are dark clouds on the horizon) = 0.9 on Sunday morning, and then the agent learns that there are dark clouds on the horizon Sunday afternoon, after learning this their cr(it will rain tonight) will equal 0.9.

On the Preference Relation

Value Maximization:

Strict Value Maximization $(\forall x, y \in X)(u(x) > u(y) \Rightarrow x > y)$ If a rational agent values 'x' more than 'y', then they will prefer 'x' to 'y'.

Weak Value Maximization $(\forall x, y \in X)(u(x) \ge u(y) \Rightarrow x \ge y)$ If a rational agent values 'x' at least as much as 'y', then they will weakly prefer 'x' to 'y'

von Neumann-Morgenstern Utility Axioms:

Completeness $(\forall x, y \in X)(x \gtrsim y \lor y \gtrsim x)$

For any 'x' and 'y' in a rational person's preference distribution, either 'x' is preferred to 'y' or 'y' is preferred to 'x'.

Transitivity $(\forall x, y, z \in X)([x \gtrsim y \& y \gtrsim z] \Rightarrow x \gtrsim z)$ If a rational agent prefers 'x' to 'y' and 'y' to 'z', then they will prefer 'x' to 'z'.

Continuity

 $(\forall x, y, z \in X) [(x \gtrsim y \gtrsim z) \Rightarrow (\exists \alpha \in (0, 1))(y \gtrsim \alpha x + (1 - \alpha)z)]$ $(\forall x, y, z \in X) [(x \gtrsim y \gtrsim z) \Rightarrow (\exists \beta \in (0, 1))(\beta x + (1 - \beta)z) \gtrsim y]$

For any set of preferences like $x \ge y \ge z$, there will be some gamble involving 'x' and 'z' (x if e_1 , z if e_2) such that 'y' is preferred to it. For example, if you prefer driving to work rather than biking and biking rather than walking, continuity states: there exists some α such that you will prefer biking to work rather than than an α % of driving to work or a (1- α)% chance of walking to work. e.g. if your car breaks down you will have to walk, and you estimate a 95% chance that your car will break down, so you prefer to simply take your bike. The same is true for a gamble involving 'x' and 'z' being preferred to 'y'.

Strong Independence $(A \gtrsim A^*) \Rightarrow (A \text{ IF } e_1, B \text{ IF } e_2) \text{ to } (A^* \text{ IF } e_1, B \text{ IF } e_2)$ If a rational agent prefers A to A*, then they will prefer a gamble between A or B to a gamble between A* or B, if the chance of recieving A is the same as the chance of recieving A*. e.g. If you prefer hot dogs to hamburgers and someone offers you a choice between the following two gambles conditional on a coin flip: hot dog if heads, salad if tails; or hamburger if heads, salad if tails; then you ought to prefer the first gamble.

Other Axioms

Reflexivity $(\forall x \in X)(x \gtrsim x)$

A rational agent will prefer any proposition at least as much as itself.

 $\begin{array}{l} \textit{Probabilistic (Credential) Equivalence} \quad [cr(e_n) = cr(f_n) \text{ for } n = 1 \ \dots \ N] \Rightarrow [(A_1 \text{ IF } e_1 \ \dots \ A_N \text{ IF } e_N) \sim (A_1 \text{ IF } f_1 \ \dots \ A_N \text{ IF } f_N)] \end{array}$

A rational agent will be indifferent between two lotteries if they have the same prizes and the same odds of receiving each prize.

The "Sure-Thing" Principle $(A_n \gtrsim A_n^* \text{ for } n = 1 \dots N) \Rightarrow (A_1 \text{ IF } e_1 \dots A_N \text{ IF } e_N) \gtrsim (A_1^* \text{ IF } e_1 \dots A_n^* \text{ IF } e_N)$

A rational agent will prefer lottery A to lottery A* if lottery A has the same odds of yielding prizes that are always better.

INTRODUCTION

Much of the literature on the formalization of moral theories is recent. This contemporary surge is at least in part due to the rise in popularity of *rational decision theory* as a model of the degrees of belief and desire of rational agents, paired with the idea, right or wrong, that good moral agents ought to be rational.² Despite this surge, the application of decision theory to virtue ethics has received little attention from philosophers.³ In this thesis I will (1) argue that if we believe virtue, phronesis, and Eudaimonia are useful concepts for discussing ethics, then we have good reason to believe a Bayesian decision theory will benefit our ethical theorizing; and (2) offer reasons to think that there is no fundamental incompatibility between virtue ethics, decision theory, and Bayesian norms.

In the first chapter of this thesis I will define some criteria for a formal theory of ethical decision making. I will then detail the decision theory, rationality requirements, and the virtue theory that I plan to use in order to make my argument. In the second chapter I will provide reasons why decision theory is a useful tool for ethicists in general, and then argue that a mutual benefit can arise from a collaboration between decision theory and virtue ethics. The second argument will focus on the ways in which Bayesian decision theory can assist in clarifying or defining virtue theoretic concepts and articulating the requirements of virtue and phronesis. In the third chapter I will provide some cursory suggestions for overcoming apparent incompatibilities

² Note that I use the term 'belief' to refer to what is sometimes call "binary belief"—a classificatory attitude which either is or is not attributed to an agen; either I believe something or I do not. When I need to use a broad term encompassing the various attitudes agents can take toward propositions (belief, certainty, credence, etc.), I will use the term 'doxastic attitude'. Credit goes to Titelbaum (forthcoming) for emphasizing this distinction. For a discussion of the push in ethics to require good moral agents to be rational see Jan Narveson, "The Relevance of Decision Theory to Ethical Theory," *Ethical Theory & Moral Practice* 13 (5).

³ One example is Colyvan, Cox, and Steele (2010), while Bastons (2008) and Morales-Sanchez & Cabello-Medina (2013) both discuss the benefit of virtues for decision theory, but not vice versa.

between virtue frameworks, decision theoretic frameworks, and Bayesian rationality requirements.

CHAPTER ONE

FORMALIZATION, BAYESIAN DECISION THEORY, AND VIRTUE ETHICS

1.1 Formalization

The primary criterion I set out for successful formalization is:

(F1) To use a formal model to represent the decisions an ideal moral agent would make.

Two further criteria for successful representation are:

- (F2) The model should only yield decisions that the moral theory deems permissible.
- (F3) The model should *never* yield a less valuable decision over a more valuable decision.⁴

The second and third criteria only apply to coherent moral theories. If two or three are violated, it could be that the model accurately depicts the theory, but the theory itself is (perhaps in some arcane way) incoherent. One important methodological note: I am not interested in a formal model as explanatory or descriptive of the decisions of *real* moral agents (as in psychology or economics); a successful moral model will assess the utility values of an *ideal* moral agent according to a specific theory. This goes along with the distinction sometimes made between *empirical* decision theory and *rational* decision theory.⁵ The primary difference being that empirical decision theory is descriptive, while rational decision theory is normative.

⁴ This is because we are modeling the theory by proxy of a moral perfectionist who simply values what the theory argues is *good*. A moral perfectionist will not consider any non-moral factors. So, even if it is *permissible* on a given theory for a perfect agent to decide on something other than the best (because of non-moral factors), the perfect agent will never do so.

⁵ e.g. Christoph Lumer, "Introduction: The Relevance of Rational Decision Theory for Ethics," *Ethical Theory and Moral Practice* 13 (2010): 485.

I take (F1) to be the primary criterion because I take ethics to be mainly concerned with making 'ought' claims. Specifically, to give moral agents advice on how they ought to behave, how they ought to act, and what they ought to value. Such claims, I think, are inextricably tied up with the decisions that moral agents make, and as such a formal theory of the 'ought' relation will necessarily be a description of how agents interact with themselves and the world, i.e. which decisions they will be led to make by the things they ought to embody, do, and value. I take the second and third criteria to be apparent; if we are modelling the decisions that an ideal moral agent would make, we should not accept decisions that even a non-ideal good moral agent would find impermissible (F2), nor decisions that are less valuable than others according to what is valued by the moral theory (F3).

Note that if we accept such criteria as reasonable, the second and third highlight some formal concerns pertinent to deontological and utilitarian ethics. Namely, because of (F2) a pure, non-intentional deontological model could never recommend a decision involving even an infinitesimally small chance of a prohibited act. This can be skirted slightly by making rules conditional on intention, but then a deeper issue arises; if one has a rule against intentionally killing and innocent or lying, and one is faced with a decision that could, with some small probability, result in killing an innocent or the telling of a falsehood, then *if* the small probability obtains, was the killing of the innocent or the telling of this falsehood intentional? After all, the decision maker *knew* what they were doing *could* result in these things.

The deontologist can respond: such rules only apply under certainty conditions; it is wrong to lie *if and only if* you know you are lying, it is wrong to kill an innocent *if and only if* you know you will kill an innocent. But then, what ought we to do if we are not certain? What of the murderer who is *pretty sure* that his decision will terminate in the death of an innocent, but is not certain? Has he broken a rule? The unconfident murderer can be handled by shifting the

focus even more firmly onto intentions—that he intended to kill an innocent is enough to have broken the rule against (intending to) kill an innocent—but this does nothing to assuage the concerns of the morally conscious agent who wants to know: Should I try to save these trapped miners, even though there is a chance that my attempting to save them could result in their deaths? She does not *want* the miners to die, but she *intends* to take an action that could result in either their safety or their doom. If we shift the focus of our rules onto the *wants* of our agents, then we run into a whole different set of problems.

