Franz Dietrich and Christian List

What matters and how it matters: a choice-theoretic representation of moral theories

Article (Accepted version)
(Refereed)

What Matters and How it Matters:
A Choice-Theoretic Representation of Moral Theories*

Franz Dietrich
CNRS & Paris School of Economics

Christian List
London School of Economics

This version: March-April 2017

Abstract

We present a new "reason-based" approach to the formal representation of moral theories, drawing on recent decision-theoretic work. We show that any moral theory within a very large class can be represented in terms of two parameters: (i) a specification of which properties of the objects of moral choice matter in any given context, and (ii) a specification of how these properties matter. Reason-based representations provide a very general taxonomy of moral theories, as differences among theories can be attributed to differences in their two key parameters. We can thus formalize several distinctions, such as between consequentialist and non-consequentialist theories, between universalist and relativist theories, between agent-neutral and agent-relative theories, between monistic and pluralistic theories, between atomistic and holistic theories, and between theories with a teleological structure and those without. Reason-based representations also shed light on an important but under-appreciated phenomenon: the "underdetermination of moral theory by deontic content".

*Earlier versions of this work were presented at the Workshop on "Social Dynamics of Information Change", University of Amsterdam, 12/2013; the 2nd International Conference on Economic Philosophy, Strasbourg, 10/2014; a seminar on "Sciences, Normes, Décisions", Université Paris-Sorbonne, 5/2015; the Workshop on "Ethics, Uncertainty, and Decision Theory", Australian National University, 7/2015; the Conference on "Deontic Logic and Ethics", Venice International University, 11/2015; a seminar on "Décision, Rationalité et Interaction", ENS, Paris, 3/2016; a seminar on "Engineering Moral Agents", Schloss Dagstuhl, 6/2016; a PHARE seminar, Paris, 6/2016; and the 13th Meeting of the Society for Social Choice and Welfare, Lund, 7/2016. We thank the audiences as well as Mikaël Cozic, Thomas Scanlon, Kai Spiekermann, and Laura Valentini for valuable feedback. We are especially grateful to John Broome, Seth Lazar, and the anonymous readers for extensive comments, which have greatly helped us to improve the paper. Dietrich was supported by the French Agence Nationale de la Recherche (ANR-12-INEG-0006-01); List was supported by a Leverhulme Major Research Fellowship.

1 Introduction

The aim of this paper is to propose a new approach to the formal representation of moral theories. We show that any moral theory within a very large class can be represented in terms of two parameters:

(i) a specification of which properties of the objects of moral choice matter in any given context, and

(ii) a specification of how these properties matter.

The first parameter tells us what the normatively relevant properties are, the second which sets of properties "outweigh" or "defeat" which others. We call a representation of a moral theory in terms of these two parameters a *reason-based representation*; we give a precise definition below. Reason-based representations encode not only a theory's *action-guiding recommendations* (i.e., how we should act, according to the theory), but also the *reasons* behind those recommendations (i.e., why we should act in that way).

We show that reason-based representations provide a very general taxonomy of moral theories, since differences among theories can be attributed to differences in their two key parameters. In particular, this way of representing moral theories can clarify a number of salient distinctions, such as between consequentialist and non-consequentialist theories, between universalist and relativist theories, between agent-neutral and agent-relative theories, between monistic and pluralistic theories, between atomistic and holistic theories, and between theories with a teleological structure and those without.

Reason-based representations also shed light on an important, but still under-appreciated phenomenon: different moral theories may coincide in all their action-guiding recommendations, despite arriving at them in different ways (e.g., Dreier 1993; Broome 2004, ch. 3; Portmore 2011). Put differently, the same action-guiding recommendations may be explained in more than one way. For example, some deontologists and some consequentialists may agree on all "ought" statements, but offer different explanations for them. Our reason-based approach allows us to formalize, and investigate the generality of, this phenomenon: the *underdetermination of moral theory by deontic content*.

In developing our approach, we build on the existing debate on whether all moral theories can be "consequentialized".[1] Roughly speaking, a moral theory is "consequentializable" if its action-guiding recommendations are the same as those of some "counter-

[1]See, among others, Vallentyne (1988), Broome (1991, 2004), Oddie and Milne (1991), Dreier (1993, 2011), Louise (2004), Portmore (2007, 2011), Sachs (2010), Brown (2011), and Tenenbaum (2014).

part theory" that is structurally consequentialist.[2] Some scholars, such as Jamie Dreier (1993, 2011) and Douglas Portmore (2007), suggest that every moral theory – or at least every plausible one – can be represented in a consequentialist format, provided we employ a sufficiently broad notion of *consequences*. Others, such as Campbell Brown (2011, 750), argue that "[t]here are in fact limits to consequentialization": whenever a moral theory does not satisfy certain formal constraints, it defies consequentialization, in a sense that can be made precise.

We confirm that there are limits to consequentialization, unless we permit unilluminating ways of redescribing the options of moral choice. This raises the question of whether a theory that falls outside those limits can be represented in some other canonical way. Our framework gives a positive answer to this question. Moreover, even when a theory *can* be consequentialized, the framework enables us to represent not only the action-guiding recommendations but also the underlying reasons. Finally, going beyond the debate on consequentialization, the framework allows us to determine, for each of the attributes *consequentialist*, *universalist*, *agent-neutral*, *monistic*, *atomistic*, and *teleological*, which theories have that attribute and which are redescribable in a form that has it. We can call such theories "A-izable", where A is the attribute in question.[3]

2 What do we mean by a moral theory?

We begin with some basic terminology and informal background to our discussion. We define our central concepts more precisely in the subsequent formal exposition.

2.1 Normative versus axiological theories

Moral theories, broadly construed, can be of at least two kinds: they can be *axiological* or *normative*. An *axiological theory* is a theory of how good or bad (or better or worse relative to one another) certain objects of assessment are. The objects of assessment can be, for instance, possible worlds, states of affairs, actions, or consequences. A *normative theory*, by contrast, is a theory of which actions (or policies, plans, arrangements) are permissible or impermissible, right or wrong. Action-guidance is usually delivered by normative theories, not by axiological ones.

[2]The "counterpart" terminology can be found, e.g., in Louise (2004), Portmore (2007), and Brown (2011).

[3]Our analysis, while self-contained, builds on our earlier work on reason-based choice. Specifically, we build on the formal framework in Dietrich and List (2016), offering a normative rather than positive-explanatory interpretation and application. For more distantly related works on reason-based preferences, see Dietrich and List (2013), Liu (2010), Pettit (1991), and Osherson and Weinstein (2012).

Of course, there can be connections between theories of these two kinds. Many normative theories are based, directly or indirectly, on an axiological theory. Consequentialist theories, such as utilitarianism, are like this. They are normative theories that are defined on the basis of an underlying axiological theory which refers to the goodness of consequences (total welfare in the case of utilitarianism). Typically, they deem actions permissible or right if and only if they bring about the best feasible consequences. As already mentioned, the "consequentialization debate" concerns the question of whether all normative theories can be re-expressed in this format, under a suitable interpretation of "best consequences".

More broadly, a theory is *teleological* if its criterion for the permissibility of actions is that they are the best feasible ones, according to *some* underlying axiological theory, though not necessarily one that focuses on consequences alone. Assessments of goodness or betterness could focus, for instance, on how the acts relate to the context of choice. We can then say that a normative theory can be *teleologized* if it can be re-expressed in a teleological format, i.e., if we can construct a teleological counterpart theory with the same action-guiding recommendations (for earlier discussions, see, e.g., Broome 2004, ch. 3, and Vallentyne 1988).

We here focus on normative theories, rather than axiological ones. Therefore, when we speak of a moral theory, this should be understood to refer to a normative theory, unless otherwise stated.

2.2 Normative theories and their deontic content

Any normative theory – perhaps together with some auxiliary assumptions – entails a body of permissibility verdicts: verdicts about which actions are permissible in any given context, and which are not. These are the theory's action-guiding recommendations. Let us call this the *deontic content* of the theory.[4]

Typically, the theory itself is more than just an enumeration of its permissibility verdicts. It goes beyond its deontic content. This is because the theory offers a systematization or explanation of the implied permissibility verdicts, for example by identifying the reasons and general principles underpinning them. It is entirely possible, for instance, that different normative theories entail the same permissibility verdicts, despite arriving at them in different ways. Several authors have recognized this phenomenon, sometimes under the label of "extensional equivalence" (e.g., Portmore 2011, Dreier 2011, and ear-

[4]This follows Brown's (2011) and Portmore's (2011) terminology (*deontic outputs* or *verdicts*). For related terminology, see also Oddie and Milne (1991).

lier, Lyons 1965).[5] A striking suggestion of extensional equivalence can be found in Derek Parfit's book, *On What Matters* (2011). Parfit argues that his favourite versions of consequentialism, Kantianism, and Scanlonian contractualism essentially coincide in their recommendations and can be seen as attempts to climb the same mountain from different sides. Similarly, John Broome (2004, ch. 3) observes that the same normative recommendations may be derived from different axiologies.

If we are interested not only in how we ought to act, but also in why we ought to act in that way, then we cannot generally consider two extensionally equivalent theories as equivalent *simpliciter*. As Portmore (2011, 109) observes:

> "[E]ven if two theories agree as to which acts are right, that does not mean that they agree on what makes those acts right... [M]oral theories are in the business of specifying what makes acts right. And so even two moral theories that are extensionally equivalent in their deontic verdicts can constitute distinct moral theories – that is, distinct theories about what makes acts right."[6]

The present problem is analogous to the case of science, where two or more distinct theories may explain the same observations and thus be observationally equivalent, despite being explanatorily different: W. V. Quine's famous *empirical underdetermination problem* (1975). Proponents of an instrumentalist view of science typically deny that there is much at stake in our choice among observationally equivalent theories. They think that the main point of a scientific theory is to accommodate the empirical observations (see, e.g., van Fraassen 1980); our theoretical constructs are just instrumentally useful representation devices. By contrast, scientific realists insist that there is a fact of the matter as to which theory offers the right explanation of the observations: no more than one of the rival theories can be true (see, e.g., Psillos 1999). Here, it is not only the theory's observable implications than can be true or false, but also the theoretical constructs offered as an explanation: the unobservables.

Similarly, normative theories are underdetermined by their deontic content. Some scholars, such as Dreier (2011), think that there is not much at stake in such cases of underdetermination. This view parallels the instrumentalist one in science. However, we think that those who attach significance to normative reasons ought to disagree. From their perspective, an accurate representation of a moral theory should capture not only the theory's deontic content, but also the underlying reasons or principles.

[5] See also the references in footnote 1.

[6] Dreier (2011) also discusses this passage.

3 How do we formalize a theory's deontic content?

We first explain how to formalize a theory's deontic content, following Broome (1991) and Brown (2011) in taking a decision-theoretic approach, and then we briefly revisit the limits of consequentialization.[7]

3.1 Choice contexts, options, and rightness functions

As noted, a theory's action-guiding recommendations are encoded by its deontic content: a specification of which actions are permissible in each context, and which are not. We formalize this as follows.

Let $\mathcal{K}$ be a set of possible *choice contexts* that an agent may be faced with. Each context K (an element of $\mathcal{K}$) is a situation in which the agent has to choose among, or appraise, some options, such as different actions or prospects. Let $[K]$ denote the set of *available options* in context K. It is most natural to interpret these as the options that are "feasible" in that context, where readers may plug in their preferred notion of feasibility.[8] The set $[K]$, in turn, is a subset of a universal set X of possible options.[9]

For each context K, a normative theory specifies which of the available options are permissible, and which not. To capture this, we introduce the notion of a *rightness function*. This is a function, denoted R, which assigns to each context K the set $R(K)$ of "permissible" or "right" options in that context, where $R(K)$ is a subset of $[K]$. In decision-theoretic terms, the function R is a *choice function*, reinterpreted to capture "permissible" or "right" choice, rather than "actual" or "formally rational" choice.[10]

A rightness function expresses a theory's deontic content. If, according to the theory, every available option is permissible in context K, then the set of permissible options $R(K)$ coincides with $[K]$. If there is a unique permissible option in K, then $R(K)$ is singleton. If there is no permissible option, then $R(K)$ is empty, in which case the

[7]Related decision-theoretic approaches can also be found in other contributions to the consequentialization debate, as cited in footnote 1.

[8]On a thinner interpretation, the options in $[K]$ are simply those that are candidates for appraisal in context K, whether feasible or not. Our default interpretation of $[K]$, as the set of "feasible" options, is natural if the normative theory in question obeys an "ought implies can" constraint. This interpretation is required for our definition of a "dilemma-free" rightness function below. The thinner interpretation (under which "available" means "being a candidate for appraisal") is compatible with the idea that we may appraise options as permissible or impermissible even before assessing their feasibility.

[9]In particular, X contains every option that is available in at least one context, i.e., $X \supseteq \cup_{K \in \mathcal{K}}[K]$.