One might wish to argue that the murderer and the rescuer clearly have different intentions, even in risky scenarios, but I use intention here to mean: the prospect that is intended to be chosen. Suppose Adam wants to kill John, and Jane wants to save John. If Adam and Jane choose different prospects (e.g. Jane chooses a 90% chance of saving, 10% chance of killing, while Adam chooses a 90% chance of killing, 10% chance of saving), their choice is clearly different. But, the concern is that we can have a would-be-savior and a would-be-murderer who both intend to make the same decision (e.g. the probability of killing/saving is unknown when no action is taken, but if action is taken there is a 90% chance of saving/10% chance of killing).⁶

Adam, really disliking John, might think a 10% chance of killing his target is enough to justify this prospect, while Jane might think a 90% chance of saving her target does the same for her. Their interpretation of the unknown probability is what leads them to their different justifications; Adam thinks if he does nothing there is less than a 10% chance that John will die, while Jane thinks if she does nothing there is more than a 10% chance that John will die. They see the opposing prospect as different, but the prospect they are actually choosing as the

⁶ For a deeper discussion see Richard Holton "Partia Belief, Partial Intention," where he introduces the concept of partial intention. With our case of murder, Holton would say Adam's intention to make such a decision is partial if it is part of a larger plan to bring about John's death. That is, if 'x' does not work, then he will try 'y', and if that does not work, he will try 'z', and so on, all aimed at trying to kill John. i.e. Adam has a partial intention to kill John if his decision is part of a larger plan to kill John. Making rules against "larger plans" of this sort might be one path a deontologist could take.

same. It is possible to say: a rule against killing means, "When deciding among multiple prospects, you ought to choose the one with the lowest probability of killing someone" in order to capture the difference between Adam and Jane. This, however, threatens to turn a deontological theory into nothing more than a complex set of decision rules without any true prohibitions or obligation. To say that Adam would like the death outcome and Jane would like the living outcome is, with the vocabulary I am using, to differentiate their actions by their preferred outcome—what they would like to happen—, rather than by the actual decision they are making.

On the other hand, (F3) highlights a concern for a simple ethical consequentialist. Brad Hooker gives an example of a simple act-consequentialist criterion of right: "[A]n act is morally permissible if and only if the actual (or expected) overall value of *that particular act* would be at least as great as that of any other act open to the agent."⁷ Let us ignore for a moment the "or expected" caveat and assume permissibility is determined by actual consequences. If a consequentialist agent is facing a pair of risky decisions, each with a chance of either increasing or decreasing total value, it is impossible at the moment of decision for the utilitarian to know which decision is right.⁸ A standard decision model dealing with risk will recommend making the decision with the highest probability of increasing value, the lowest probability of decreasing value, the highest amount of expected value, or something along these lines. However, with a pair of risky prospects, any of these recommendations could end up being worse in terms of total value than the alternative. Thus, unless the agent is working under

⁷ Brad Hooker, "Act-consequentialism" in *Ideal Code, Real World*. (Oxford: Oxford University Press, 2000), 142.

⁸ For an overview of thinkers with a similar complaint see Sven Ove Hansson, "The Harmful Influence of Decision Theory on Ethics," *Ethical Theory & Moral Practice* 13 (2010): 589.

certainty conditions, without a decision procedure separate from this simple criterion of right, the model could easily yield an inferior moral decision to a superior one.⁹

If we pay attention to the caveat, we might shift our criterion of right to use *expected* value. If we do this instead of coming up with a new decision procedure, then we are saying "what makes something *right* is the expected value of it." But why is it valuable in the first place? Standard justifications of the value will not apply, because the *expected* value is what makes decisions right. To argue that expected value is right because the value in question simpliciter is right is to provide a criterion of right for our criterion of right. Providing an account that *expected* value is *simpliciter* right is highly counter-intuitive, and perhaps impossible; how can an expectation of something be valuable if it is not because that thing itself is valuable? In other words, it seems like we have simply pushed the real criterion of right into the background and split our decision procedure into two parts, both concerned with our best ways of bringing about what we value *in practice*. It is tantamount to saying: what is right is to function such that we make the best decisions given our values. Such a position is viable, but it is an adaptation to uncertainty, and blurs the lines between a criterion of right and a decision procedure.

Then, what if we keep our simple, non-expected, criterion of right and come up with a new decision procedure that uses expected value? Brad Hooker argues that no consequentialist should believe that decisions ought to be made by evaluating which decision has the highest expected value.¹⁰ He suggests, instead, that all consequentialist decision procedures should make use of general rules. The first two reasons he gives for this are of particular interest to us. Those are: we often lack probabilistic information, and we often lack the time to procure

⁹ Thanks to Simon Rippon for suggesting that the distinction between a decision procedure and a criterion of right would be useful here.

¹⁰ Hooker, "Act-consequentialism," 142-143.

probabilistic information, even when it is within our power. There are generally taken to be three types of decision spaces: certainty, risk, and uncertainty.¹¹ If we know what will happen given any choice we make, then we are making a decision under *certainty*. If some of our choices have probabilistically well-defined possible outcomes (e.g. a coin flip), then we are making a decision under *risk*. If we do not know all of the relevant possible outcomes, or we do not know how likely some known possible outcomes are, then we are making a decision under *uncertainty*. Decisions made under risk trouble an act-consequentialist without a concept of expected value (either in their decision procedure or their criterion of right), and decisions made under uncertainty trouble consequentialist decision procedures in general.

I highlight these concerns to show that deontological and consequentialist theories also experience difficulty with formalization under decision theory, and so such an initial resistance from virtue theory should not dissuade us from trying. The specific concern with deontological and consequentialist theories was that a theory of decision does not simply fall out of a theory of right under risk and uncertainty. This is not directly a problem for virtue theory; their theory of right is primarily dependent upon the nature of the agent's motivation and the extent of their virtuous character, and so uncertainty about consequences does not, to the same extent, require an adapted theory of decisions. The more general concern for all three theories is that they face some difficulty with formalization; in deontology and consequentialism it is with uncertainty, and with virtue ethics the worry is with certainty. Deontology and consequentialist theories have well-defined rules, while the advice virtue ethics gives does not seem, at first glance, to be so ordered. Perhaps it is natural, then, that the former types of theories work well under certainty (where it is easy to apply rules), while the latter works well under uncertainty (where it is difficult to apply rules which refer to agent-neutral value). Part of the goal of this thesis is

¹¹ e.g. see John Harsanyi, "Bayesian decision theory and utilitarian ethics," *American Economic Review, Papers and Proceedings* 68 (1978): 223.

to impose some formal structure upon virtue ethics such that it functions more clearly in certainty conditions, while still retaining its freedom to direct us when we, as we often do, lack sufficient knowledge to enact our rules.

1.2 Bayesian decision theory

The formal model I intend to use is decision theory paired with Bayesian rationality requirements. There are roughly two key features of decision theory: prospects and preferences. Prospects are the options under consideration in any given decision problem, and preference is a relational attitude between prospects. Weak preference (\gtrsim) means x is at least as preferred as y (x \gtrsim y). Being at least as preferred as means: either x is strictly preferred to y (x>y) or the agent is indifferent (~) between x and y (x~y). Strong preference (\succ) means x is strictly preferred to y (x>y). In this thesis, when I talk of preference I will mean in the weak sense. The Bayesian rationality requirements are primarily about constraining two things: the preference relation and rational credence distributions.

The most basic requirement imposed on the preference relation is that rational agents are value maximizing.¹² Value maximization is achieved through preferring (and making) the decisions with the highest expected value. Expected value is usually calculated with degrees of desire

¹² For examples of this see the list of Bayesian commitment in Bradley (2001) page 263 and Hansson (1994) page 38, and the better prizes lottery axiom in Dreier (2004) page 3. Harsanyi (1978) introduces value maximization as the first theorem of Bayesian rationality, referring to a proof by Gerard Debreu from a complete preorder and continuity to value maximization. Ramsey (1931) claims value maximization is a law of psychology, and Titelbaum (forthcoming) gives this assumption a much deeper discussion throughout his book. Joyce (2004) takes value maximization to be part of the expected utility thesis, which is in turn part of an argument for probabilistic consistency among beliefs. Here I will treat value maximization as a basic requirement, but even if we frame the problem like Joyce—if probabilistic consistency is seen to be a fundamental requirement for rational agents and value maximization is the reason *why* this is true (of rational agents), then value maximization is necessarily as fundamental as probabilistic consistency among beliefs, if not more so.

and degrees of belief. Degrees of desire—often represented as a utility function—most fundamentally represent the *values* of the agent. Note that by value I mean normative value, not mathematical value.¹³ Degrees of belief—often portrayed as some sort of probability function—as the name suggests, represent the strength of an agent's confidence in the truth of a proposition. Expected value gives us a picture of *how* rational agents ought to begin forming their preference set. The von Neumann-Morgenstern axioms: transitivity, reflexivity, completeness, and continuity—are four further structural requirements on the preference relation.¹⁴ David McCarthy also argues that to have an expected utility theory, we need strong independence, i.e. if an agent holds $A \gtrsim A^*$, then they should also hold (A IF e₁, B IF e₂) \gtrsim (A* IF e₁, B IF e₂).¹⁵ For example, if an agent prefers receiving books written by Plato rather than Aristotle, they ought to prefer *being given a book by Plato if heads, or a poem by Sappho if tails*, to *being given a book by Aristotle if heads, or a poem by Sappho if tails*.

Credence is a term standing for a "degree of belief" or "degree of confidence", sometimes called a "partial belief" or "graded belief" in opposition to "binary beliefs".¹⁶ If I am not certain that it will rain tomorrow, but I think there is some chance that it will, then I have a degree of belief in the proposition "It will rain tomorrow." In other words, rather than a binary/classificatory doxastic attitude (e.g. either I am certain that it will rain or I am not

¹³ For an overview of the Bayesian arguments in favor of value maximization *as a normative theory* of rationality (specifically the Dutch book argument and theories of representation), see James Joyce, "Bayesianism," in *The Oxford Handbook of Rationality*, (Oxford: Oxford University Press, 2004), 5.