[10]Formally, R is a function from the set $\mathcal{K}$ of contexts into a set of subsets of X, where, for each K, $R(K) \subseteq [K]$. Note that a standard choice function assigns to each context a *non-empty* set of chosen options. We do not require non-emptiness. Our definition is similar to Brown's (2011).

agent faces a moral dilemma. We call the rightness function R *dilemma-free* if $R(K)$ is non-emtpy for every context K. (Note that some theories distinguish not only between permissible and impermissible options, but also between "merely permissible" options and "supererogatory" ones, which are beyond the call of duty. We set this complication aside, though our framework can be extended to accommodate it.[11])

The present formalism permits a variety of interpretations. Contexts can be specified as richly as needed for an adequate description of the choice situation.[12] Even the agent's identity can in principle be built into the notion of a context, as discussed in Section 5.3. Similarly, options can be specified in a variety of ways, though it is best not to pack contextual features into the specification of the options themselves. Contextual features should be included in the specification of the contexts in which those options are available. It will then be possible, at least theoretically, to encounter the same option in more than one context. This makes it meaningful to ask questions such as the following: "would option x, which is permissible in context K, still be permissible in a different context K'?" Later, we explicitly distinguish between properties that options have intrinsically and properties they have in relation to the context.

3.2 Consequentialization revisited

Our goal is to find a canonical way of systematizing or explaining the permissibility verdicts encoded by a given rightness function. One approach to this problem, familiar from the literature on consequentialization, is to try to identify a binary relation over the options, ideally an ordering, such that, for any context, the permissible options are the highest-ranked available options according to that relation. Formally, given a rightness function R, we are looking for a binary relation $\succeq$ on the set X such that, for any context K,

$$R(K) = \{x \in [K] : x \succeq y \text{ for all } y \in [K]\},$$

[11] Some of the permissible options may stand out as "saintly" or "heroic" (using the terms of Urmson 1958). We can capture this by introducing a *saintly or heroic choice function* H which assigns to each context K the set $H(K)$ of those options that a saint or hero would choose, subject to the constraint that $H(K) \subseteq R(K)$. For some contexts K, $H(K)$ may be a *proper* subset of $R(K)$ (i.e., $H(K) \subsetneq R(K)$), so that we can interpret the options in $H(K)$ as the supererogatory ones. Here the saint or hero does something that goes beyond the call of duty. For other contexts K, $H(K)$ and $R(K)$ may coincide. A saint's or hero's choices will not always differ from those of an "ordinary" moral agent; sometimes there is no scope for supererogation. An option is *supererogatory* in context K if it is in $H(K)$ *and* $H(K) \subsetneq R(K)$. There is some debate about whether the phenomenon of supererogation exists, i.e., in the present terms, whether there is an H function as distinct from the R function. See Heyd (2015).

[12] To make this explicit, we could define a context K as a pair $\langle Y, \Phi \rangle$ of (i) a set Y of available options (i.e., $[K] = Y$) and (ii) a set Φ of other contextual features (such as the time, the past history, certain background facts, the cultural environment, or even facts about the agent).

where "$x \succeq y$" means "x beats, or ties with, y" or "x is ranked weakly above y", according to $\succeq$.[13] When there exists such a relation, we say that the rightness function is *representable by a binary relation*. Although this definition does not require $\succeq$ to be transitive and reflexive, the case in which it has these properties is the most important one. The relation $\succeq$ is then naturally interpretable as a *betterness relation* (recall that betterness is commonly taken to be transitive), and it is conventional to call the representation *structurally consequentialist* (see, e.g., Brown 2011). A normative theory is *consequentializable* if its rightness function admits a structurally consequentialist representation. (Some philosophers, such as Broome, use the term "teleological" instead of "consequentialist" to refer to this notion. In Section 5, we formally distinguish between "consequentialism" and "teleology". Arguably, the conventional definition of structural consequentialism combines both notions.)

When can a rightness function be represented by a binary relation? Since rightness functions are formally the same as choice functions, we can bring a large body of work in decision theory to bear on this question (as in Brown's 2011 analysis). In particular, there are well-known necessary and sufficient conditions for the representability of a choice function by a binary relation, under a variety of constraints on that relation. These results apply equally to rightness functions. In Appendix A, we state one illustrative such representation theorem, which gives us a precise dividing line between those rightness functions that can be represented by a binary relation, and those that cannot. The key point is that *if, and only if*, a rightness function satisfies a particular structural condition, it can be represented by a binary relation.

3.3 Two problems

There are at least two problems with the attempt to systematize or explain rightness functions by representing them in terms of binary relations over the options. The first is that some reasonable rightness functions cannot be represented in this way. For a simple example (due to Amartya Sen 1993), consider a rightness function that encodes norms of politeness. When you are offered a choice between different pieces of cake at a dinner party, politeness commands that you do not choose the biggest piece, because that would be greedy. So, when a big (x), a medium-sized (y), and a small piece of cake (z) are available, it is permissible to choose any of the three, except the biggest. In particular, choosing the medium-sized piece is perfectly ok. Formally, if $[K] = \{x, y, z\}$,

[13] The relation $\succeq$ induces an asymmetrical ("strict") relation $\succ$ and a symmetrical ("indifference") relation $\sim$. For any x and y, we have $x \succ y$ if and only if $x \succeq y$ and not $y \succeq x$, and we have $x \sim y$ if and only if $x \succeq y$ and $y \succeq x$. For any $x \in [K]$, we call x *highest-ranked* in K if $x \succeq y$ for all $y \in [K]$.

then $R(K) = \{y, z\}$. However, when the big piece is unavailable (so you are now choosing between the two smaller pieces), then choosing the medium-sized piece (y) is no longer permissible, because it is now the biggest on offer. Formally, if $[K'] = \{y, z\}$, then $R(K') = \{z\}$. No binary relation over the different pieces of cake could represent this rightness function. For a binary relation to represent it, it would have to place z ahead of y and also not do so.[14]

Of course, one might try to respond to this problem by redescribing the options in a richer way, but this response is problematic too. If we build the entire choice context into the description of the options, so that each option can occur in only one context, we can trivially represent *any* rightness function in terms of an artificially constructed binary relation (as shown in Appendix A), but that binary relation will be completely uninformative. It will simply be a cumbersome redescription of the rightness function itself, enumerating all its recommendations in a relation-theoretic format. Moreover, it will still be true that the rightness function *as originally defined* admits no representation in terms of a binary relation.

The second problem with the attempt to explain rightness functions in terms of binary relations is that, even when such a representation exists, it tells us very little about the *reasons* underpinning the permissibility verdicts encoded by the given rightness function. Suppose we are asked: "why is x permissible and y is not?" If we simply say "because x is better than y", this is not very informative. It would be legitimate to ask a further question, namely: "what is it about x and y that makes x better than y?" A representation of a rightness function in terms of a binary relation is silent on that further question. We would like go beyond merely enumerating "brute" goodness facts or betterness facts; we would like to say something about each option's right-making or wrong-making features.

4 Reason-based representation

As announced, our idea is to represent any moral theory in terms of two parameters:

(i) a specification of which properties of the options matter in any given context, and

(ii) a specification of how these properties matter.

To make this precise, we first introduce the notion of a *property* and give a taxonomy of different kinds of properties. We then define the notion of a *reasons structure*, which is

[14] We have to explain (1) $R(K) = \{y, z\}$ and (2) $R(K') = \{z\}$. A necessary condition for explaining (1) is $y \sim z$, while a necessary condition for explaining (2) is $z \succ y$. We cannot have both.

our formalization of (i) and (ii). We finally explain how a reasons structure entails, and thereby explains, a rightness function.[15]

4.1 Properties

At a first gloss, a property is a feature that an option may or may not have, so that a property picks out the set of those options that have that property. For example, if the options are possible meal choices, the property *vegetarian* picks out the set of those meals that involve no meat. As will become clear, however, this understanding of properties is not sufficiently general for our purposes.

Instead of taking properties to be features of options *simpliciter*, we take them to be features of option-context pairs. An *option-context pair* is a pair of the form $\langle x, K\rangle$, where x is an option (an element of X) and K is a context (an element of $\mathcal{K}$). Formally, a *property* is a primitive object P that picks out a set of option-context pairs, called the *extension* of P and denoted $[P]$. Whenever a pair $\langle x, K\rangle$ is contained in $[P]$, this means that option x has property P in context K. Sometimes we also say: the option-context pair $\langle x, K\rangle$ has property P. A property P may be of three different kinds:[16]

- P is an *option property* if its possession by an option-context pair depends only on the option, not on the context.[17]
- P is a *context property* if its possession by an option-context pair depends only on the context, not on the option.[18]
- P is a *relational property* if its possession by an option-context pair depends on both the option and the context.[19]

Suppose, for example, that X is a set of meal choices, and $\mathcal{K}$ is a set of menus from which one may choose. The property *vegetarian* is an option property. If a particular meal option is vegetarian in the context of one menu, then it is still vegetarian in the context of another. The property *offering two or more options* is a context property. Whether an option-context pair has that property depends solely on the context (here the menu), irrespective of the particular option we are considering. Finally, the property *most calorific* is a relational property, since the same meal option can be most calorific

[15] Our formalism draws on the positive, decision-theoretic framework in Dietrich and List (2016).

[16] In what follows, we will only consider properties whose extension is "non-trivial", in that it is neither empty nor total: i.e, at least one, but not all, option-context pairs have the property in question.

[17] Formally, for all x in X and all K, K' in $\mathcal{K}$, $\langle x, K\rangle \in [P]$ if and only if $\langle x, K'\rangle \in [P]$.

[18] Formally, for all K in $\mathcal{K}$ and all x, x' in X, $\langle x, K\rangle \in [P]$ if and only if $\langle x', K\rangle \in [P]$.

[19] Formally, it is neither an option property nor a context property.

relative to one menu, but not relative to another. Likewise, properties such as *polite* or *norm-conforming* are relational properties, since the same act can be polite or norm-conforming in one context, but not in another. Think about the difference between public and private contexts or between contexts involving different cultures.

We write $\mathcal{P}$ to denote the set of those properties that may be candidates for normatively relevant properties: we call these the *admissible properties*. If there are no constraints on the properties that might turn out to be normatively relevant according to some moral theory, then $\mathcal{P}$ could, in principle, be the universal set of all logically possible properties (which contains at least one property for every possible extension). If, on the other hand, some properties are so far-fetched that they could never be normatively relevant, then $\mathcal{P}$ could be more restricted.[20]

4.2 The notion of a reasons structure

We can now define the notion of a reasons structure. It specifies what the normatively relevant properties in each context are and which sets of properties "outweigh" or "defeat" which others. Formally, a *reasons structure* is a pair $\mathcal{R} = \langle N, \trianglerighteq \rangle$ consisting of:

- A *normative relevance function*, denoted N, which assigns to each context $K \in \mathcal{K}$ a set $N(K)$ of *normatively relevant* properties in that context.[21]

- A *weighing relation* (or *defeat relation*) over sets of properties, denoted $\trianglerighteq$, formally a binary relation whose relata are subsets of $\mathcal{P}$. When one set of properties S stands in this relation to another set S', formally $S \trianglerighteq S'$, we say that S *weakly outweighs* S', or S is *ranked weakly above* S', or S *defeats* S'.[22]

For example, in the case of a utilitarian theory, the function N assigns to every context the set of all "welfare properties"; and the relation $\trianglerighteq$ ranks sets of such properties by stipulating that more welfare is better than less; we make this more precise below.

To regiment our formalism, we impose one invariance constraint on the normative relevance function N: whenever two contexts K and K' have the same context properties,

[20] The set $\mathcal{P}$ can be partitioned into the subsets $\mathcal{P}_{\text{option}}$, $\mathcal{P}_{\text{context}}$, and $\mathcal{P}_{\text{rel}}$ of all option, context, and relational properties in $\mathcal{P}$, respectively. For any option x and any context K, we write $\mathcal{P}(x, K)$ for the set of all properties of $\langle x, K \rangle$ (among those in $\mathcal{P}$); $\mathcal{P}(x) = \mathcal{P}(x, K) \cap \mathcal{P}_{\text{option}}$ for the set of all option properties of x; and $\mathcal{P}(K) = \mathcal{P}(x, K) \cap \mathcal{P}_{\text{context}}$ for the set of all context properties of K.

[21] Formally, N is a function from $\mathcal{K}$ into $2^{\mathcal{P}}$.

[22] Some of these interpretations work best if $\trianglerighteq$ is transitive and reflexive, as discussed below. We write $\triangleright$ and $\equiv$ for the asymmetrical ("strict") and symmetrical ("indifference") relations induced by $\trianglerighteq$. We are grateful to John Broome for suggesting the "defeat" interpretation.

then the same properties are normatively relevant in those contexts.[23] By contrast, we initially impose no restrictions on the relation $\trianglerighteq$. It could fall well short of the requirements of an ordering; it may be incomplete and even intransitive, for example. However, if one wishes to interpret this relation as capturing betterness comparisons among sets of properties and one accepts the view that "betterness" is transitive, then one might restrict attention to transitive relations.