¹⁴ For a utility theory without completeness see Robert Aumann, "Utility Theory Without the Completeness Axiom," *Econometric Research Program: Research Memorandum No.* 26 (1961). For a utility theory without continuity see Melvin Hausner, "Multidimensional Utilities," in *Decision Processes*, ed. R. M. Thrall, C. H. Coombs, and R. L. Davis, (New York: John Wiley and Sons, 1954). For a discussion of the necessity of transitivity in decision theory see Bernd Schauenberg, Roger Dunbar and Rudi Bresser, "The Role of Transitivity in Decision Theory," *International Studies of Management & Organization* 11, no. 1 (1981). See the list at the beginning of this thesis for formal definitions of these and any other requirements that are mentioned but not defined in the text.

¹⁵ David McCarthy, "Probability in ethics," in Alan Hájek & Christopher Hitchcock (eds.), *The Oxford Handbook of Philosophy and Probability*, (Oxford: Oxford University Press, forthcoming).
 ¹⁶ Michael G. Titelbaum, "Beliefs and Degrees of Belief" in *Fundamentals of Bayesian Epistemology*, (under

¹⁰ Michael G. Titelbaum, "Beliefs and Degrees of Belief" in *Fundamentals of Bayesian Epistemology*, (unc contract with Oxford University Press), 15.

certain, there is no "partial certainty"), it is a comparative (e.g. I am more confident that it will rain than not) and/or quantitative (I am 70% confident that it will rain) doxastic attitude. For the purpose of this thesis, I will assume that credences can be meaningfully represented numerically (that they are quantitative doxastic attitudes), and when I say "credence" I will mean "numerical credence" unless I say otherwise.

The most basic requirement on rational credence distributions is that they are probability distributions.¹⁷ A probability distribution is usually taken to be one that satisfies Kolmogorov's three probability axioms: non-negativity, normalization, and finite additivity. These three axioms also have parallels for conditional probabilities. John Harsanyi argues that for uncertain or risky decision situations, the Bayesian needs two further constraints: probabilistic equivalence and the sure-thing principle.¹⁸ The latter of these was earlier argued for by Leonard Savage.¹⁹

Probabilistic equivalence states that one should be indifferent between two lotteries if they are of the form (A₁ IF $e_1 ldots A_N$ IF e_N) and (A₁ IF $f_1 ldots A_N$ IF f_N), where $pr(e_n) = pr(f_n)$ for n = 1 ldots. N. This notation means that, to take a traditional ticket-based lottery, if you draw ticket e_1 you will get prize A₁, and so on until the Nth ticket. To take a very inconsequential example demonstrating probabilistic equivalence, suppose the tickets in the e-lottery are blue and tickets in the f-lottery are red, a rational agent will be indifferent between these two lotteries, and the same should hold for any change as long as the probabilities remain equivalent. The sure-thing principle states that a rational agent will always prefer a lottery of the form (A₁ IF $e_1 ldots A_N$ IF e_N) to (A*₁ IF $e_1 ldots A_N$ IF e_N), where A_n is strictly preferred to A*_n for n = 1 ldots. N. This means that if you prefer every prize in one lottery respective to every prize in another, and the

¹⁷ See Joyce, "Bayesianism." Hansson, *Decision Theory a Brief Introduction*; and Titelbaum, *Fundamentals of Bayesian Epistemology*.

¹⁸ Harsanyi, "Bayesian decision theory," 224.

¹⁹ Leonard Savage, The Foundations of Statistics, 1954.

probabilities of getting each prize are identical, then you ought to prefer the lottery with the better prizes. Titelbaum also includes the ratio formula among his core Bayesian rules.²⁰

A final requirement on credence is updating by conditionalization. While there are various forms of conditionalization (e.g. Jeffrey conditionalization), the basic idea is that if an agent holds a certain degree of belief, p, in α conditional on β at t_1 , and they learn β at t_2 , then they ought to have a degree of belief p in α at t_2 .²¹ For example, if I have a conditional credence of 90% that you will come to work with an umbrella given that it is raining on Monday morning, then if it happens to be raining on Monday morning, I ought to have an unconditional credence of 90% that when you arrive at work you will have an umbrella. This requirement is different from the others as it applies to credence over time, and not to the time-independent structure of one's credences. As such, it is not as important for generating any stand-alone slice of the model unless we are predicting a rational agent's future credences based on their prior ones. These slices are often what decision theoretic analyses of ethical problems provide, but for virtue theory the conditionalization requirement, or something like it, is integral for representing virtuous people; virtuous people are necessarily extended through time; they have a certain ingrained *character* and that character cannot be assessed simply through one decision scenario.²²

Many of these constraints on rationality are contested. For the purposes of this thesis, I only wish to show that these constraints, and constraints like them, can interact with the core concepts of virtue ethics in valuable ways: in other words, how axiomatic rationality interacts

²⁰ Titelbaum, *Fundamentals*, ix.

²¹ For Jeffrey's account see Richard Jeffrey, "The Logic of Decision," in 1st. McGraw-Hill series in probability and statistics, (New York: McGraw-Hill, 1965). For a general account, and an overview of Jeffrey conditionalization, see Titelbaum, *Fundamentals*.

²² See Aristotle, *Nicomachean Ethics*, trans. Robert C. Bartlett and Susan D. Collins (Chicago: The University of Chicago Press, 2012); Rosalind Hursthouse, "Virtue Ethics," *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (2013); and Richard Kraut, "Aristotle's Ethics," *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (2016).

with phronesis and virtue. I think the axioms I have outlined above are sufficient for this cursory purpose, even if one wishes to abandon or reformulate some of them.

1.3 Virtue Ethics

Virtue ethics is primarily concerned with three things: virtues, phronesis (practical wisdom), and Eudaimonia (flourishing, contentment).²³ In this thesis I will use a simple Aristotelian framework of virtue ethics, as this is enough for my purposes. The basic definition of virtue that I will use is: Those dispositions or character traits which, when manifest with practical wisdom, and barring bad luck, lead to the flourishing life.²⁴ I will not use a robust definition of practical wisdom. To paint the very general picture with Aristotle's own words: "Virtue makes the goal right, practical wisdom the things leading to it."²⁵ The things which lead to the goal are an important part of decision theory, and I will return to the nature of phronesis in chapter two. Whatever it is that constitutes the flourishing life extends beyond my bare-bones framework of virtue ethics, and this should make what I say amenable to many different theories, so long as they originate from some Aristotelian roots. A terminological note: When I write "virtue" alone or "the virtues," I am referring to the specific dispositions or character," I

²³ See Hursthouse, "Virtue Ethics"; and Kraut, "Aristotle's Ethics."

²⁴ In the *Nicomachean Ethics* Aristotle refers to virtue as *hexis* (1105b25–6), which translates to state, condition, or disposition. See Kraut (2016) for a discussion. This is also similar to a general definition of virtue given in Hursthouse (2013). In the same article, Hursthouse writes: "For Aristotle, virtue is necessary but not sufficient—what is also needed are external goods which are a matter of luck." See the Chapter V of Book III of the Nicomachean Ethics for Aristotle's discussion of luck. Cohen (1990) argues that bad luck can undermine virtuous character, but that bad luck is not inextricably tied to external goods, i.e. bad luck can affect our flourishing by removing external goods, but it can also affect our moral character without relying on external goods. I find Cohen's argument persuasive here, and so I will assume a more general framing of 'luck' which does not make necessary reference to external goods.

²⁵ Aristotle, Nichomachean Ethics, 1144a7-8.

mean the sum of these dispositions (however this sum is calculated) such that a *person* with all of these dispositions would be virtuous, and when I write "virtue ethics" or "virtue theory," I mean the comprehensive moral theory arising from virtue, virtuous character, phronesis, and Eudaimonia.

CHAPTER TWO

HOW DOES DECISION THEORY CONTRIBUTE TO ETHICS SIMPLICITER, AND HOW COULD IT CONTRIBUTE TO VIRTUE ETHICS?

2.1 How does decision theory contribute to ethics simpliciter?

(1) Making risky and uncertain decisions Certainty conditions are relatively easy to work in for most ethical theories, but a large portion of our decisions, ethical and otherwise, are not made under certainty; we do not *know* what will happen given that we act in this or that way. We are often guessing or simply at a loss. Decision theory is a very good tool for assessing and analyzing decisions made under risk and uncertainty.²⁶ The reasoning here goes as follows:

1) Ethics is concerned with what decisions we ought to make.

2) Our decisions are often (if not always) made under risk or uncertainty.

- 3) Probabilistic reasoning is our best way of making decisions under risk and uncertainty.
- 4) Probabilistic reasoning is often (if not always) our best way of making decisions. (2, 3)

5) Decision theory is a good model of probabilistic reasoning.

6) Decision theory is a good model of one of our best ways of making decisions. (4, 5)

 \therefore Ethics should, barring a better model, make use of decision theory. (1, 6)

²⁶ See McCarthy, "Probability in ethics," for an overview of arguments supporting the use of probability in ethics. See Lumer, "The Relevance of Rational Decision Theory for Ethics," for an overview of arguments supporting the use of decision theory in ethics.