We use the term "weighing relation" rather than "defeat relation" to refer to $\trianglerighteq$ throughout this paper and frequently speak of "weighing" and "ranking". This terminology is most natural when the relation $\trianglerighteq$ is transitive and reflexive. In the absence of these properties, the relation $\trianglerighteq$ can simply be interpreted as specifying which sets of properties "defeat" which others. Our next step is to explain how a reasons structure entails permissibility verdicts.

4.3 The entailed rightness function

In any context K, the question is: which options among the available ones are permissible, according to a given reasons structure $\mathcal{R}$? To answer this question, we look at the available options through the lens of their normatively relevant properties, as picked out by the function N. Specifically, for each option x and each context K, we write $\mathcal{P}(x, K)$ to denote the set of *all* properties of this option-context pair (among the properties in $\mathcal{P}$). Since $N(K)$ is the set of all *normatively relevant* properties in context K, the normatively relevant properties of option x in context K are obviously those that lie in the intersection of $\mathcal{P}(x, K)$ and $N(K)$, formally

$$\mathcal{P}(x, K) \cap N(K).$$

Let $N(x, K)$ denote this set. We then assess different available options by comparing their sets of normatively relevant properties, using the weighing relation $\trianglerighteq$. More precisely, option x beats or at least ties with option y in context K if and only if the set $N(x, K)$ at least weakly outweighs the set $N(y, K)$, i.e., $N(x, K) \trianglerighteq N(y, K)$. Then the permissible or "right" options are the ones that beat, or tie with, all available options. In short, for an option to be permissible, its set of normatively relevant properties must weakly outweigh the corresponding set for every available option. Formally:

$$R(K) = \{x \in [K] : N(x, K) \trianglerighteq N(y, K) \text{ for all } y \in [K]\}.$$

[23] If we assert that different properties are normatively relevant in contexts K and K', then we should be able to point to some properties of those contexts to which this difference can be attributed. If two contexts are property-wise indistinguishable, then it is hard to explain how they could give rise to different normatively relevant properties. Formally, if $\mathcal{P}(K) = \mathcal{P}(K')$, then $N(K) = N(K')$.

We call the function R thus defined the *rightness function entailed by the reasons structure* $\mathcal{R}$.[24] (In Appendix B, we consider some alternative ways in which a reasons structure may entail a rightness function, e.g., by deeming an option permissible if and only if its set of normatively relevant properties is not strictly outweighed by the corresponding set for any available option.)

Let us give some informal illustrations. First, recall our example of a utilitarian theory. Here, the set $N(K)$ of normatively relevant properties is always the set of all "total welfare properties", i.e., all properties of the form "the total welfare produced by the option's consequences is such-and-such". The weighing relation $\trianglerighteq$ then ranks property sets in terms of total welfare. So, the singleton property set {"the total welfare is 15"} outweighs the set {"the total welfare is 10"}, for instance. In each context K, the function R now selects the welfare-maximizing option(s) among the available ones.

Next consider a theory which defines permissibility in terms of the minimization of rights violations; Robert Nozick (1974) calls this "utilitarianism of rights". Here, in each context, the set $N(K)$ consists of those "rights-violation properties" in terms of which options are to be assessed. For any property P in $N(K)$, an option's having that property in context K means that, by choosing that option in that context, the agent would violate some right. The weighing relation may then rank sets of properties by size: i.e., one set of properties is ranked above a second set if the first set is smaller than the second (i.e., it consists of fewer rights-violation properties than the second). For a more strictly deontological theory, we may replace this weighing relation with one in which only the empty set of properties – standing for no rights violations – is ranked above every set of properties, while no other sets of properties are weakly or strictly ranked above any sets of properties (yielding a non-reflexive relation). In this case, only options that involve no rights violations are ever deemed permissible, so $R(K)$ can sometimes be the empty set.

We say that a rightness function has a *reason-based representation* if there exists some reasons structure that entails it.[25] In Appendix B, we show that every rightness function within a very large class can be represented in this way, including rightness functions that defy consequentialization in the conventional sense.[26] Reason-based representations

[24] To refer to this rightness function, we sometimes also use the notation $R^{\mathcal{R}}$.

[25] Formally, there exists some reasons structure $\mathcal{R} = \langle N, \trianglerighteq \rangle$ such that $R = R^{\mathcal{R}}$.

[26] The more properties we are willing to invoke in constructing a reason-based representation, the more rightness functions we can represent. If the set $\mathcal{P}$ of admissible properties contains all logically possible properties, then every logically possible rightness function – however far-fetched – can be *formally* represented by a reasons structure. Of course, far-fetched rightness functions may be representable only in a gerrymandered and unilluminating way, as they may fail to correspond to any plausible moral theories.

avoid the two problems of standard consequentialist representations we have discussed. First, the rightness functions of interesting moral theories can be represented in a reason-based format, regardless of whether they are plausibly consequentializable. Second, a reason-based representation encodes not only the entailed rightness function itself, but also the underlying right-making features. Our framework thus allows us to represent not only a moral theory's deontic content, but also its explanation for it.

5 A taxonomy of moral theories

As we have seen, a moral theory can be represented canonically as a reasons structure, which, in turn, entails the theory's rightness function. We now show that this way of representing moral theories yields a very general taxonomy. In particular, we characterize moral theories in terms of six key distinctions. The first four refer to a theory's normative relevance function (i.e., they have to do with *which properties matter*); the last two refer to its weighing relation (i.e., they have to do with *how these properties matter*).

5.1 Consequentialist versus non-consequentialist theories

We suggest that the distinction between consequentialist and non-consequentialist theories can be drawn in terms of the nature of the properties these theories deem normatively relevant. On a simple definition, a theory is *consequentialist* if, according to that theory, only consequences matter;[27] it is non-consequentialist if, according to it, some things other than consequences matter, at least sometimes. We formalize this as follows.

A theory with reasons structure $\mathcal{R} = \langle N, \trianglerighteq \rangle$ is *structurally consequentialist* if it never deems any properties other than option properties normatively relevant; i.e., for every context K, $N(K)$ consists of option properties alone. The theory is *structurally non-consequentialist* if it sometimes deems relational or context properties normatively relevant; i.e., for some context K, $N(K)$ includes some relational or context property. We also call such properties *context-related properties*.

One might quarrel whether our distinction between option properties and context-related properties matches the standard distinction between "consequence properties" and "non-consequentialist properties". However, we suggest that any doubts on this

For example, it would be unilluminating to explain the permissibility of option x in context K simply by saying that the singleton property set {"being option x in context K"} outweighs {"being option y in context K"} for all available options y. We may therefore exclude properties like "being option x in context K" from $\mathcal{P}$. By imposing reasonable constraints on $\mathcal{P}$, we may rule out implausible reasons structures.

[27] Sometimes a distinction is drawn between *causal* and *constitutive* consequences of an action; see, e.g., Dreier (2011). Our formal analysis below is ecumenical.

front should lead us, not so much to challenge our distinction, but rather to specify the options in X and the contexts in $\mathcal{K}$ more carefully. A plausible requirement is that, for any option-context pair $\langle x, K\rangle$, x should encode everything that is deemed to belong to the consequences of choosing x, while K should encode only the situation in which the choice takes place, including the range of alternative options. If our specification of the options and contexts meets this requirement, then the distinction between option properties and context-related properties will be aligned with the distinction between "consequence properties" and "non-consequentialist properties".

Some examples help to illustrate our distinction between consequentialist and non-consequentialist theories. Suppose the options are welfare distributions across a society of n individuals. So, each option is some n-tuple of the form $\langle w_1, w_2, ..., w_n\rangle$, where w_i is the welfare level of the i^{th} individual. The agent might be a social planner choosing among these options. Consider three theories.

- **Utilitarianism:** This says that we should choose a distribution $\langle w_1, w_2, ..., w_n\rangle$ which maximizes the *total welfare*, defined as $w_1 + w_2 + ... + w_n$.
- **Entitlement satisfaction:** Here, in each context K, there is an n-tuple of entitlements, $\langle e_1, e_2, ..., e_n\rangle$, where e_i is the welfare level to which the i^{th} individual is entitled (perhaps on grounds of effort or desert); the n-tuple of entitlements may differ from context to context.[28] Now the theory says that we should choose a distribution $\langle w_1, w_2, ..., w_n\rangle$ that maximizes the number of individuals whose entitlements are satisfied, i.e., for whom $w_i \geq e_i$.[29]
- **Satisficing utilitarianism:** This says that we should choose a distribution $\langle w_1, w_2, ..., w_n\rangle$ in which the total welfare is at least 0.8 times as great as the total welfare in every available alternative.[30]

All of these theories can be easily represented in our framework. To represent utilitarianism, we must invoke properties of the form

$$P_{\text{wel}=w}: \text{ "The total welfare is } w\text{"},$$

where w is some real number. Call such properties *total-welfare properties.* The reasons structure is $\mathcal{R} = \langle N, \trianglerighteq\rangle$, where:

[28] Formally, we can think of each context as a pair $\langle Y, \langle e_1, e_2, ..., e_n\rangle\rangle$, where Y is the set of available options and $\langle e_1, e_2, ..., e_n\rangle$ is the context-specific n-tuple of entitlements.

[29] If there are ties between distributions under this criterion, we might add some tie-breaking criterion, for instance one that requires minimizing the sum-total shortfall.

[30] Brown (2011) and Dreier (2011) discuss this as an example of a non-consequentialist theory.

- for every context K, $N(K)$ is the set of all total-welfare properties;
- the weighing relation ranks singleton sets of total-welfare properties such that $\{P_{\text{wel}=w}\} \trianglerighteq \{P_{\text{wel}=w'}\}$ if and only if $w \geq w'$.

To represent the entitlement-satisfaction theory, we must invoke properties of the form

$$P_i: \text{ “The entitlement of individual } i \text{ is satisfied”},$$

where i ranges over the n individuals in the society under consideration. Call such properties *entitlement-satisfaction properties*. The reasons structure is $\mathcal{R} = \langle N, \trianglerighteq \rangle$, where:

- for every context K, $N(K)$ is the set of all entitlement-satisfaction properties;
- the weighing relation is defined as follows: for any two sets of entitlement-satisfaction properties – call them S and S' – we have $S \trianglerighteq S'$ if and only if S contains at least as many entitlement-satisfaction properties as S'.[31]

Finally, to represent satisficing utilitarianism, we must invoke the following property:

$$P_{\text{satisf}}: \text{ “The total welfare is at least 0.8 times as great as in every alternative”}.$$

Call this the *satisficing property*. The reasons structure is simply $\mathcal{R} = \langle N, \trianglerighteq \rangle$, where:

- for every context K, $N(K)$ consists of the satisficing property alone;
- the weighing relation ranks $\{P_{\text{satisf}}\}$ above the empty set.

It is easy to see that utilitarianism is consequentialist in the sense we have defined, while the other two theories are not. This is because the normatively relevant properties according to utilitarianism, total-welfare properties, are option properties, while the normatively relevant properties according to the other theories (entitlement-satisfaction properties and the satisficing property) are not. Whether a welfare distribution has a particular total-welfare property – it offers such-and-such total welfare – depends only on the distribution itself. By contrast, whether the individuals' entitlements are satisfied depends on the n-tuple of entitlements in the context: it is a relational property. Likewise, whether a distribution's total welfare is at least 0.8 times as great as that in

[31] To include the tie-breaking criterion from footnote 29, we must add to each set $N(K)$ all properties of the form $P_{\text{sh}=\delta}$ (“the total shortfall is δ”) for some $\delta \geq 0$, where the *total shortfall* for any distribution $\langle w_1, w_2, ..., w_n \rangle$ relative to $\langle e_1, e_2, ..., e_n \rangle$ is $\sum_{i=1}^{n} \min(0, e_i - w_i)$. The weighing relation must then be as follows: $S \trianglerighteq S'$ whenever (i) S contains more entitlement-satisfaction properties than S', *or* (ii) S and S' contain equally many entitlement-satisfaction properties and $\delta \leq \delta'$, where $P_{\text{sh}=\delta} \in S$ and $P_{\text{sh}=\delta'} \in S'$.

every alternative depends on the other available distributions.[32] Along similar lines, theories that deem some "essentially comparative" properties normatively relevant, as suggested by Larry Temkin (1996, 2012), qualify as non-consequentialist. We give an example later.

Importantly, we have not built transitivity of the weighing relation into our definition of consequentialism. Although a transitive ordering of consequences is sometimes considered one of the defining characteristics of consequentialism (as in the conventional definition discussed in Section 3.2), the question of whether the normatively relevant properties are restricted to option properties is orthogonal to the question of whether the weighing relation is transitive (and/or reflexive). We take the former question to pick out the difference between consequentialism and non-consequentialism, and the latter to pick out the difference between teleology and non-teleology; more on this in Section 5.5.

5.2 Universalist versus relativist theories

Informally, universalism can be understood as the view that what matters is always the same, regardless of the context in which the assessment takes place, while relativism can be understood as the view that different things matter in different contexts. For example, different cultural contexts or social practices may make different properties normatively relevant. Later we comment on a special form of relativism, *agent-relativity*.

Our framework offers a natural formalization of the distinction between universalism and relativism. A theory with reasons structure $\mathcal{R} = \langle N, \trianglerighteq \rangle$ is *structurally universalist* if the set of normatively relevant properties $N(K)$ is the same across all contexts; i.e., the normative relevance function N is constant. The theory is *structurally relativist* if $N(K)$ is not the same across all contexts; i.e., N is non-constant. Different contexts render different properties normatively relevant.

The three theories discussed in the last subsection, utilitarianism, entitlement satisfaction, and satisficing utilitarianism, are all examples of structurally universalist theories; they each specify a constant set of normatively relevant properties: the set of all total-welfare properties, the set of all entitlement-satisfaction properties, and the satisficing property, respectively. If we constructed a theory that deems welfare properties relevant in some contexts and entitlement-satisfaction properties in others, this would be structurally relativist. Communitarians sometimes endorse structurally relativist theories, insofar as they take the normatively relevant properties to depend on social conventions and social meanings in the contexts in question.

[32] These classifications of the properties are supported by our formal definitions in footnotes 17-19.

While the present distinction between universalism and relativism focuses on the normative relevance function, one might also draw a similar distinction along a second dimension. Recall that, under our definition of a reasons structure, the weighing relation does not depend on the context. One might amend the definition by permitting a context-dependent weighing relation. We might then call a moral theory *weighing-universalist* if the weighing relation does not vary across contexts, and *weighing-relativist* if it does. This new distinction is orthogonal to our earlier one between what one might call *normative-relevance universalism* and *normative-relevance relativism.* In principle, we can thus recognize two dimensions on which theories may be universalist or relativist. Since this richer classification requires a less parsimonious definition of a reasons structure, however, we set it aside for now. In Appendix C, we return to it and examine the extent to which weighing relativism can be formally re-modelled in a weighing-universalist format (e.g., by shifting the relativism from the weighing relation to the normative relevance function). In what follows, the terms "universalism" and "relativism" will continue to refer to universalism and relativism with respect to the normative relevance function.

It is worth noting that the two main distinctions we have drawn so far, namely between consequentialism and non-consequentialism and between universalism and relativism, each concern the question of whether a theory's reasons structure is context-dependent in a particular way. The first distinction – between consequentialism and non-consequentialism – concerns the question of whether the normatively relevant properties include context-related properties. Call a reasons structure *context-related* if the answer to this question is positive, and *context-unrelated* otherwise. The second distinction – between universalism and relativism – concerns the question of whether the normatively relevant properties vary across contexts. Call a reasons structure *context-variant* if the answer to this second question is positive, and *context-invariant* otherwise. Table 1 shows how these two distinctions can be combined.

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Context-variant?</td></tr>
<tr><td>No</td><td>Yes</td></tr>
<tr><td rowspan="2">Context-related?</td><td>No</td><td>Universalist consequentialism
(e.g., utilitarianism)</td><td>Relativist consequentialism
(e.g., practice-dependent theories)</td></tr>
<tr><td>Yes</td><td>Univ. non-consequentialism
(e.g., entitlement satisfaction, deontological theories)</td><td>Relativist non-consequentialism
(e.g., cultural-norms-based theories)</td></tr>
</table>

Table 1: Two kinds of context-dependence

5.3 Agent-neutral versus agent-relative theories

Roughly speaking, a moral theory is agent-neutral if the identity of the agent makes no difference to its prescriptions (other things being equal), while it is agent-relative if it does. Utilitarianism is the standard example of an agent-neutral theory, while ethical egoism – which recommends that each agent should pursue his or her own self-interest – is a familiar example of an agent-relative one: its action-guiding recommendations depend on who the agent is. There are many ways of making the distinction between agent-neutrality and agent-relativity precise (see, e.g., Ridge 2011 and Schroeder 2007).

To explicate the distinction in our framework, we first revisit the notion of a context. We have already noted that contexts can be specified as richly as needed for an adequate description of the choice situation, and we have acknowledged that even the agent's identity can be built into the context. We may thus think of a context K as a triple $\langle i, Y, \Gamma\rangle$, where i is the agent, Y is the set of available options (i.e., $[K] = Y$), and Γ is a set of other situational or environmental features.[33] We can now refine our taxonomy of two kinds of context-dependence from the end of the last subsection.

We begin by reconsidering the distinction between context-invariant and context-variant reasons structures, our dividing line between universalist and relativist theories.[34] Recall that a reasons structure is context-invariant if the normatively relevant properties – those in $N(K)$ – do not vary across contexts, and context-variant if they do. If we take each context K to be a triple $\langle i, Y, \Gamma\rangle$, we can refine this distinction by asking

(i) whether $N(K)$ varies with changes in the agent i;[35]

(ii) whether $N(K)$ varies with changes in the set of available options, Y;[36] and

(iii) whether $N(K)$ varies with changes in the other situational or environmental features, as specified by Γ.[37]

The answer to question (i) yields the distinction between *agent-variant* and *agent-invariant* reasons structures; the answer to question (ii) yields the distinction between *menu-variant* and *menu-invariant* reasons structures (where the "menu" is the set of available options); and the answer to question (iii) yields the distinction between *situation/environment-variant* and *situation/environment-invariant* reasons structures.

[33] In footnote 12, we suggested thinking of a context as a pair $\langle Y, \Phi\rangle$ of a set of available options and a set Φ of other contextual features. We can now interpret Φ as subsuming i and Γ.

[34] See Appendix C for a discussion of an agent-relative weighing relation.

[35] Formally, are there $K=\langle i, Y, \Gamma\rangle$ and $K'=\langle i', Y', \Gamma'\rangle$, with $Y=Y'$ and $\Gamma=\Gamma'$, such that $N(K)\neq N(K')$?

[36] Formally, are there $K=\langle i, Y, \Gamma\rangle$ and $K'=\langle i', Y', \Gamma'\rangle$, with $i=i'$ and $\Gamma=\Gamma'$, such that $N(K)\neq N(K')$?

[37] Formally, are there $K=\langle i, Y, \Gamma\rangle$ and $K'=\langle i', Y', \Gamma'\rangle$, with $i=i'$ and $Y=Y'$, such that $N(K)\neq N(K')$?

A theory that deems different properties normatively relevant depending on the social or cultural context exhibits situation/environment-variance. A theory that instructs different agents to focus on different properties exhibits agent-variance. Ethical egoism, for instance, may be an illustration, deeming only individual i's welfare properties normatively relevant for individual i. Accordingly, we might define *agent-neutrality* as the absence of agent-variance in a theory's reasons structure, and *agent-relativity* as the presence of such agent-variance. Since agent-variance is a special case of context-variance, agent-relativity is thus a special form of relativism.

However, there is also another way of distinguishing between agent-neutral and agent-relative theories. To see this, let us go back to the distinction between context-unrelated and context-related reasons structures, our dividing line between consequentialist and non-consequentialist theories. Recall that a reasons structure is context-unrelated if the normatively relevant properties – those in $N(K)$ – are always option properties, while it is context-related if they include context-related properties, at least sometimes. And recall that a property P is context-related if its possession by an option-context pair depends, at least in part, on the context.[38] If we interpret each context K as a triple $\langle i, Y, \Gamma \rangle$, we can refine these definitions too:

(i) a property P is *agent-related* if its possession by an option-context pair depends, at least in part, on the agent i;[39]

(ii) a property P is *menu-related* if its possession by an option-context pair depends, at least in part, on the set of available options, Y;[40]

(iii) a property P is *situation/environment-related* if its possession by an option-context pair depends, at least in part, on the other situational or environmental features, as specified by Γ.[41]

Call a reasons structure *agent-related*, *menu-related*, and *situation/environment-related* if the normatively relevant properties in $N(K)$ include, respectively, agent-related, menu-related, and situation/environment-related properties, at least for some K. The reasons structure of utilitarianism is free from any such properties. By contrast, the reasons

[38] Formally, for some x in X and some K, K' in $\mathcal{K}$, $\langle x, K \rangle \in [P]$ and $\langle x, K' \rangle \notin [P]$.

[39] Formally, for some x in X and some $K = \langle i, Y, \Gamma \rangle, K' = \langle i', Y', \Gamma' \rangle$ in $\mathcal{K}$ with $Y = Y'$ and $\Gamma = \Gamma'$, $\langle x, K \rangle \in [P]$ and $\langle x, K' \rangle \notin [P]$.

[40] Formally, for some x in X and some $K = \langle i, Y, \Gamma \rangle, K' = \langle i', Y', \Gamma' \rangle$ in $\mathcal{K}$ with $i = i'$ and $\Gamma = \Gamma'$, $\langle x, K \rangle \in [P]$ and $\langle x, K' \rangle \notin [P]$.

[41] Formally, for some x in X and some $K = \langle i, Y, \Gamma \rangle, K' = \langle i', Y', \Gamma' \rangle$ in $\mathcal{K}$ with $i = i'$ and $Y = Y'$, $\langle x, K \rangle \in [P]$ and $\langle x, K' \rangle \notin [P]$.

structure of the entitlement-satisfaction theory is situation/environment-related, as the entitlement-satisfaction properties are situation/environment-related. Whether someone's entitlements are satisfied depends on what he or she is entitled to in the given situation. The reasons structure of satisficing utilitarianism is menu-related, since the satisficing property is menu-related. Whether the sum-total welfare in a given distribution is at least 0.8 times as great as that in every alternative depends on the available "menu" of distributions. However, none of these reasons structures deem any agent-related properties normatively relevant. We might define *agent-neutrality* as the absence of agent-relatedness in a theory's reasons structure, and *agent-relativity* as the presence of agent-relatedness. Defined in this way, agent-relativity is a special form of non-consequentialism, not of relativism.

To illustrate these two orthogonal ways of defining the distinction between agent-neutrality and agent-relativity, consider ethical egoism. Specifically, we use the setup of our earlier examples where the options are welfare n-tuples of the form $\langle w_1, w_2, ..., w_n \rangle$. We now assume that the agent is one of the n individuals. Ethical egoism requires individual i to choose a distribution which maximizes w_i. According to one reason-based representation, the theory is agent-relative in the relativist sense (rather than the non-consequentialist one). Here, we invoke properties of the form

$$P_{\text{wel}_i=w} : \text{ "Individual } i\text{'s welfare is } w\text{"},$$

where i is some individual and w is some real number. Call these the *welfare properties for individual i*. Now the reasons structure is $\mathcal{R} = \langle N, \trianglerighteq \rangle$, where:

- for every context K, $N(K)$ is the set of all welfare properties for the individual i named in $K = \langle i, Y, \Gamma \rangle$.
- the weighing relation ranks singleton sets consisting of individual welfare properties such that $\{P_{\text{wel}_i=w}\} \trianglerighteq \{P_{\text{wel}_i=w'}\}$ if and only if $w \geq w'$.

Since $N(K)$ varies with the agent, the reasons structure is agent-variant and thus context-variant, our defining condition of relativism. At the same time, $N(K)$ contains only option properties. Thus the reasons structure is context-unrelated, our defining condition of consequentialism. Contrast this with a second representation of ethical egoism, according to which the theory is agent-relative in the non-consequentialist sense. Here, we invoke properties of the form

$$P_{\text{own wel}=w} : \text{ "The agent's welfare (or perhaps 'my' welfare) is } w\text{"},$$

where w is some real number. Call these the *agent-centred welfare properties.* Now the reasons structure is $\mathcal{R} = \langle N, \trianglerighteq \rangle$, where:

- for every context K, $N(K)$ is the set of all agent-centred welfare properties;
- the weighing relation ranks singleton sets consisting of agent-centred welfare properties such that $\{P_{\text{own wel}=w}\} \trianglerighteq \{P_{\text{own wel}=w'}\}$ if and only if $w \geq w'$.

Note that $N(K)$ does not vary with the identity of the agent. Regardless of who makes the choice, the theory always deems all agent-centred welfare properties normatively relevant. In fact, the reasons structure is completely context-invariant, and thus universalist as we have defined it. At the same time, the properties contained in $N(K)$, agent-centred welfare properties, are agent-related: options do not have these properties "intrinsically", but only in relation to the agent in question.[42] The reasons structure is therefore context-related, our defining condition of non-consequentialism.