Hansson's warning Sven Ove Hansson argues that decision theory has harmed ethics by emphasizing "means-ends relationships between an individual action and its possible outcomes" when philosophers talk about 'ought' claims.²⁷ Using group-action problems as an example of this, he suggests contemporary moral philosophers assume that arguments of the form "It would be a disaster if everyone did like that" must be backed up by some sort of (decision theoretic, in spirit, if not form) reasoning from the point of view of the individual agent.²⁸ To use his example: it is taken for granted that we cannot condemn flushing a small amount of paint thinner down the drain simply because in combination with other such acts it can have a large detrimental effect. That, instead, we *must* condemn flushing a small amount of paint thinner down the drain by merit of that single action and its possible consequences.²⁹

By means-end relationship here, Hansson is thinking of a type of causal relationship where a means is a possible cause of an end.³⁰ While he does not see this as a particular problem for decision theory, which is primarily concerned with guiding *individual* actions, he argues that restricting moral theory to the effects that individuals can bring about with any specific action is "much more questionable".³¹ His main worry with this move is that binary causal reasoning (from an individual cause to its effects) ignores the broader network of interrelated events simultaneously influencing each other which are often at play in moral problems.

First answer to Hansson's warning A focus on simplistic binary causal relationships is the result of the emphasis on a certain type of consequentialism, not decision theory. Decision theoretic utility is merely a degree of desire, and does not need to be found directly in the outcome or consequence of a prospect-at least, not as outcome is traditionally understood (as

²⁷ Hansson, "The Harmful Influence," 585.²⁸ Hansson, "The Harmful Influence," 589.

²⁹ "It is usually taken for granted that a moral requirement on an individual to act in a certain way has to be based on an appraisal of the effects of that very action." Hansson, "The Harmful Influence," 589.

³⁰ Hansson, "The Harmful Influence," 589.
³¹ Hansson, "The Harmful Influence," 590.

a sort of effect or end). This sort of utility is, at its most basic level, the value an agent associates with a *prospect*, for the purpose of mapping preference, and this can sometimes be the sums of weighted value directly assigned to the related outcomes, but this is not always so. It is important here to note that we should take care when defining what exactly constitutes an 'outcome'.

Suppose we are considering two prospects: quickly murdering someone, or slowly ruminating on their death before murdering them. If we are successful, and the death itself is identical from other points of view, there is a sense in which the outcome is the same: our victim has been murdered, our victim is dead, we killed our victim, etc. There is another sense in which it is not; in one case our victim has been murdered with little deliberation, and in the other our victim has been murdered as the result of slow rumination. Our victim is dead because we quickly murdered them. Our victim is dead because we slowly plotted their demise. Note, however, that in order to distinguish these outcomes we necessarily refer to causes or motivations. So, either the outcomes are identical and the prospects are different, or the outcomes are different *because* we have built features of the prospect into the outcome.

One might be tempted to say, "Wait a minute! Utility just *is* the value an agent assigns to *an outcome* by definition." This, however, is simply a very common practice and not a necessary feature of decision theory. Imagine a person who values a prospect because they feel like similar prospects in the past have ended up helping them in some non-definite way. There is no necessary reference to any specific outcome of the prospect in its associated utility; the utility is derivative of the prospect. Similarly so for a deontological utility: any prospect which

is a rule might be assigned an infinite utility, while any prohibition an infinite disutility.³² The utility value does not reference the outcomes at all.

This is a slippery problem, however, as there are many cases where it is not clear what is valued, and whether or not it should count as something *in* the outcome. If we go back to killing: suppose we can defend ourselves with a knife or with a gun. Suppose we know that the outcome in both cases will be that we kill someone. We might still prefer to do so with a gun rather than with a knife. However, it could be that what we value is some semblance of psychological health, and it is less damaging psychologically to fatally shoot someone than to stab them to death. Thus, what we are actually concerned with are two further types of outcome: one where we are psychologically harmed more, and one where we are psychologically harmed less. Thus we value not being harmed, and this is distinct from the prospect.

Psychological harm might be something that is in outcomes, but this does not seem to be so for everything we value. Take the case of quickly murdering someone and slowly ruminating first; the second could indicate a deeper malfeasance within ourselves, and indeed we generally hold people less accountable for crimes of passion than for premeditated wrongs. Yet, we would be hard pressed to find 'premeditation' *in* the outcome unless we build parts of the 'decision', 'motivation', 'cause', 'action', or the like into it.³³

A virtue ethicist can go one of two ways here: they can insist that virtue is something like deontological utility (something that *always* exists independent of the outcome), or say that, at

³² For a deontological decision theory with infinite utilities see Mark Colyvan, Damian Cox, and Katie Steele, "Modelling the Moral Dimensions of Decisions," *Nous* 44 (2010): 512-515.

³³ I raise this as a concern because the virtue ethicist does not want to say that the things virtuous people value reside only in outcomes. This would be a problem for two reasons: (1) the virtuous person would then not value their own virtuous character (being an honest, just, generous, and so on, person), and this seems like an impossibility; in chapter 4 of book IX of the *Nicomachean Ethics*, Aristotle argues that vicious people do not enjoy their own character, nor do they love themselves, while virtuous people necessarily do both; and (2) if the *proper* things for the virtuous person to value were *simply* things that resided in outcomes, there would be no need for a fuller account of virtuous character (to *justify* actions).

least sometimes, the virtuous person values something directly in the outcome. The first path removes a lot of the force from decision theory; there would be no need to discuss outcomes in a virtuous person's decision making process. I will focus instead on the claim that virtuous people do take into account outcomes when making decisions. If this path s taken, the virtuous person can still ask *why* the virtuous person values certain outcomes in these cases, and the path diverges in two again; they can say that something independent of rational decision theory (e.g. an independent account of the virtues) justifies valuing certain outcomes, or they can assert that some component of rational decision theory itself can justify such values. I will focus on a rational decision theory paired with something independent.

Second answer to Hansson's warning However, even if decision theory *does* emphasize means-ends relationships between individual actions and their possible outcomes, this does not mean we, the users of decision theory (the reflective decision makers and moral theorizers), must emphasize those relationships when we talk about what we ought to do. It should be noted that this is an answer I think Hansson would be happy to adopt, as he concludes his paper by remarking that decision theory has provided a great deal of benefit to moral theory, and that he only advocates a careful, purposeful use when applying it to moral problems.

(2) Attainment of values If we accept that rational decision theory is a good way of achieving the things we value on average, then it can point out flaws or inconsistencies in the pursuit of our values.

(3) Justification of values Rational decision theory is sometimes taken not to simply bolster our moral efficiency or to point out moral errors, but to do real justificatory work.

Christoph Lumer's introduction to a special issue of *Ethical Theory and Moral Practice* focused on the relevance of rational decision theory to ethics goes through a number of ways that rational decision theory can justify ethical positions. The closest he gets to virtue theory is what he calls an "Ethic of Rational Moral Value." Such a view might tie in nicely with *Eudaimonia* and *phronesis*. However, in what follows I will focus more on the commensurability of the virtuous person and a decision theoretic framework, rather than whether or not decision theory can justify part (or all) of virtue ethics. In other words, I will focus on the interactions and relationships between the parts of decision theory (prospects and preferences) and virtue ethics (virtue and phronesis), rather than try to provide any sort of proof for virtue theory via decision theory.

2.2 How can Bayesian decision theory contribute to, or enhance, theories of virtue ethics?

Modelling the virtuous person will not often help with immediate decision making; virtue ethics is not designed to yield direct and simple answers to moral problems. The 15-year-old girl who was raped and considering abortion will not find a coherent answer by thinking, "What would Aristotle do if he was a 15-year-old girl who was raped and considering abortion?"³⁴ Decision theory will not change this; it will yield direct advice in paradigmatic cases of virtuous behavior, just as virtue ethics does, but when it comes to the young girl making that difficult choice—and most real life applications—the answers will still be vague. Such answers are vague under virtue theory because the method of measurement for the relevant values has these

³⁴ As argued in Martha Nussbaum, "Non-Relative Virtues: An Aristotelian Approach," *Midwest Studies in Philosophy* 13(1988).

qualities; there is no simple calculus or function that outputs all the *proper* credence and utility values of a specific 15-year-old girl. This does not mean that there is no answer in such cases, and virtue ethics may indeed give us a better way of evaluating uncertain decision scenarios than competing theories by giving us something else to measure: the virtue of the agent. Even when we are uncertain about what will happen as a result of our decisions, we can be sure of our own character.

The simple definition of virtue I provided in chapter one still functions under uncertainty; those character traits and dispositions which lead a person to flourish do not change just because an agent does not know exactly what will come of what she does. One may worry that we do need to know something about what our *will result* of what we do in order to be virtuous. I think the best response to this has two parts: a practical and a theoretical. The practical answer is that all virtuous agents will be good epistemic agents and will make use of all of their faculties and the evidence they have, including information about what will result of their behavior. The theoretical answer is that it is impossible for an agent to be completely in the dark about what the prospects in their decision scenario *could* result in. Throwing a bowling ball off of a tall building could kill someone or damage a car or fly straight into the sun. Pressing a mysterious red button could cause a catastrophe or it could do nothing. If an agent is presented with a prospect in their mind—a necessity for their making a decision in the first place—they will have information about possible worlds resulting from such a decision, even if such information is plainly "y could cause x" where x is any possible world.

A decision theoretic analysis will—given a robust set of preferences belonging to people who already *have* virtue and practical wisdom, or who we believe have virtue and practical wisdom—allow us to measure the degrees of a virtuous person's beliefs and desires.³⁵ It will provide us with a model that allows us to interpret the internal structure of the virtuous person's preference relation. In the following three sections I will discuss ways in which Bayesian decision theory can enhance and supplement virtue ethics. First, how we can clarify two of the key concepts in virtue theory: phronesis and virtue.