For a further example of agent-relativity as a form of non-consequentialism, consider Bernard Williams's famous thought experiment of "Jim and the Indians" (1973). An evil army captain is about to kill ten innocent people. Jim passes by, and the captain offers Jim the option of killing one of the ten while the captain would then spare the rest. Should Jim accept the offer? According to a consequentialist theory which says "minimize the number of deaths", he ought to do so. By contrast, according to a theory which prohibits killing and draws a strong act-omission distinction, Jim ought to decline the offer, so as to avoid being the perpetrator of an act of killing.

To formalize this example, let choice contexts be of the form $K = \langle i, Y \rangle$, where i is the agent (such as "Jim" or "Captain") and Y is the set of feasible options. For simplicity, we take options to be of the form $\langle n, j \rangle$, where n is the number of deaths and j is the perpetrator of the killing (such as "Jim" or "Captain"). Note that the agent of the moral choice (i) need not be the same as the perpetrator of the killing (j). Jim faces the tragic choice between accepting the captain's offer, in which case Jim himself would kill one person ($\langle 1, \text{Jim} \rangle$), and declining the offer, in which case the captain would kill ten ($\langle 10, \text{Captain} \rangle$). Thus Jim's choice context is $K = \langle i, Y \rangle$, where $i = \text{Jim}$ and $Y = \{\langle 1, \text{Jim} \rangle, \langle 10, \text{Captain} \rangle\}$.[43] We consider two kinds of properties:

$$P_{\text{deaths}=n} : \text{ "The number of deaths is } n\text{" (where } n \geq 0\text{), and}$$

$$P_{\text{act}} : \text{ "The option involves an act of killing by the chooser".}$$

Properties of the first kind are option properties: for each numerical value of n, all options of the form $\langle n, j \rangle$ (and no others) have the property $P_{\text{deaths}=n}$, regardless of the

[42] Formally, whether an option-context pair $\langle x, K \rangle$ has the property $P_{\text{own wel}=w}$ depends on both x and the agent i in the triple $\langle i, Y, \Gamma \rangle = K$.

[43] We can think of $\langle 1, \text{Jim} \rangle$ and $\langle 10, \text{Captain} \rangle$ as the consequences of accepting and declining the offer.

context. The property P_{act} is an agent-related property: an option of the form $\langle n, j\rangle$ (with $n > 0$) has this property in context $K = \langle i, Y\rangle$ if and only if $i = j$.[44]

The consequentialist theory which mandates minimizing the number of deaths can be represented by a reasons structure $\mathcal{R} = \langle N, \trianglerighteq\rangle$, where:

- for every context K, $N(K)$ is the set of all properties of the form $P_{\text{deaths}=n}$;
- the weighing relation places singleton property sets of the form $\{P_{\text{deaths}=n}\}$ in a linear order of decreasing n, i.e., $\{P_{\text{deaths}=n}\}\trianglerighteq\{P_{\text{deaths}=n'}\}$ if and only if $n\leq n'$.

The deontological theory which combines a prohibition on killing with an act-omission distinction can be represented by a reasons structure $\mathcal{R} = \langle N, \trianglerighteq\rangle$, where:

- for every context K, $N(K)$ consists of all properties of the form $P_{\text{deaths}=n}$ and the property P_{act};
- the weighing relation places singleton property sets of the form $\{P_{\text{deaths}=n}\}$ in a linear order of decreasing n, and it ranks any property set containing the property P_{act} strictly below any set not containing that property; furthermore, sets containing the property P_{act} do not weakly outweigh any sets.

The first of these two theories is a regular consequentialist theory; in Jim's context, it mandates accepting the captain's offer ($\langle 1, \text{Jim}\rangle$). The second is a non-consequentialist agent-relative theory: it deems the agent-related property P_{act} relevant. In Jim's context, this theory makes declining the captain's offer the only permissible option, with the result that the captain would kill ten ($\langle 10, \text{Captain}\rangle$). In a context in which the captain is the agent, the option $\langle 10, \text{Captain}\rangle$ is impermissible, as it now has the property P_{act}.[45]

We have identified two ways of drawing the line between agent-neutral and agent-relative theories: we can view agent-relativity either as a special form of relativism, or as a special form of non-consequentialism.

5.4 Monistic versus pluralistic theories

Like the previous distinctions, the next one concerns a theory's normative relevance function. Moral philosophers distinguish between monistic and pluralistic theories. Roughly,

[44] We stipulate that $\langle n, j\rangle$ does not have the property P_{act} if $n = 0$.

[45] One could also formalize the example of "Jim and the Indians" differently, by (i) taking options to encode just the number of deaths, so $X = \{0, 1, 2, ...\}$, and (ii) taking each context K to encode the agent i, the set Y of feasible options, and a specification of which of the feasible options are available through action and which through omission. We do not explore this alternative formalization here.

monistic theories assess options in terms of a single normatively relevant property, while pluralistic theories assess them in terms of two or more normatively relevant properties, which may need to be traded off against one another. In our framework, a theory with reasons structure $\mathcal{R} = \langle N, \trianglerighteq \rangle$ is *structurally monistic* if each option has precisely one normatively relevant property in each context; i.e., $N(x, K)$ is singleton for every option-context pair $\langle x, K \rangle$. The theory is *structurally pluralistic* if some options have two or more normatively relevant properties in some contexts; i.e., there exists at least one option-context pair $\langle x, K \rangle$ for which $N(x, K)$ contains more than one property.[46] Of course, this distinction is useful only to the extent that we attach significance to the individuation of properties; we discuss this issue further in Section 6.4.

Some of our illustrative theories – e.g., utilitarianism, satisficing utilitarianism, and ethical egoism – are monistic in the present sense, assuming we individuate the relevant properties as in our formalizations of those theories. The entitlement-satisfaction theory and our deontological theory that defines permissibility in terms of the avoidance of rights violations are pluralistic; they take options to have more than one normatively relevant property, given the way we have individuated those properties. Similarly, we obtain a pluralistic theory if we add side-constraints to utilitarianism, requiring the maximization of sum-total welfare, subject to some additional requirements. Options would then have two or more normatively relevant properties, such as *their sum-total welfare being such-and-such* and *the side-constraint being satisfied/violated*. Another example is a functioning-or-capability theory, which assesses each option in terms of the functionings or capabilities that it generates. Here, each option could have several functioning-or-capability properties in a given context, which are all normatively relevant.

While the present distinction between monism and pluralism refers to each option's *number* of normatively relevant properties (which is why the distinction is sensitive to how we individuate properties), one could revise the definition of monism by requiring that the normatively relevant properties of all options (and in all contexts) be *of the same kind*. So $N(K)$ would only ever include properties of a single kind: e.g., only welfare properties, or only rights properties, or only duty-properties, and so on. To make this precise, we would have to introduce an equivalence relation over properties, capturing sameness of kind, and demand that, for a monistic theory, the normatively relevant properties always fall into the same equivalence class. This approach would be sensitive to the individuation of property *kinds*, but not to that of *individual* properties.

Finally, we can also represent those pluralistic theories that associate pluralism with

[46] For the purpose of distinguishing between monism and pluralism, we set aside the limiting case in which $N(x, K)$ is empty, i.e., where x has no normatively relevant properties in context K.

moral dilemmas, by deeming some options mutually incomparable in light of their disparate properties. We do not require the weighing relation to be complete (where *completeness* means that no two sets of properties are left mutually unrelated). It is possible, then, that some sets of properties are mutually incomparable. If those properties turn out to be the normatively relevant properties of some available options, this may generate a situation in which the rightness function deems no options permissible.[47]

5.5 Teleological versus non-teleological theories

We have so far focused on distinctions that refer to a theory's normative relevance function; we now turn to some distinctions that refer to the weighing relation.[48] We begin with the distinction between teleological and non-teleological theories. As already noted, a theory is conventionally called *teleological* if its criterion for the permissibility of options is that they are the best feasible ones, according to *some* underlying axiology, though not necessarily an axiology that is concerned with consequences alone.

In our framework, this requires the weighing relation over sets of properties to admit a "betterness" interpretation. If we accept the view that betterness is transitive, then such an interpretation is available only in those cases where the weighing relation is transitive and reflexive; for any sets of properties S and S', "$S \trianglerighteq S'$" should be interpretable as "S is at least as good as S'". We thus call a theory with reasons structure $\mathcal{R} = \langle N, \trianglerighteq \rangle$ *teleological* if $\trianglerighteq$ is transitive and reflexive, and *non-teleological* otherwise.[49]

Practically all the examples of theories we have discussed are teleological in this sense. (Minor exceptions are some instances of non-reflexivity in $\trianglerighteq$.) But even some theories that may at first sight seem non-teleological – perhaps because of their non-consequentialism – can be expressed in a teleological format. Larry Temkin's work offers some examples (though Temkin himself would interpret them differently, namely as evidence for the intransitivity of betterness). We here give a Temkin-inspired example that has been discussed by Alex Voorhoeve (2013, 413).[50] Let the set X consist of three options:

terminal: "curing one young person of a terminal illness;"

[47]For further discussion of pluralism, see Mason (2015). On incomparability, see also Appendix B.5.

[48]Note that the distinction between weighing universalism and weighing relativism, which we have set aside, also refers to the weighing relation.

[49]One might restrict the requirements of transitivity and reflexivity to "relevant" property sets, namely those $S \subseteq \mathcal{P}$ such that $S = N(x, K)$ for some option x and some context K.

[50]Voorhoeve, like us, retains the transitivity of betterness. We quote Voorhoeve's wording to describe the options.

considerable: "saving 10000 from a considerable impairment;"

slight: "saving several billion from a very slight impairment."

We assume, for the sake of the example, that the total-welfare properties of these options are as follows. The welfare of *terminal* is 5000; the welfare of *considerable* is 10000; and the welfare of *slight* is 15000. So, the three options have the option properties $P_{\text{wel}=5K}$, $P_{\text{wel}=10K}$, and $P_{\text{wel}=15K}$, respectively. Let us grant the following Temkin-inspired intuitions:

- When faced with a choice between *terminal* and *considerable*, we ought to choose *considerable*, on total-welfare grounds.
- When faced with a choice between *considerable* and *slight*, we ought to choose *slight*, again on total-welfare grounds.
- When faced with a choice between *terminal* and *slight*, we ought to choose *terminal.* Given the dramatic difference between "death" and "very slight impairment", it would be disrespectful not to save the one from death for the sake of saving the billions from a very slight impairment.

We can make sense of these intuitions in a reason-based way. In addition to the total-welfare properties, we introduce a "disrespect property".[51] An option x in context K is "disrespectful" (for short, D) if and only if there is at least one other available option y such that someone's welfare loss from choosing x rather than y is very significantly greater than anyone's welfare loss from choosing y rather than x. Suppose, for the sake of our example, that option *slight* has this property if and only if *terminal* is also among the available options; no other option among our three options has this property. Now consider the reasons structure $\mathcal{R} = \langle N, \trianglerighteq \rangle$ where:

- for every context K, $N(K)$ consists of all the total-welfare properties and the disrespect property;
- the weighing relation ranks sets of properties in a linear order satisfying:

$$\begin{gathered}\{P_{\text{wel}=15K}\} \triangleright \{P_{\text{wel}=10K}\} \triangleright \{P_{\text{wel}=5K}\} \\ \triangleright \{P_{\text{wel}=15K}, D\} \triangleright \{P_{\text{wel}=10K}, D\} \triangleright \{P_{\text{wel}=5K}, D\}.\end{gathered}$$

In a context in which only *terminal* and *considerable* are available, their sets of normatively relevant properties are $\{P_{\text{wel}=5K}\}$ and $\{P_{\text{wel}=10K}\}$. So, the reasons structure

[51] Voorhoeve (2013) attributes this kind of move to Kamm (2007).

recommends choosing *considerable* over *terminal*. In a context in which only *considerable* and *slight* are available, their sets of normatively relevant properties are $\{P_{\text{wel}=10K}\}$ and $\{P_{\text{wel}=15K}\}$, so the recommendation is to choose *slight* over *considerable*. However, when both *terminal* and *slight* are available, their sets of normatively relevant properties are $\{P_{\text{wel}=5K}\}$ and $\{P_{\text{wel}=15K}, D\}$, and so the recommendation is to choose *terminal* over *slight*, in line with the Temkin-style intuitions. Interestingly, the weighing relation of this reasons structure is transitive (and, being a linear order, reflexive). So, the theory counts as teleological, although it is clearly non-consequentialist. What is going on is simply that D is a relational property (it is "essentially comparative" in Temkin's terminology), and the option *slight* has this property in some contexts but not in others.