Conceptual clarification of phronesis and virtue

If we have a successful formal model of the virtuous person, it will show us whether or not her preferences and credences adhere to plausible norms of rationality. If the preference structure of a virtuous person satisfies only some or none of these norms, then we can start to ask more pointed questions of interest: Why does the virtuous person violate very plausible norms of rationality? Is this violation a necessary part of the virtue account? If the preference structure satisfies all of our norms, and such a model is possible, then we can ask: If we could have a robust model of virtuous decisions that other people could apply to their own actions, why are these other people, if they do so, not virtuous? Why must the actions be performed by a virtuous person?

The decision theoretic framework offers interesting insight into this question. Under uncertainty conditions, the use of standard models and probabilistic reasoning yields partial answers at best because we do not have access to well-defined probabilities. Even with a complete and fleshed out model of how virtuous people act under certainty and risk, we still only have partial assistance under uncertainty. This is a part of what virtue ethics points to

³⁵ For a method of measuring the variables of preference (degrees of belief and desire) see Ramsey, "Truth and probability"; and Bradley, "Ramsey and the Measurement of Belief," for an in-depth discussion of Ramsey's method, including proofs for many of his assumptions.

when it says we cannot simply look at what virtuous people do and copy them in order to be virtuous, that to some extent we cannot rely only on probability and models, and that we simply ought to act as our virtuous character dictates. We rely upon our intuitions, or our sentiment, or some implicit understanding of what is right. Such decisions, under uncertainty, may form a large portion of our "everyday decisions", and it would explain why something like consequentialism seems to be most effective in policy making or in other situations that often have probabilistically well-defined variables.

The virtue theorist might want to argue for a stronger point: that our moral character matters not only because it helps us act when we are uncertain, but that even if we had complete, accurate, and consistent information describing all possible outcomes of all possible actions (or if some super-computer had this information) the character of the person, and their motivations, *still* matter. This might be the case, as with the intuition that visiting a friend in the hospital because you *care* about them is more important than doing so because of a calculation or rule.

There seem to be two ways we could include the "sick friend" intuition in the model: either through prospects or utility values. If I am visiting my friend in the hospital because I care about them, we can represent this as (1) me *choosing* to care for my friend rather than not; or (2) that *because* I care for my friend, I can include that caring in the value of the care-neutral prospect 'visiting my friend in the hospital'. Choosing to care about someone might seem like an implausible description of human psychology, at least when it comes to immediate decision making (in a broader sense you might be able to convince yourself to stop caring for someone, or vice versa). However, if we change the thinking so that it is about what we can *not* choose, then: if I do not care about whoever is in the hospital, I *know* showing my friend I care about

them by visiting them in the hospital is not an option. I might be able to mislead them, but I cannot affect an emotion that does not exist.

This highlights an interesting feature of the psychology of agents: certain decision theoretic utility values are only available to certain agents; the compassion that you show to someone cannot be valuable if you do not show someone compassion. Such a position would be another argument for the importance of virtuous character: having a virtuous character opens up new (valuable) prospects that people who do not have a virtuous character cannot access. An argument in the reverse might also be made: that lacking virtue opens up new prospects which may often appear to be valuable (for bettering our position in society, etc.). However, the latter type of argument will always fail if our virtue theory is worth its salt; if the virtues just are what result in human flourishing, then whatever value can be brought about by not being virtuous cannot be the sort of value which makes something right or good, by definition. If, for example, a virtuous person is never hateful toward, or motivated by disdain for, innocent people, then a businessman who makes a fortune by being motivated by disdain for innocent people cannot be flourishing. He is mistaken if he thinks he is.

Returning to the question I began this section with: If the virtuous person's credences satisfy our rationality postulates, we can ask: Is phronesis an extension of the rationality requirements? Is it something entirely different?

One aspect of phronesis that seems to extend beyond Bayesian rationality requirements is the ability to not only reason the best way to our ends, but to balance our ends themselves. i.e., the young person who accepts a virtue and acts according to it, but does so at the expense of all other virtues to the extent that they are entirely excluded, is not being guided by phronesis, even if he meets the Bayesian rationality requirements. This is similar to what is called the framing problem in decision theory.

The framing problem can be summed up by the question: What prospects *ought* we to entertain? For example, if a child frames his decision as, "Should I give away all of my cookies or none of my cookies?" and then decides to give them all away at his own expense, he might be acting generously in the management of his personal property, but he may fail to see a third option: distributing his cookies equally. Or a fourth: distributing his cookies based upon who deserves cookies the most. And so on. Thus, perhaps he acts from generosity when he ought to have acted from justice or "greatness of the soul".³⁶ The rationality postulates, however, say nothing about how we ought to frame our decisions, and so while a Bayesian decision theoretic model might remind us to ask, "Am I considering all the relevant prospects?" it will not shed any light on what prospects are actually relevant. Perhaps this is an area where virtue theory can bolster decision theory. While consequentialism might call out more naturally for decision theoretic modelling, neither consequentialism nor rational decision theory offer direct insight into how we ought to frame our decisions, whereas virtue theory is directly engaged with such a problem.

It could be argued that some types of consequentialism offer insight into this problem, but their advice for management of our potential options boils down to: "consider all possibilities and choose the one with the best consequences." This is, I take it, not a very effective solution when we cannot consider, or do not know all of the possibilities. Virtue, on the other hand, prunes our available choices in at least two ways. First, the virtuous person will not give any deep thought to options which are not virtuous, as a product of habits inculcated over time. It seems more likely to be true that humans can develop dispositions to act in certain ways (e.g. in honest, just, generous, and so on ways) than dispositions to filter options by the contents of their consequences. This is because dispositions to be certain ways only make reference to the ways in which we are, they are relative to us as agents, while the content of the consequences

³⁶ Nussbaum, "Non-Relative Virtues."

of our actions are independent of us; we cannot by mere introspection know what our decisions will result in.

This idea might be strengthened by appealing to a distinction Jonathan Raz draws between first- and second-order reasons. The essence of first-order reasons is captured by decision theoretic utility; competing first-order reasons are weighed (either cardinally or ordinally), and the stronger reason prevails.³⁷ Second-order reasons, on the other hand, always trump first-order reasons.³⁸ A general example given by Raz is avoiding making important financial decisions while tired and stressed.³⁹ Thus, for a virtuous person any action which is vicious could be seen as prohibited by second-order reasons. There is no need to weigh the options; any prospect arising from a vicious motivation is simply dispelled by the fact that virtuous people are virtuously motivated. As Raz points out (not specifically in relation to virtue), this does not mean virtuous people cannot, by definition, entertain viciously motivated decisions or recognize that they are possible decisions in a given scenario, only that such prospects are simply ruled out by a higher-order reason.

The second way that virtue prunes our prospects is that, because virtues represent useful habits for attaining Eudaimonia, they include useful habits of inference. Such habits give us general rules for how to frame our decisions. I will discuss this further in the following section.

³⁷ Hzr Murat Kose, "Razian Exclusionary Reasons," Journal of Academic Studies 4, no. 14 (2002): 78.

³⁸ Kose, "Exclusionary Reasons," 78.

³⁹ Kose, "Exclusionary Reasons," 79.

On Ramsey, useful habits, and intellectual virtue

"Induction is such a useful habit, and so to adopt it is reasonable."⁴⁰ -F.P. Ramsey

The fourth and fifth sections of F.P. Ramsey's seminal essay *Truth and Probability* discuss what he calls "the logic of consistency" and "the logic of truth".⁴¹ I think a discussion of this yields an interesting insight into the connection between virtue ethics and decision theory. Ramsey argues that we hold two sorts of knowledge: (1) that arising from our memories and perceptions; and (2) that which we hold because of an argument based on an analysis of (1). This analysis is the business of deductive logic, and is concerned with the consistency of our memories and perceptions, and with avoiding self-contradiction.⁴² It is wholly derivative upon knowledge of the first sort; it does not give us new knowledge, but merely rearranges what we already know. Ramsey claims that inductive logic gives us knowledge like that we get from memory or perception. It is part of the "logic of discovery" or "a human logic." I think he calls such a logic 'human' because he is specifically trying to engage with what we ought to believe as creatures who operate mostly under risk and uncertainty, which is something that formal deductive logic cannot do.

This is the crux of Ramsey's argument: deductive logic is designed to function only under certainty conditions. It tells you when to set your degree of belief in a proposition to one. That is, when you are certain of the premises of a valid deductive argument, you ought to be certain of the conclusion. However, it says nothing about credence values less than one.⁴³ Yet, he says

⁴⁰ Ramsey, "Truth and Probability," 24.

⁴¹ Ramsey, "Truth and Probability."

⁴² Ramsey uses 'formal logic' and 'deductive logic' interchangeably, but contemporary Bayesians claim to have a formal system of inductive logic, and so here I will use 'deductive logic' to refer to what Ramsey sometimes simply calls 'formal logic'.

⁴³ Ramsey uses the language "partial belief" here to mean a degree of belief less than one, e.g. "believing q to the extent of 1/3." However, since I am using "belief" to represent a classificatory doxastic attitude, I will instead write "credence" for consistency.

it seems to be true that we ought to hold such credence values, and I think he is clearly right about this; as I stated earlier, most of our decisions occur under uncertainty or risk, and so any proper normative theory ought to address such situations.⁴⁴ From this he concludes that to justify numerical credence values less than one, we *must* sometimes contradict deductive logic, and that while it may be foolish for an omniscient being to do such a thing, for humans it is entirely rational. A Humean-like consequence of this is that inductive logic cannot be reduced to deductive claims. On this Ramsey writes, "Induction is one of the ultimate sources of knowledge just as memory is: no one regards it as a scandal to philosophy that there is no proof that the world did not begin two minutes ago and that all our memories are not illusory."⁴⁵

If we adopt a definition of virtue that is something like "those character traits which, barring bad luck, lead to a flourishing life", there is an interesting parallel between the rationality requirements and flourishing. Ramsey takes inductive habits (def. of habit: a "rule or law of behavior") to be better or worse for humans depending upon certain facts about our nature. One core assumption when discussing rationality is that the human mind operates in some regulated way; it must follow some rules in order to function such that we can describe reasonable and unreasonable processes. To function, humans must, in the words of John Finnis, make reasonable judgments with the intention of terminating those judgments in action.⁴⁶ Ramsey phrases the core question when investigating a human logic as: "What habits [rules or laws of behavior] in a general sense would it be best for the human mind to have?"⁴⁷ One plausible answer is, just those habits which would lead the mind to flourish; an inferential virtue is just that which follows most closely the truth, and truth is valuable because it helps us flourish.