5.6 Atomistic versus holistic theories

Our final distinction is that between "atomistic" and "holistic" theories. Informally, a theory is *atomistic* if its assessment of any set of properties is simply the "sum" (or some other simple combination) of its assessments of the individual properties in this set, while a theory is *holistic* if this is not the case; we will make this more precise below. Atomistic theories are "generalist" insofar as the "valence" they give to any property is the same, irrespective of which other properties are present as well. A property that is at least *pro tanto* right-making within one combination of properties remains *pro tanto* right-making when combined with other properties. By contrast, holistic theories are "particularist", insofar as the "valence" they give to some properties may depend on the other properties they are combined with. According to such a theory, a property may be *pro tanto* right-making in the presence of one particular set of properties, and *pro tanto* wrong-making in the presence of another. Obviously, one can draw the distinction between atomism and holism only once one has settled the individuation of properties.[52]

In our framework, the distinction between atomistic and holistic theories can be drawn in terms of the nature of those theories' weighing relations. We call a theory with reasons structure $\mathcal{R} = \langle N, \trianglerighteq \rangle$ *atomistic* if the weighing relation $\trianglerighteq$ is *separable*. Separability of $\trianglerighteq$ means that, for any comparable subsets S, S' of $\mathcal{P}$ and any subset T of $\mathcal{P}$ that does not overlap with S or S',

$$S \trianglerighteq S' \text{ if and only if } S \cup T \trianglerighteq S' \cup T,$$

provided $S \cup T$ and $S' \cup T$ are also comparable.[53] The theory is *holistic* if its weighing relation is not separable. A special case of separability is the *additive* one. Here, there exists

[52]See also Section 6.6.

[53]Two sets of properties, such as S and S', are *comparable* if $S \trianglerighteq S'$ or $S' \trianglerighteq S$.

a numerical *weighing function* w, which assigns to each property P in $\mathcal{P}$ a real number $w(P)$, its positive or negative "weight", such that, for any two subsets S, S' of $\mathcal{P}$,

$$S \trianglerighteq S' \text{ if and only if } \sum_{P \in S} w(P) \geq \sum_{P \in S'} w(P).$$

Monistic theories, as we have defined them, are automatically atomistic.[54] But there are also pluralistic theories that are atomistic. For an easy construction of such a theory, we can simply define a positive or negative numerical weight for each normatively relevant property and then assess any set of properties by summing up their weights.

The kinds of "particularist" theories defended by Jonathan Dancy and others are holistic in the present sense. Dancy (2013) says:

> "A feature can make one moral difference in one case, and a different difference in another. Features have ... variable relevance. Whether a feature is relevant or not in a new case, and if so what exact role it is playing there (the 'form' that its relevance takes there) will be sensitive to other features of the case. This claim emerges as the consequence of the core particularist doctrine, which we can call the holism of reasons. This is the doctrine that what is a reason in one case may be no reason at all in another, or even a reason on the other side... [I]n some contexts the fact that something is against the law is a reason not to do it, but in others it is a reason to do it (so as to protest, let us say, against the existence of a law governing an aspect of private life with which the law should not interfere)."

What Dancy describes here entails holism as we have defined it (a non-separable weighing relation), but it also seems to entail relativism (a context-variant set of normatively relevant properties). Logically, one can have holism without relativism and vice versa: non-separability of the weighing relation and context-variance of the normatively relevant properties are independent from one another.

6 The underdetermination of moral theory by deontic content

Our reason-based framework confirms the existence of a broadly Quinean underdetermination problem in moral philosophy: the same rightness function will often admit more

[54]To be precise, this assumes (without loss of generality) that $\trianglerighteq$ is defined only over singleton property sets. In Section 6.6, we note that, just as a pluralistic theory may be re-expressed as a monistic one by re-individuating properties, so a holistic theory may be similarly re-expressed as an atomistic one.

than one reason-based representation, in analogy to the way in which a body of empirical observations in science may admit more than one theoretical explanation. In sum, a moral theory's deontic content underdetermines the reasons structure underwriting it. This observation supports Parfit's claim that different moral theories can in principle climb the same mountain from different sides, reaching the same action-guiding recommendations at the summit, albeit via different routes. Moreover, we can accept this general structural point, irrespective of whether we are persuaded by Parfit's own example of it: the purported convergence of consequentialism, Kantianism, and Scanlonian contractualism.

The recognition of this underdetermination problem raises the question of how the six dimensions along which we have categorized moral theories manifest themselves in a theory's deontic content. Can we tell from the deontic content whether or not the underlying theory is consequentialist, universalist, agent-neutral, monistic, atomistic, and/or teleological? Or are some of these attributes "deontically inert"?

We address this issue by considering each of the six attributes and asking whether every rightness function – assuming it has a reason-based representation at all – admits a representation that has that attribute. This, in turn, sheds light on the question of whether, for every moral theory, there exists an extensionally equivalent counterpart theory with that attribute. So, we are asking, in effect, whether every moral theory can be "consequentialized", "teleoligized", "universalized", and so on. While the question of consequentialization has received much attention in formal ethics, the analogous questions for the other attributes have been less salient.[55] In what follows, we focus on moral theories that are representable in our framework (i.e., whose rightness functions admit a reason-based representation, relative to the given set $\mathcal{P}$ of properties).[56]

6.1 Does every rightness function admit a consequentialist representation?

Here, our analysis suggests the answer is "no". We obtain this negative answer not only if we define consequentialism in the traditional way – as representability in terms of a betterness ordering over the options – but also if we define it as we have proposed, namely in terms of the normative relevance of no properties other than option properties.

[55]For recent references to the idea of representing a moral theory in line with some target attribute other than consequentialism, see, e.g., Portmore (2007), who mentions "Kantianizing" and "contractualizing", and Hurley (2013), who discusses "deontologizing". Generally, however, consequentialization has received the most attention. Recall the references in footnote 1.

[56]When we refer to a rightness function in each of the following subsection titles, we are therefore, strictly speaking, referring to a *reason-based representable* rightness function.

The reason for this negative answer is that, even when a rightness function has some reason-based representation, it need not have a reason-based representation in which only option properties are normatively relevant. Formally, the conditions for reason-based representability in a context-unrelated format (corresponding to the top row of Table 1 above) are more demanding than those for reason-based representability *simpliciter*.[57] For instance, the norms of politeness in Amartya Sen's dinner-party example can be represented by a reasons structure in which the relational property *polite* is normatively relevant, while they cannot be represented by a reasons structure in which only option properties are relevant. Similarly, several of our other examples of normative theories – such as the entitlement-satisfaction and satisficing theories and the simple deontological theory in Jim and the Indians – could only be represented by reasons structures involving context-related properties.[58]

How does this finding relate to the "Extensional Equivalence Thesis", defended by scholars such as Dreier (2011) and Portmore (2007), the claim that every plausible moral theory is extensionally equivalent to a consequentialist theory? Clearly, *if* one defines consequentialism in the way we do – namely as the normative irrelevance of any properties other than option properties – *then* one must conclude that the Extensional Equivalence Thesis is not generally true. Of course, since Dreier and Portmore restrict that thesis to *plausible* moral theories, one might deny the plausibility of those normative theories that defy consequentialization, but we are not persuaded that all the relevant examples, including Sen's simple theory of politeness, can be dismissed as implausible. Alternatively, one could try to defend the Extensional Equivalence Thesis by adopting a more permissive definition of consequentialism. For example, Dreier and Portmore take consequentialism to be compatible with the normative relevance of agent-related properties.[59] If we were to define consequentialism as requiring that only option properties *and* agent-related properties (but no other context-related properties) be normatively relevant, then our illustrative deontological theory in Jim and the Indians would be reclassified as consequentialist. But even under this permissive definition of consequentialism, we would still have to conclude that not every moral theory can

[57]This follows from Theorem 5 in Dietrich and List (2016). The application to moral philosophy is new.

[58]These claims about our examples assume that the set $\mathcal{P}$ of admissible properties is not so artificially rich as to allow gerrymandered forms of relativism as a substitute for non-consequentialism. In Section 6.2, we say more about why some relativist theories may be deontically equivalent to some non-consequentialist ones.

[59]Dreier says: "I count agent-centered consequentialism as a kind of consequentialism" (ibid., 100). Similarly, Portmore (2007) recognizes agent-relative forms of consequentialism. For critical discussions of agent-neutral and agent-relative approaches to consequentialization, see Louise (2004), Schroeder (2007), and Hurley (2013). For a defence of agent-neutral consequentializability, see Oddie and Milne (1991).

be consequentialized. Once again, the simple normative theory of politeness is a case in point: a property such as *politeness*, which is indispensable for any reason-based representation of the rightness function here, is menu-related, not agent-related.

One might try to respond by accepting even further properties as admissible candidates for normatively relevant properties in a consequentialist theory. Generally, one might ask: what *kinds* of normatively relevant properties must we minimally invoke in order to represent a given rightness function in a reason-based format? Are option properties alone sufficient? Or do we need to invoke agent-related properties? Or even other context-related properties? Now the problem is this: if we were to redefine consequentialism so as to permit the normative relevance of more and more kinds of context-related properties, this would run the risk of making the notion of consequentialism vacuous: *any* theory that has a reason-based representation might then end up being called "consequentialist". By contrast, as we have seen, our preferred definition of consequentialism – which requires the normative relevance of option properties alone – is non-vacuous.

That said, one would be able to consequentialize every normative theory by redescribing the options themselves. As shown in Appendix A, such a redescription is always logically possible – a result that could be viewed as capturing the core idea behind the Extensional Equivalence Thesis. However, as already noted, this would not generally be very informative and would involve a significant departure from the original description of the moral choice problem. It thus seems reasonable to conclude that the distinction between consequentialist and non-consequentialist theories is deontically significant. But whether or not one agrees with this substantive conclusion, the reason-based framework can offer a useful diagnosis of what is at stake in the debate about consequentializability.

6.2 Does every rightness function admit a universalist representation?

The answer to this question, which has received less attention in formal ethics than the question about consequentializability, is "yes", provided that (as assumed) the rightness function has some reason-based representation at all. Surprisingly, the conditions for the existence of a reason-based representation *simpliciter* are logically equivalent to the conditions for the existence of a reason-based representation *without context-variant normative relevance*.

Fact: A rightness function R has a reason-based representation with a constant normative relevance function if and only if it has a reason-based representation *simpliciter*.[60]

[60] Formally, this is a subtly strengthened corollary of Proposition 3 in Dietrich and List (2016). The earlier result – stated in terms of choice functions – requires the function R to be dilemma-free. The

In other words, every reasons structure is *deontically equivalent* to some reasons structure that is context-invariant (relative to the same admissible set $\mathcal{P}$ of properties). Thus, every relativist moral theory does indeed have a universalist counterpart theory with exactly the same deontic content. There is a crucial caveat, however. The universalist counterpart theory may have to be non-consequentialist: the price of avoiding relativism is the normative relevance of context-related properties. Even if the original relativist theory deems only option properties normatively relevant (though different ones in different contexts), its universalist counterpart may need to deem suitable context-related properties normatively relevant in all contexts, so as to arrive at the same action-guiding recommendations. Schematically, the universalization of the given theory may involve a shift from the top right quadrant in Table 1 into the bottom left one.

These observations show that the distinction between universalist and relativist theories is really a distinction at the level of the reasons structure, not at the level of the deontic content. They further illustrate that there may sometimes be a tradeoff between universalism and consequentialism: in order to "universalize" a given moral theory, we may have to represent it in a non-consequentialist format (by deeming context-related properties normatively relevant). In Appendix C, we derive an analogous universalizability result for weighing relativism, the alternative form of relativism we have mentioned: all weighing-relativist theories, except those in a very special class, have a weighing-universalist counterpart theory with the same deontic context.

6.3 Does every rightness function admit an agent-neutral representation?

Here, the answer depends on which of our two orthogonal ways of drawing the distinction between agent-neutrality and agent-relativity we adopt. Recall that, on one interpretation, agent-relativity is a special case of relativism and, on another, a special case of non-consequentialism (again, for discussion, see Ridge 2011 and Schroeder 2007).

If we define agent-neutrality as agent-invariance – so that agent-relativity becomes a special case of relativism – then the answer to the question of agent-neutral representability is a qualified "yes". As we have seen, every rightness function that has *some* reason-based representation also has one with a constant normative relevance function; and so, *a fortiori*, it has one that is agent-invariant: the set of normatively relevant properties will not vary with the agent i. However, just as the typical cost of re-expressing a relativist theory in a structurally universalist format is context-relatedness

application to moral philosophy is new.

in the normatively relevant properties, so the typical cost of the present exercise will be agent-relatedness in those properties, which implies agent-relativity in the second, non-consequentialist sense. Our second representation of ethical egoism in Section 5.3 illustrates this. The reasons structure in that example is agent-invariant, but it deems agent-related properties normatively relevant.

If we define agent-neutrality as agent-unrelatedness – so that agent-relativity becomes a special case of non-consequentialism – then the answer to our present question is "no": not every rightness function admits an agent-neutral representation. Just as it is not true that every non-consequentialist theory can be consequentialized – such that only option properties are normatively relevant – so the elimination of the normative relevance of agent-related properties will not generally be possible if the given theory has an agent-related reasons structure. This is consistent with the familiar observation that the normative relevance of some agent-related properties may sometimes be indispensable if we wish to account for certain verdicts of commonsense morality (cf. Scheffler 1982, Dreier 1993, and the consequentialization debate).

6.4 Does every rightness function admit a monistic representation?

Here, the answer is a qualified "no". As should be evident, for instance from our Temkin-inspired example, there are some rightness functions which only admit pluralistic reason-based representations, at least relative to a given set $\mathcal{P}$ of admissible properties. Thus, if we attach significance to the way we have specified the properties in $\mathcal{P}$ – perhaps because those properties meet some criterion of naturalness, simplicity, or salience from a human perspective – then pluralism in the reasons structure may be unavoidable for representing certain rightness functions.

Suppose, on the other hand, we impose no restrictions on the set $\mathcal{P}$. We would then be free to specify and re-individuate properties in any logically possible way. For instance, for any two properties P and P', we might also construct the conjunctive and disjunctive properties $P \wedge P'$ and $P \vee P'$ with extensions $[P] \cap [P']$ and $[P] \cup [P']$, respectively. In Appendix B, we show that, by suitably re-individuating properties like this, we can easily represent every rightness function in a monistic format. (A trivializing case of such a representation is one in which the only normatively relevant property of each option x in each context K is "being option x in context K".) However, the resulting monistic representation will often rely on rather gerrymandered properties and will therefore be unilluminating. Given a more disciplined specification of the set $\mathcal{P}$ (e.g., by invoking some criterion of which properties are natural, simple, or salient for humans), the distinction between monistic and pluralistic theories is indeed deontically significant.

As Elinor Mason (2015) notes, "[p]luralists argue that there really are several different values, and that these values are not reducible to each other or to a super value". Our framework confirms that not every pluralistic theory can be recast in a monistic format, unless we invoke an artificially gerrymandered "super value".

6.5 Does every rightness function admit a teleological representation?

Recall that we have drawn a formal distinction between "consequentialism" and "teleology": the former requires that only option properties be normatively relevant (it is a constraint on *which properties matter*), while the latter requires that the weighing relation be transitive and reflexive (it is a constraint on *how these properties matter*). For this reason, the present question – about whether every rightness function admits a teleological representation – is distinct from our earlier question about consequentializability. Like our earlier answer, however, the present answer is also a qualified "no", although this "no" is less obvious than in the earlier case.

We have seen that a very large class of rightness functions – even the ones in Temkin-inspired examples, contrary to Temkin's own preferred interpretation – can be represented in terms of a reasons structure with a transitive weighing relation. Yet, if we hold the set $\mathcal{P}$ of admissible properties fixed, then the conditions for reason-based representability with a transitive weighing relation are formally more demanding than those for reason-based representability *simpliciter*. In Appendix B, we give an example to illustrate this point.

By contrast, the conditions for reason-based representability with a reflexive weighing relation are the same as those without this constraint, provided we restrict our attention to dilemma-free rightness functions; see also Appendix B. A non-reflexive weighing relation is needed, however, to capture certain moral dilemmas. For instance, to generate the verdict that murder is impermissible even if it is the only available option, this option's set of normatively relevant properties must not stand in the relation $\trianglerighteq$ to itself.

Although we have seen that, formally, not all rightness functions admit a teleological representation for a given set $\mathcal{P}$, we are unsure whether many *plausible* moral theories defy teleologization in the strong sense of requiring an intransitive weighing relation. It is still possible that most plausible theories are representable with a transitive weighing relation.

6.6 Does every rightness function admit an atomistic representation?

Again, the answer is "no", though subject to a similar qualification as in our discussion of monistic representations above. Suppose we take the set $\mathcal{P}$ of admissible properties to be fixed, perhaps because we consider its elements the most plausible or natural candidates for normatively relevant properties and we are unwilling to re-individuate them. Then we may well have no choice but to represent a given rightness function in terms of a reasons structure with a non-separable weighing relation. If, on the other hand, we are free to specify the set $\mathcal{P}$ as permissively as we like, for instance by re-individuating properties, then we can always construct an atomistic representation of a given rightness function, which may coincide with the monistic representation we have mentioned (again, see Appendix B). Of course, this may be a very artificial representation. Once we set this possibility of gerrymandering properties aside, the distinction between atomistic and holistic theories is deontically significant. This is broadly consistent with the point – often made by moral particularists – that some bodies of normative judgments may be irreducibly holistic, in the sense of not being amenable to an atomistic systematization (for a survey, see Dancy 2013).

7 Concluding remarks

We have argued that any moral theory within a very large class can be represented by a reasons structure, which encodes the theory's answer to the two-part question of which properties matter and how they do so. This reasons structure not only entails the theory's verdicts about which options are permissible in each context, but also expresses the theory's explanation for those verdicts. We have shown that this way of representing moral theories yields a very general taxonomy of such theories, in which many salient distinctions can be drawn. Our approach to the formal representation of moral theories is significantly more general than the classical approach involving consequentialization, and, we have argued, it is explanatorily richer too. Our paper ends with some technical appendices, which further substantiate our claims. The final appendix explains the resources our framework offers for handling moral choices under uncertainty.

References

Bhattacharyya, A., P. K. Pattanaik, and Y. Xu. 2011. "Choice, Internal Consistency and Rationality." *Economics and Philosophy* 27: 123–49.

Bossert, W., and K. Suzumura. 2010. *Consistency, Choice, and Rationality.* Cambridge, MA: Harvard University Press.

Brandt, F., M. Brill, and P. Harrenstein. 2016. “Tournament Solutions.” In *Handbook of Computational Social Choice*, ed. F. Brandt et al., 57–84. Cambridge: Cambridge University Press.

Broome, J. 1991. *Weighing Goods: Equality, Uncertainty and Time.* Oxford: Blackwell.

Broome, J. 2004. *Weighing Lives.* Oxford: Oxford University Press.

Brown, C. 2011. “Consequentialize This.” *Ethics* 121: 749–71.

Buchak, L. 2013. *Risk and Rationality.* Oxford: Oxford University Press.

Dancy, J. 2013. “Moral Particularism.” In *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), ed. E. N. Zalta. Available at: http://plato.stanford.edu/archives/fall2013/entries/moral-particularism/

Dietrich, F., and C. List. 2013. “A Reason-Based Theory of Rational Choice.” *Noûs* 47: 104–34.

Dietrich, F., and C. List. 2016. “Reason-Based Choice and Context-Dependence: An Explanatory Framework.” *Economics and Philosophy* 32: 175–229.

Dreier, J. 1993. “Structures of Normative Theories.” *The Monist* 76: 22–40.

Dreier, J. 2011. “In Defense of Consequentializing.” In *Oxford Studies in Normative Ethics* 1, ed. M. Timmons, 97–118. Oxford: Oxford University Press.

Heyd, D. 2016. “Supererogation.” In *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition), ed. E. N. Zalta. Available at: http://plato.stanford.edu/archives/spr2016/entries/supererogation/

Hurley, P. 2013. “Consequentializing and Deontologizing: Clogging the Consequentialist Vacuum.” In *Oxford Studies in Normative Ethics* 3, ed. M. Timmons, 123–53. Oxford: Oxford University Press.

Kamm, F. M. 2007. *Intricate Ethics: Rights, Responsibilities, and Permissible Harm.* Oxford: Oxford University Press.

Liu, F. 2010. “Von Wright’s ‘The Logic of Preference’ revisited.” *Synthese* 175: 69–88.

Louise, J. 2004. "Relativity of Value and the Consequentialist Umbrella." *The Philosophical Quarterly* 54: 518–36.

Lyons, D. 1965. *Forms and Limits of Utilitarianism.* Oxford: Clarendon Press.

Mason, E. 2015. "Value Pluralism." In *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition), ed. E. N. Zalta. Available at: http://plato.stanford.edu/archives/sum2015/entries/value-pluralism/

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

Oddie, G., and P. Milne. 1991. "Act and value: Expectation and the representability of moral theories." *Theoria* 57: 42–76.

Osherson, D., and S. Weinstein. 2012. "Preference based on reasons." *Review of Symbolic Logic* 5: 122–47.

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

Pettit, P. 1991. "Decision Theory and Folk Psychology." In *Foundations of Decision Theory*, ed. M. Bacharach and S. Hurley, 147–75. Oxford: Blackwell.

Portmore, D. 2007. "Consequentializing Moral Theories." *Pacific Philosophical Quarterly* 88: 39–73.

Portmore, D. 2011. *Commonsense Consequentialism: Wherein Morality Meets Rationality.* Oxford: Oxford University Press.

Psillos, S. 1999. *Scientific Realism: How Science Tracks Truth.* Abingdon: Routledge.

Quine, W. V. 1975. "On Empirically Equivalent Systems of the World." *Erkenntnis* 9: 313–28.

Richter, M. 1971. "Rational Choice." In *Preferences, Utility, and Demand*, ed. J. S. Chipman et al., 29–58. New York: Harcourt Brace Jovanovich.

Ridge, M. 2011. "Reasons for Action: Agent-Neutral vs. Agent-Relative." In *The Stanford Encyclopedia of Philosophy* (Winter 2011 Edition), ed. E. N. Zalta. Available at: http://plato.stanford.edu/archives/win2011/entries/reasons-agent/

Sachs, B. 2010. "Consequentialism's Double-Edged Sword." *Utilitas* 22: 258–71.

Savage, L. 1954. *The Foundations of Statistics.* New York: Wiley.

Scheffler, S. 1982. *The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions*. Oxford: Clarendon Press.

Schroeder, M. 2007. "Teleology, Agent-Relative Value, and 'Good'." *Ethics* 117: 265–95.

Sen, A. K. 1993. "Internal Consistency of Choice." *Econometrica* 61: 495–521.

Temkin, L. S. 1996. "A Continuum Argument for Intransitivity." *Philosophy and Public Affairs* 25: 175–210.

Temkin, L. S. 2012. *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. Oxford: Oxford University Press.

Tenenbaum, S. 2014. "The Perils of Earnest Consequentializing." *Philosophy and Phenomenological Research* 88: 233–40.

Urmson, J. O. 1958. "Saints and Heroes." In *Essays in Moral Philosophy*, ed. A. I. Melden, 198–216. Seattle: University of Washington Press.

Vallentyne, P. 1988. "Gimmicky Representations of Moral Theories." *Metaphilosophy* 19: 253–63.

van Fraassen, B. C. 1980. *The Scientific Image*. Oxford: Oxford University Press.

von Neumann, J., and O. Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

Voorhoeve, A. 2013. "Vaulting Intuition: Temkin's Critique of Transitivity." *Economics and Philosophy* 29: 409–23.

Williams, B. 1973. "A Critique of Utilitarianism." In *Utilitarianism: For and Against*, ed. J. J. C. Smart and B. Williams, 77-150. Cambridge: Cambridge University Press.

Appendix A: When can a rightness function be represented by a binary relation?

A.1 An illustrative representation theorem

We state one illustrative representation theorem, adapted from Marcel Richter's work (1971):

Theorem 1. A rightness function R is representable by some binary relation $\succeq$ on X if and only if R satisfies the following axiom.[61]

Axiom 1. For every context K in $\mathcal{K}$ and any available option x in $[K]$, if, for every available option y in $[K]$, there exists some context K' in $\mathcal{K}$ in which x is right while y is available (i.e., $x \in R(K')$ and $y \in [K']$), then x is right in K (i.e., $x \in R(K)$).

For the purposes of this paper, we need not worry too much about the interpretation of Axiom 1. It is simply a formal condition that some rightness functions satisfy and others violate. Its point is only to provide a dividing line between those rightness functions that can be represented by a binary relation, and those that cannot. Axiom 1 by itself does not guarantee representability of a rightness function R by a transitive and reflexive binary relation, i.e., a betterness relation in the narrower sense. For the latter, stronger requirements on R are needed (see, e.g., Bossert and Suzumura 2010). The details need not concern us here.

A.2 Consequentializing by redescribing the options

The limits to consequentialization established by results such as Theorem 1 depend on holding the set X of options fixed. If we permit redescriptions of the options, we can trivially represent any rightness function in terms of a binary relation. This can be viewed as a variant of the Extensional Equivalence Thesis defended by Dreier (2011) and Portmore (2007), according to which every set of action-guiding recommendations has some consequentialist representation, under a sufficiently permissive notion of consequences.[62]

Let R be *any* rightness function defined on $\mathcal{K}$. Formally, R maps each context K in $\mathcal{K}$ to a subset of $[K]$. We redescribe the options as follows. Construct a new set X' of options, consisting of all pairs of the form $\langle x, K\rangle$, where x is an option in X and K is a context in which x is available. For each context K, let us reinterpret the set of available options in K as $\{\langle x, K\rangle : x \in [K]\}$. We can now recast the rightness function R as a function R' from $\mathcal{K}$ into the set of subsets of X'. For each K in $\mathcal{K}$, let

$$R'(K) = \{\langle y, K\rangle : y \in R(K)\}.$$

Then R' is trivially representable by a binary relation on X'. This is because each redescribed option in X' is available in only one context K; it is described so richly that

[61] Richter proved the theorem for choice functions, the equivalents of dilemma-free rightness functions. We have verified that the result still holds if the "dilemma-free" restriction is dropped.