⁴⁴ Any proposed theory of rationality which wishes to maintain that rationality is valuable is a normative theory, for it says, at least on some level or in some cases, that we ought to behave rationally; we ought to believe the thing which it is reasonable to believe.

⁴⁵ Ramsey, "Truth and Probability," 31.

⁴⁶ John Finnis, "Aquinas' Moral, Political, and Legal Philosophy," *The Stanford Encyclopedia of Philosophy* (2014).

⁴⁷ See Ramsey's discussion of yellow toadstools in "Truth and Probability," 29.

This answer accomplishes three things: (1) rationality requirements which are useful inductive habits simply become new intellectual virtues; (2) thereby our account of the virtues is more robust and defined; and (3) helps us to unify our general account of the virtues with phronesis. Ramsey writes:

We all agree that a man who did not make inductions would be unreasonable: the question is only what this means. In my view it does not mean that the man would in any way sin against formal logic or formal probability; but that he had not got a very useful habit, without which he would be very much worse off, in the sense of being much less likely to have true opinions.⁴⁸

If we replace "to have true opinions" with "to flourish", then we neatly tie induction to virtue and provide a reason *why* some inductive habits are more valuable than others (without supposing true opinions have intrinsic value). With this in mind, we can integrate formal rationality postulates concerning credence into our theory of virtue.

Imagine Leila, a young woman living in a forest. One day she comes across a type of mushroom that she has never seen before. She considers the risks of eating such a mushroom and wants to know what preferences and credences would be rational for her to hold. We can ask, what kind of habits would be useful for Leila to have? There are two types of habits we can consider: preferential habits and credential habits. Credential habits are those which map on to Ramsey's "human logic"; those habits which make us more likely to have true opinions. However, let us start with preferential habits, as these are easier to observe. Imagine Leila violates probabilistic equivalence and prefers "red ticket" lotteries to "blue ticket" lotteries (where color is understood as being value-neutral). Provided that she greatly prefers red *x*'s to blue *x*'s, she would be willing to travel to a far away and dangerous forest to eat from a patch of red

⁴⁸ Ramsey, "Truth and Probability," 31.

mushrooms rather than eat from the patch of, otherwise identical, blue mushrooms right outside her door. Such a habit, I think, is not a useful one given most accounts of flourishing.

Less fanciful examples of this can be found in real life. In a discussion about confabulation, Lisa Bortolotti mentions a 2005 study which found that, when offered a choice between two identical articles of clothing, people repeatedly showed a preference for the piece of clothing on the right. When asked later about this preference, "subjects produced plainly confabulated explanations, such as that the item they had chosen was softer, it had appeared to be better made, or that it had a more attractive colour."49 Here the choice itself does not imply irrationality; people need some, arguably any, rule of thumb to take over in matters of indifference. However, when the agents engage in confabulation after the fact, it seems to indicate more than a throw-away rule for tie-breaking. Rather, they have a general habit (a rule of behavior) to choose things on the right when their other rules do not give them an answer, and such a habit is subject to a psychological compulsion to justify our decisions; the dress on the right must have been softer, prettier, and so on. In parallel, Leila simply insists the red mushrooms fight off disease, when, other than the color, there is no difference in their composition. It is merely a justification Leila gives for an irrational preference. Similar examples can be concocted for the von Neumann-Morgenstern utility axioms, the "sure-thing" principle, and strong independence.

In-depth proofs have been attempted which would show that violating the von Neumann-Morgenstern axioms is necessarily a bad preferential habit. The reasoning goes roughly: if you violate any of the axioms, then you will be willing to take a bet which is a guaranteed loss of something you value. There is some push back against this type of proof.⁵⁰ I think, however,

⁴⁹ Lisa Bortolotti, "Psychiatric classification and diagnosis: delusions and confabulations," Paradigmi

XXXIX(1) (2011): 108.

⁵⁰ See Joyce, "Bayesianism" for arguments against Dutch book violability as a good measure of rationality.

this endeavor is unnecessarily concerned with proving formally that preferential habits are *necessarily*, in all cases, good or bad habits to have. It is enough for the purpose of assessing the usefulness of any particular habit if it is *generally* useful; we need not show, for example, that having transitive preferences is always and necessarily a good habit. I use general to mean both in the scope of its application in known scenarios, and its accuracy in uncertain scenarios. This project of formal proof might arise because 'rationality' is seen as an ultimate and absolute form of reasoning, rather than as describing useful habits in reasoning. I think this is a mistake; as Ramsey points out, the entire purpose of a human logic is to function when we are uncertain. Requiring useful habits to *always* be useful is not what we ought to be concerned with when giving moral advice.

Now let us consider credential habits. Imagine Leila believes that there is less than a 0% chance that eating unknown, bright mushrooms will kill her. It is hard to imagine how such a credence would even function, but suppose her decision scenario could be accurately represented:

prospect	Х	cr(x)	u(x)	EU(x)	EU _{total}
eat mushroom	death	-0.5	-100	50	125
	life	1.5	50	75	
do not eat mushroom	death	0.2	-100	-20	20
	life	0.8	50	40	20

Such a strange structure leads Leila to inordinately value death by eating mushrooms in a completely nonsensical way. This habit seems so un-useful that I cannot imagine anyone actually behaving in such a way, which perhaps is our greatest testament to *bad* credential habits. In fact, it seems so un-useful that I cannot imagine a real example of anyone who violates non-negativity (i.e. where $(\forall x \in X)(cr(x) \ge 0)$ is false). Having negative credences essentially reverses the direction (pos./neg.) of the agent's values; it is as if something is *so* unlikely that the *opposite* will happen instead. This might seem absurd and even impossible for the average person to entertain, but there is nothing inherently contradictory about such a belief

system. It simply is not useful at all for the type of creatures we are, living in the type of world we live in.

Similarly, if she were to only be somewhat confident that the mushroom will either kill her or not, we might think she is prone to making very strange choices; e.g. if her credence distribution could be accurately represented as follow:

cr(eating bright unknown mushrooms will kill me) = 0.2

cr(eating bright unknown mushrooms will not kill me) = 0.2

cr(eating bright unknown mushrooms will result in something else happening) = 0.6

She might be willing to risk the 40% chance that she will die or not die in the hopes that whatever else it is that could happen would be good. Having a credence of 0.4 in a tautology violates normality ($\forall T \in X$)(cr(T) = 1), and this seems to result in her underestimating the chance that she will die from eating an unknown mushroom in the absence of further evidence. Similar examples can be thought up for finite additivity and the ratio formula. Thus we might think, at the very least, violating normality is a *bad* habit to have. If these habits are unquestionably useful, then what is the point of discussing them? What falls out of all of this is that, if generally adhering to the probability axioms are almost undeniably useful habits to have, then we have good reason to think following the consequences of the probability axioms are useful habits. This opens up a great swath of content for the virtue ethicist to make use of.⁵¹

A natural concern following from this discussion is that there are many useful inductive habits which are not rationality postulates, nor parts of being virtuous. For example, never eating mushrooms from a certain grove, because you think that particular grove was cursed by evil magic, might be a useful habit if all of those mushrooms just happened to be poisonous because

⁵¹ See Titelbaum, *Fundamentals*, ix for a list of consequences of the probability axioms & the ratio formula.

of their natural properties. Yet, we do not wish to say 'abstaining from food because one thinks it is cursed by evil magic' is a useful habit to have. The important part of this example is that the mushrooms just so happened, by luck, to be poisonous: there was no deeper connection between the habit and the benefit wrought. To be a useful habit in a broad and normative sense, it must be extended beyond one local domain. But, what of habits which simply happen to be useful in general? e.g. a person who thinks all brightly colored mushrooms are cursed by evil magic in a world where all brightly colored spotted mushrooms are poisonous. These are *better* habits because they are more general and seem to track something true, albeit by some sort of accident. They are better because they are more likely to yield true opinions and—given a virtuous character and a rational preference relation—thus are more likely to lead to a flourishing life.

This line of thought leads us back to my emphasis on uncertainty conditions; a habit which has some more general justification will allow for useful decisions in a broader set of cases. e.g. if we have some concept of a toxin, and have observed that many brightly colored things contain toxins, we might be wary of bright mushrooms *because* we are connecting it to other experiences we have had, even if we know nothing specific of the brightly colored mushrooms in *this* forest. This is a very simple example, but the reasoning can be carried into other domains as well; if we wish to be a good and honest person in as many situations as possible, we need general habits of goodness and honesty. An interesting note is that if we generalize far enough, it seems like we reach a point where it does not matter if we believe that something is so because of evil magic; if we explain the concept of toxin to the best of our ability and simply append "and this is because of evil magic" to the end, we cannot argue against that *particular* habit of belief formation by saying it is not general enough to be useful. Though if appending "and this is because of evil magic" to the end of all explanations is one of your habits, such a habit does not seem eminently useful when it comes to leading a flourishing life.

One might want these postulates to be necessary requirements for rationality and not *just* useful habits. I think, however, this is again a mistake; we will strip the essence of rationality away if we attempt to define it without justifying it as picking out useful habits for humans to have when reasoning.

Credence and Subjective Bayesianism

Recall that credences are degrees of belief, which I am treating as quantitative doxastic attitudes. For the purpose of this thesis, I will make certain assumptions about credence. The first of these is that credence is an attitude toward a proposition, and when I speak of assigning numerical value to a virtuous agent's degree of belief, this numerical value is part of the attitude itself, not in the proposition. That is, if we examine the virtuous person's preferences and find that their cr(it will rain tomorrow)=0.5, the propositional content of that credence is wholly contained within the parentheses; the "50%" is part of the doxastic attitude that the agent adopts toward the proposition "It will rain tomorrow." i.e. they have a credence *of* 50%.⁵² The second is that ideal moral agents do not necessarily think in explicit numerical terms. However, just as we impose numerical structure upon the world when measuring distances between stars in a galaxy, imposing a numerical structure upon the credences of agents can be a useful and informative practice, especially when functioning as moral theorists.

According to Titelbaum, the decisive question between (normatively) objective and subjective Bayesians is: Is there more than one rationally acceptable credence assignment for an identical proposition given the same set of evidence? In other words, if I adopt cr(it will rain tomorrow)=0.5 and you adopt cr(it will rain tomorrow)=0.6, given that we have all the same

⁵² This distinction is again emphasized in Titelbaum, *Fundamentals*.

evidence, is one of us necessarily wrong? The objective Bayesian says Yes, while the subjective Bayesian says No. For the Bayesian this question is closely tied to the discussion of hypothetical prior distributions, or the ur-prior, of an agent who has been updating by conditionalization. As a brief reminder, conditionalization means that if I hold the conditional credence $cr_i(x | E) = p$ and learn E at time j, then my *un*conditional credence in x at time j ought to be p (i.e. $cr_j(x) = p$). The ur-prior is a distribution from which we can calculate an agent's belief in any proposition at any time. This is done by conditionalizing her ur-prior on the sum of her total evidence at that time. i.e. where *t* represents any point in time, $cr_t(x) = cr_{ur}(x | E_t)$. The ur-prior is a logical consequence of an agent whose credence distribution satisfies Kolmogorov's axioms and the ratio formula, and whose credences have been updated by conditionalization. Titelbaum writes:

Yet being regular, the hypothetical prior does not assign any contingent certainties itself. This means that it contains no contingent evidence, and is not influenced by empirical experiences. So when we are confronted with an agent's credences at any given time, we can cleanly separate out two distinct influences on those credences: her total evidence and her ultimate evidential standards.

Her evidential standards are tied to the question posed above, which splits the objective/subjective normative Bayesian camps; if one thinks there is only one ur-prior that rational agents *ought* to have, then they are an objective Bayesian, and vice versa. This has an interesting connection to virtue ethics: namely, if there is only one correct ur-prior belonging to virtuous agents, then we have an easy litmus test for virtue. This is not to say a certain assignment of credences is sufficient for being virtuous, only that it would be one necessary part of being virtuous. Thus, if someone failed to have the correct ur-prior with which to conditionalize upon, they could not be virtuous. If there is not one correct ur-prior, and I think this is likely the case, then we have an epistemic argument for the vagueness present in phronesis *and* we loosen the more strict requirements of rational decision theory that seem to

clash with the un-quantifiable aspects of virtuous character, because any number of ur-priors are rational, provided certain useful habits are followed (Kolmogorov's axioms and the ratio formula) and credences are updated by a general rule over time (conditionalization). This move simultaneously lays out strict rules for practical wisdom/intellectual virtue *and* allows for the robustness that many thinkers may find attractive within virtue ethics.

CHAPTER THREE

IS IT POSSIBLE TO HAVE A BAYESIAN DECISION THEORY REPRESENTING VIRTUE?

Many of the benefits I discussed in chapter 2 make use of the suggestions in what follows, however this will simply be a cursory look at the possibility of having a model which satisfies (F1)-(F3). Answering the following two questions is a good start to laying the groundwork for future developments of a virtue theoretic formal decision model: (1) How do the Bayesian rationality requirements on credence and preference interact with virtue?⁵³ And, the largest hurdle for a formal virtue theory, (2) How can virtue and virtuous character be represented by degrees of desire?

Rationality requirements and virtue

I think it is plausible, given the nature of rationality and my argument in the spirit of Ramsey from chapter 2, that the requirements on credence can fit quite naturally into accounts of virtue, so long as we tie the virtues to the ability to attain the Eudaimonic life. I think it also plausible that requirements on preference can fit naturally into accounts of virtue, as I suggested with credential equivalence. However, value maximization is one rationality requirement that we might worry is at odds with virtue ethics; what is it exactly that the virtuous person maximizes,

⁵³ I ask this question because some "rationality" restrictions seem necessary for any normative decision theory.

and is maximization even part of what the virtuous person does? To the former part of this question, the answer is simply: their virtuous character.

This might seem like an unhelpful answer, but it contains an important assertion: the virtuous person, like all people, has values that she pursues. Since these values guide the virtuous person's actions, and the virtuous person acts rightly, these values guide right action. Her values need not be traditional ends, and perhaps most often are not. e.g. the virtuous person acts honestly because she values honesty itself; she values *being* honest, not the consequences of being honest. She values virtues in and of themselves because having such values leads her to flourish. The 'because' does not indicate that she first thinks "behaving thusly will lead me to flourish", but only that in some way the reason why she behaves thusly is because it is connected to human flourishing. An important note to make here: valuing something does not entail maximizing it.⁵⁴ However, the type of maximization here is of those things which lead to flourishing; if having too much of one quality diminishes flourishing, then it is no longer a virtue. In other words, the sort of maximizing a virtuous person does cannot, by definition, be excess; what the virtuous person maximizes are her odds of flourishing, and this is directly proportional to the degree of her virtuous character.

What the virtuous person maximizes can be thought of as agent-relative value. Jennie Louise defines this as a value which cannot be specified without referring to the agent who is doing the valuing.⁵⁵ In addition to being agent-relative, the value of the virtuous agent is also time-relative; the virtuous person will not, or perhaps cannot, act viciously (i.e. from vice) in order to bring a more virtuous character for themselves, or others, at a future time. That is, the

⁵⁴ Thanks to Simon Rippon for pointing this out.

⁵⁵ Jennie Louise, "Relativity of Value and the Consequentialist Umbrella," *The Philosophical Quarterly* 54 (2004): 519.

virtuous person values being virtuous and thus *maximizes* temporally-local value which is relative to her own character.

There is a worry here raised by Campbell Brown that an agent-relative theory of good which orders centered worlds (opposed to simple possible worlds) is bound to be self-contradictory.⁵⁶ One potential response to this is: Having a virtuous character includes considering, and acting in the benefit of, the state of virtue of your society. Aristotle argues for a position resembling this in Books VIII and IX of the Nichomachean Ethics. We get a similar line from Plato in the Apology, when Socrates defends himself against the charge of corrupting the youth. If we extended this principle out to entire worlds rather than local social groups, the question becomes: Does the virtuous person *aim* to bring about the possible world in which *she* is most virtuous (in a time-relative way), which simply happens to be the biggest contribution she can make in order to shift the actual world toward the possible world with the most virtuous people in it? This question is open for a more robust theory of virtue to answer, however I think most virtue theories would deny the first, and if the second is true—i.e. maximizing time-relative agent-relative virtuous character is the best way to maximize the amount of virtuous people in the world—then either aim is identical from a the view of a decision procedure.

Degrees of desire and virtue

On to the second and more difficult question: How can virtue and virtuous character be represented by degrees of desire? The most common approach for answering these questions in general (non-virtue theoretic) ethics is with a representation theorem. Representation

⁵⁶ Campbell Brown, "Consequentialize This," *Ethics* 121 (2011): 762.

theorems in decision theory share the following assertion: If an agent's preference relation satisfies certain requirements, then there will be a credence (or for non-Bayesians, probability) function and a utility function which yield expected utilities that describe the agent's preference relation.⁵⁷ That is, $(\forall a, b \in A)(EU(a) \ge EU(b) \Rightarrow a \gtrsim b)$; a greater expected utility value entails preference. This means that the expected utility function provided by a theorem will be a *numerical representation* of the agent's preference relation. Richard Bradley makes an important observation about the different roles representation theorems play, first, in *justifying* rational degrees of belief and desire, and second, in *measuring* the related variables: an agent's degree of belief or desire.⁵⁸ The second use underlies what I have called the biggest hurdle for unifying virtue with utility: how do we measure the values of the virtuous agent? The primary question here is: can the relevant virtues be represented by the degrees of desire of virtuous agents? The answer, I believe, is Yes in at least one way. That is, whatever value it is that the virtuous agent pursues when making their decisions can be tracked by a representation theorem applied to their preferences.

The second chapter of this thesis discussed justifying *representation theorems* as formally capturing useful habits, and so justifying the degrees of belief and desire that they represent as rational. I will elaborate briefly on that here: There are at least two ways that the virtue theorist can deal with the justification of value (of some utility distributions over others) in decision theory; they can try to incorporate it in the model such that the value associated with prospects contains the justifying material, or they can hold that something external to the model-value justifies that value (e.g. a virtuous person might value some welfare in an outcome, but that something else external to the model justifies acting upon that value). The first would be

⁵⁷ Ramsey laid the foundation for representation theorems in 1926, and Savage provided a popular proof in 1954. Later representation theorems were provided by Jeffrey (1965) and Joyce (1999).

⁵⁸ Bradley, "Ramsey and the Measurement of Belief."

achieved by representing, in some way, virtuous character *in* the value associated with prospects. e.g. having a compassionate character would have an effect on the utility value of your behaving compassionately. The second would be achieved by saying something like: such a prospect is compassionately motivated, but it is not the act of a morally good agent unless it is the manifestation of a deeper virtuous character. In other words, I might help a struggling person because it is deeply ingrained in my character to care about the struggles of others, or I might help a struggling person because I follow the rule 'do unto others as you want them to do unto you.' There is a sense in which both of these can be deeply ingrained in my character, but if helping struggling people is virtuous, the virtuous person has a preference for that *in and of itself* not for a rule whose consequence is the preference for what is virtuous.

One potential solution to this problem is connecting agent-relative value to some sort of *utility* updating method; if the utilities a virtuous agent has are dependent upon her character, then her utility distribution at one time is dependent upon some deeper assignment of utilities. Say that the utility function in the representation theorem of the virtuous person tracks agent-relative values (which are defined by our account of the virtues), and that, as Aristotle thought, there are some spheres of experience in which all humans share. The virtuous person must manipulate her values in these spheres in some way in order to reach the values she has in less general spheres of experience. Such an updating rule would depend a great deal upon what sort of virtue ethics we have.

CONCLUSION

The main purpose of this thesis was to argue that thinking about virtue and phronesis in terms of rational decision theory is useful. In some ways it is useful for rational decision theory (e.g. dealing with the framing problem and clarifying why we are interested in *rational* decision theory), but mostly for the philosopher who wishes to have a full and coherent virtue ethics. A Bayesian decision theory can benefit virtue ethics by clarifying both what constitutes virtue and phronesis, and if a representation theorem of the virtuous person is created, then we have a quantitative way of testing certain qualities of the virtuous person's credences and preferences. The father of representation theorems himself speaks of the necessity of useful inductive habits, which I have claimed can be easily shifted into virtue theoretic terms, and even supplemented in the process.

Furthermore, adopting a subjective Bayesian viewpoint, which admits of many rational urpriors, both assists us in defining phronesis and in capturing much of the robustness, or the "human" aspect, of traditional virtue ethics. In other words, it helps explain why the recommendations of virtue ethics are not clean and simple; depending on the ur-prior of the virtuous agent, of which there is no single *correct* ur-prior, the degrees of belief she holds at any point in time can vary from other equally virtuous agents (all acting with phronesis). However, if they all update their beliefs by conditionalization, there will be similarities in their degrees of belief over time, and they will often converge in the limit.⁵⁹

⁵⁹ This, however, is not always the case; a clever epistemologist can pick a set of ur-priors which diverge in the limit on simple questions (such as confidence levels when pulling multiply colored balls from an urn). I think this should not dissuade us from labeling subjective Bayesianism as useful in general, even if in such cases it is not.

In the final chapter, I suggested that if we take virtuous agents to be maximizing agent- and time-relative value (determined by our account of virtue), we can create a traditional representation theorem of the virtuous agent. This, however, does not capture the full *justification* of virtue theory. I provided two suggested solutions to this problem. The first was to leave the justification out of the model. The second was to build in some sort of utility updating principal into rationality requirements such that the virtuous agent's utility distribution, at any time, is derivative or conditional upon their utility distribution in shared spheres of experience, or in some other way upon their dispositions and character.

While decision theory does help the virtue theorist clarify and refine their theory, it will not do all the work. For a decision theoretic analysis to be useful, the virtue theorist must have a good and accurate account of Eudaimonia; if they do not, or cannot, have such an account, a decision theoretic analysis will not likely provide one.⁶⁰ Furthermore, Bayesian norms of rationality only seem useful for supplementing and clarifying an account of virtue; without a pre-existing and robust account of what virtue is or ought to be, such a decision model does not extend beyond what Bayesians simply tell us is rational. Here I assumed that virtue consists of those habits which lead to flourishing, but this is far from uncontroversial.⁶¹ I have placed a great deal of responsibility on the virtue theorist. They must have an accurate and robust account of Eudaimonia, and at least provide a foundational account for virtue, but this is a responsibility we must accept if we truly wish to give virtue ethics a fair trial.

⁶⁰ The reason I say "likely" is that philosophers such as Lumer (2010), have suggested that decision theory can actually justify moral values. If this is the case, such an approach might work to provide a justified account of what it means for a human to flourish.

⁶¹ See Hursthouse, "Virtue Ethics," section 2.

Bibliography

- Aristotle. *Nicomachean Ethics*. Translated by Robert C. Bartlett and Susan D. Collins. Chicago: The University of Chicago Press, 2012.
- Aumann, Robert. "Utility Theory Without the Completeness Axiom," *Econometric Research Program*: Research Memorandum, no. 26 (1961).
- Bastons, Miquel. "The role of virtues in the framing of decisions," *Journal of Business Ethics* 78 (2008): 389-400.
- Bortolotti, Lisa. "Psychiatric classification and diagnosis: delusions and confabulations." *Paradigmi* XXXIX, no. 1 (2011): 99–112.
- Bradley, Richard. "Ramsey and the Measurement of Belief," in *Foundations of Bayesianism*, ed. D. Corfield and J. Williamson, 263-290. London: Kluwer, 2001.
- Brown, Campbell. "Consequentialize This," Ethics 121, no. 4 (2011): 749-771.
- Cohen, Sheldon M. "Luck and Happiness in the 'Nicomachean Ethics'," *Soundings: An Interdisciplinary Journal* 73 (1990): 209-218. http://www.jstor.org/stable/41178513/>.
- Colyvan, Mark, Damian Cox, and Katie Steele. "Modelling the Moral Dimensions of Decisions," *Nous* 44 (2010): 503-529. doi:10.1111/j.1468-0068.2010.00754.x.
- Finnis, John. "Aquinas' Moral, Political, and Legal Philosophy," *The Stanford Encyclopedia of Philosophy* (Summer 2014 Edition) ed. Edward N. Zalta. http://plato.stanford.edu/archives/sum2014/entries/aquinas-moral-political/>.
- Hájek, Alan. "Interpretations of Probability," *The Stanford Encyclopedia of Philosophy* (Winter 2012 Edition), Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/win2012/entries/probability-interpret/.
 ——. "What Conditional Probability Could Not Be," *Synthese* 137, no. 3 (2003): 273-323.
- Hansson, Sven Ove. "The Harmful Influence of Decision Theory on Ethics," *Ethical Theory* & *Moral Practice* 13 (2010): 585-593. doi:10.1007/s10677-010-9232-0.
- *Decision theory: a brief introduction.* Technical report, Department of Philosophy and the History of Technology, Royal Institute of Technology (2005).
- Harsanyi, John. "Bayesian decision theory and utilitarian ethics," American Economic Review, Papers and Proceedings 68 (1978): 223–228.
- Hausner, Melvin. "Multidimensional Utilities." In *Decision Processes*, ed. R. M. Thrall, C. H. Coombs, and R. L. Davis, 167-180. New York, 1954.

Holton, Richard. "Partial belief, partial intention." Mind 117, no. 465 (2008): 27-58.

- Hooker, Brad. "Act-consequentialism," in *Ideal Code, Real World*. (Oxford: Oxford University Press, 2000), 142-158.
- Hursthouse, Rosalind. "Virtue Ethics," *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. 2013. plato.stanford.edu/archives/fall2013/entries/ethics-virtue/.

Jeffrey, Richard C. "The Logic of Decision." New York: McGraw-Hill, 1965.

- Joyce, James. "Bayesianism." In *The Oxford Handbook of Rationality*. Oxford: Oxford University Press, 2004.
- Kolmogorov, A. N. Foundations of Probability, New York: Chelsea Publishing Company, 1950.
- Kose, Hzr Murat. "Razian Exclusionary Reasons,." *Journal of Academic Studies* 4, no. 14 (2002): 77-93.
- Kraut, Richard. "Aristotle's Ethics" *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. 2016. plato.stanford.edu/archives/spr2016/entries/aristotle-ethics/.
- Louise, Jennie. "Relativity of Value and the Consequentialist Umbrella," *The Philosophical Quarterly* 54 (2004): 518-536.
- Lumer, Christoph. "Introduction: The Relevance of Rational Decision Theory for Ethics," *Ethical Theory and Moral Practice* 13 (2010): 485-496.
- Narveson, Jan. "The Relevance of Decision Theory to Ethical Theory." *Ethical Theory & Moral Practice* 13, no. 5 (2010): 497-520. doi:10.1007/s10677-009-9193-3.
- McCarthy, David. "Probability in ethics." In *The Oxford Handbook of Philosophy and Probability* edited by Alan Hájek & Christopher Hitchcock. Oxford: Oxford University Press, forthcoming.
- Morales-Sanchez, Rafael, and Carmen Cabello-Medina. "The Role of Four Universal Moral Competencies in Ethical Decision Making." *Journal of Business Ethics* 116 (2013): 717-734.
- Neumann, John von and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press; New York: John Wiley and Sons, 1964.
- Nussbaum, Martha. "Non-Relative Virtues: An Aristotelian Approach," *Midwest Studies in Philosophy* 13 (1988): 32-53.
- Ramsey, Frank P. "Truth and Probability." In Ramsey, 1931, *The Foundations of Mathematics and other Logical Essays*: Ch. VII, 156-198. edited by R.B. Braithwaite. London: Kegan, Paul, Trench, Trubner & Co., New York: Harcourt, Brace and Company, (1926).

Savage, Leonard J. The Foundations of Statistics. Wiley Publications in Statistics, 1954

Schauenberg, Bernd, Roger Dunbar and Rudi Bresser. "The Role of Transitivity in Decision Theory." *International Studies of Management & Organization* 11, no. 1 (1981): 33-55. http://www.jstor.org/stable/40396885.