[62] As noted earlier, Portmore and Dreier restrict this thesis to the deontic content of "plausible" moral theories. Our analysis illustrates the generality of this thesis in formal terms.

its occurrence is not repeatable. To represent R', we simply need to construct a binary relation $\succeq$ on X' which ranks the permissible options in each context K equally and strictly above the impermissible ones. Interpretationally, R' and the original rightness function R encode the same action-guiding recommendations.[63]

Appendix B: Further results on reason-based representation

B.1 A representation theorem

Any rightness function that satisfies a relatively undemanding regularity axiom has a reason-based representation. To state the axiom, we require one preliminary definition. For any rightness function R and any two sets of properties, S and S', we say that S *is sometimes choice-worthy in the presence of* S' if there exists some context K in which some option instantiating S is right while some option instantiating S' is available. (An option x is said to *instantiate* a set S of properties in a context K if $\mathcal{P}(x, K) = S$.)

Axiom 2. For every context K in $\mathcal{K}$ and any available option x in $[K]$, if, for every available option y in $[K]$, $\mathcal{P}(x, K)$ is sometimes choice-worthy in the presence of $\mathcal{P}(y, K)$, then x is right in K (i.e., $x \in R(K)$).

As in the case of Axiom 1 above, we should not worry too much about the interpretation of Axiom 2. It is simply a verifiable condition that a rightness function may or may not satisfy. The following result holds:

Theorem 2. A rightness function R is representable by some reasons structure $\mathcal{R}$ if and only if R satisfies Axiom 2.

Axiom 2 thus allows us to draw a line between those rightness functions that can be represented in a reason-based way and those that cannot. The present theorem is a reason-based counterpart of Theorem 1 above.[64] Just as Theorem 1 only established conditions for the representability of a rightness function by *some* binary relation – not necessarily one with further properties such as transitivity and reflexivity – so Theorem 2 only establishes conditions for the representability of a rightness function by *some*

[63]Formally, the present construction is borrowed from Dietrich and List (2016). For an earlier, more sophisticated construction that establishes a similar point, see Bhattacharyya, Pattanaik, and Xu (2011).

[64]Formally, Theorem 2 is a slightly strengthened variant of Theorem 1 in Dietrich and List (2016). While the original result was proved for the choice-theoretic equivalents of dilemma-free rightness functions, we here drop the "dilemma-free" restriction. As in the case of Theorem 1, we have checked that the result still holds without it. The application to moral philosophy is new.

reasons structure – not necessarily one with properties such as transitivity and reflexivity of the weighing relation. We discuss those further properties in Appendix B.3.

B.2 How permissive is reason-based representability?

The more properties we have at our disposal in constructing a reasons structure, the more rightness functions we can represent. Technically, the demandingness of Axiom 2 depends on how large or small the set $\mathcal{P}$ is. Recall that $\mathcal{P}$ contains those properties that we admit as candidates for normatively relevant properties.

In the limit, if $\mathcal{P}$ contains all logically possible properties, then every logically possible rightness function has a reason-based representation. To see this, suppose, in particular, that the set $\mathcal{P}$ contains every property of the form:

$$P_{x,K}: \text{ ``The option is } x\text{, and the context is } K\text{''},$$

where x is an option and K a context.[65] Such a property is called *maximally specific*, as it is only satisfied by one option in one context. We can then represent any logically possible rightness function R by constructing the following reasons structure $\mathcal{R} = \langle N, \trianglerighteq \rangle$:

- for every context K, $N(K)$ is the set of *all* maximally specific properties;
- for any options x and y and any context K, $\{P_{x,K}\} \trianglerighteq \{P_{y,K}\}$ if and only if x is deemed right by R in context K while y is available (i.e., $x \in R(K)$ and $y \in [K]$).

Indeed, Axiom 2 is trivially satisfied when all maximally specific properties are admissible as candidates for normatively relevant properties. Yet, the reasons structure we have just constructed is not very illuminating. It accounts for the given permissibility verdicts simply by stipulating that "being option x in context K" is "better" than "being option y in context K" whenever x is deemed right in context K while y is available. From a substantive perspective, such a representation can be criticized for being gerrymandered and not identifying any non-*ad-hoc* "right-making" or "good-making" features.

The present considerations show that before we can usefully ask whether a given rightness function has a reason-based representation, we must at least provisionally commit ourselves to some auxiliary hypothesis concerning the set $\mathcal{P}$. In other words, we must take a view on which properties are admissible candidates for right-making or good-making features and which are not.

[65]Strictly speaking, the property $P_{x,K}$ has non-empty extension – as we normally require – only if x is available in context K. Nothing hinges on this.

B.3 How demanding is representability in a teleological format?

We show three things. First, the conditions for reason-based representability with a transitive weighing relation are more demanding than those for reason-based representability *simpliciter*. Second, the conditions for reason-based representability with a reflexive weighing relation are the same as those for reason-based representability *simpliciter*, provided we restrict our attention to dilemma-free rightness functions. (Similarly, assuming dilemma-freeness, the conditions for reason-based representability with a transitive and reflexive weighing relation are the same as those for reason-based representability with a transitive weighing relation *simpliciter*.) Third, certain moral dilemmas can be represented only if the weighing relation is non-reflexive.

To establish the first claim, we give an example of a rightness function that admits a reason-based representation relative to a given set $\mathcal{P}$ of properties, but only with an intransitive weighing relation. Let $X=\{x,y,z\}$. Suppose that, for every non-empty subset Y of X, there is a context K such that $[K]=Y$. The rightness function deems y impermissible whenever x is available; all other options are always permissible. Formally, $R(K)=[K]$ if x is not in $[K]$, and $R(K)=[K]\backslash\{y\}$ if x is in $[K]$. (Note that $[K]\backslash\{y\} = [K]$ if y is not in $[K]$.) Suppose that $\mathcal{P}$ contains only option properties and that x, y, and z instantiate distinct property sets, say $\mathcal{P}(x)=\{P\}$, $\mathcal{P}(y)=\{Q\}$, and $\mathcal{P}(z)=\{R\}$. This rightness function is representable by a reasons structure $\mathcal{R}=\langle N,\trianglerighteq\rangle$ in which

- $N(K)=\{P,Q,R\}$ for every context K, and
- $\{P\}\vartriangleright\{Q\}$, $\{P\}\equiv\{R\}$, $\{Q\}\equiv\{R\}$, $\{P\}\equiv\{P\}$, $\{Q\}\equiv\{Q\}$, $\{R\}\equiv\{R\}$.

The weighing relation is clearly intransitive, and it is easy to see that no reason-based representation of R can avoid such intransitivity. Accordingly, R admits no teleological representation, unless we enrich the set $\mathcal{P}$ of properties we are prepared to invoke.

To establish our second claim – concerning reflexivity – consider any dilemma-free rightness function R, with reason-based representation $\mathcal{R}=\langle N,\trianglerighteq\rangle$. Let $\trianglerighteq^*$ be the minimal subrelation of $\trianglerighteq$ for which $\langle N,\trianglerighteq^*\rangle$ entails the rightness function R.[66] (Such a subrelation exists and is unique.) Then construct the *reflexive closure* of $\trianglerighteq^*$. Formally, this is the binary relation $\trianglerighteq^{**}$ such that, for any sets of properties $S,S'\subseteq\mathcal{P}$, we have $S\trianglerighteq^{**}S'$ if and only if $S\trianglerighteq^{*}S'$ or $S=S'$. Since R is dilemma-free, one can show that the reasons structure $\langle N,\trianglerighteq^{**}\rangle$ also entails the rightness function R. This establishes that, for a dilemma-free rightness function, reason-based representability with a reflexive weighing

[66] A binary relation $\trianglerighteq^*$ is a *subrelation* of $\trianglerighteq$ if, for any S and S', $S\trianglerighteq^* S'$ implies $S\trianglerighteq S'$.

relation is no more demanding than reason-based representability *simpliciter*.[67]

To establish our third claim – concerning moral dilemmas – consider a rightness function that sometimes yields an empty set of permissible options even when the set of available options is singleton. Clearly, such a rightness function can at most be represented by a reasons structure with a non-reflexive weighing relation.

B.4 From pluralistic to monistic representations

Suppose the rightness function R admits a reason-based representation relative to the set $\mathcal{P}$ of admissible properties, but no such representation is monistic. We show that any pluralistic (and perhaps holistic) reasons structure $\mathcal{R} = \langle N, \trianglerighteq \rangle$ for R can be "translated" into a monistic (and atomistic) one, relative to a suitably redefined set $\mathcal{P}^*$ of properties. A trivial monistic and atomistic representation of R is the one constructed in Appendix B.2, where $\mathcal{P}^*$ is the set of all maximally specific properties (of the form "being option x in context K") and exactly one such property is normatively relevant for each option in each context. But this reasons structure uses *ad-hoc* properties and makes no use of the original pluralistic reasons structure $\mathcal{R}$, and so hardly qualifies as a "translation".

We can construct a non-trivial monistic (and atomistic) representation of R by re-individuating properties. Given the pluralistic representation $\mathcal{R} = \langle N, \trianglerighteq \rangle$, the idea is to introduce new properties that are logical combinations of the normatively relevant properties from $\mathcal{R}$. For each option x and each context K in which x is available, we introduce the property of "satisfying all properties in $N(x, K)$ and no others in $N(K)$", denoted $P_{N(x,K)}$. The extension of $P_{N(x,K)}$ is the intersection of the extensions of all properties in $N(x, K)$ and of the complements of the extensions of all properties in $N(K)\backslash N(x, K)$.[68] The new set of admissible properties $\mathcal{P}^*$ must then include (at least) all properties of the form $P_{N(x,K)}$ and all context properties from the original set $\mathcal{P}$.

We now show how to define a monistic reasons structure $\mathcal{R}^* = \langle N^*, \trianglerighteq^* \rangle$ for R, relative to $\mathcal{P}^*$. We need one further preliminary definition. For each context K, let $\mathcal{K}_{N(K)}$ be the equivalence class of those contexts K' in $\mathcal{K}$ for which $N(K) = N(K')$. Then define $\mathcal{R}^*$ as follows:

- for each context K, $N^*(K) = \{P_{N(x',K')} : K' \in \mathcal{K}_{N(K)} \text{ and } x' \in [K']\}$;

[67]Similarly, any dilemma-free rightness function R that admits a representation $\langle N, \trianglerighteq \rangle$ in which $\trianglerighteq$ is transitive also admits a representation $\langle N, \trianglerighteq^{**} \rangle$ in which $\trianglerighteq^{**}$ is transitive and reflexive. Here we must take $\trianglerighteq^*$ to be the minimal *transitive* subrelation of $\trianglerighteq$ for which $\langle N, \trianglerighteq^* \rangle$ entails R, and take $\trianglerighteq^{**}$ to be the reflexive closure of $\trianglerighteq^*$.

[68]Formally, $[P_{N(x,K)}] = \left(\bigcap_{P \in N(x,K)} [P]\right) \cap \left(\bigcap_{P \in N(K)\backslash N(x,K)} \overline{[P]}\right)$, where the *complement* $\overline{S}$ of any set S of option-context pairs consists of all option-context pairs not in S.

- for any options x and y and any context K in which x and y are available, $\{P_{N(x,K)}\} \trianglerighteq^* \{P_{N(y,K)}\}$ if and only if $N(x,K) \trianglerighteq N(y,K)$.

It is easy to verify three points. First, $\mathcal{R}^*$ satisfies the invariance constraint on reasons structures, i.e., if $\mathcal{P}^*(K) = \mathcal{P}^*(K')$, then $N^*(K) = N^*(K')$. Second, $\mathcal{R}^*$ is no more relativist than $\mathcal{R}$, i.e., whenever $N(K) = N(K')$, then $N^*(K) = N^*(K')$. Third, $\mathcal{R}^*$ represents R, formally $R^{\mathcal{R}^*} = R$. We have thereby translated the original pluralistic reasons structure $\mathcal{R}$ into a monistic reasons structure $\mathcal{R}^*$ for a suitably redefined set of properties. Furthermore, the reasons structure $\mathcal{R}^*$ is obviously atomistic as well.

B.5 Defeating, undefeated, and defeat-the-most representability

We now consider some alternative ways in which one might define the rightness function R entailed by a reasons structure $\mathcal{R}$=$\langle N, \trianglerighteq \rangle$.[69] Recall our definition: for any context K,

$$R(K) = \{x \in [K] : N(x,K) \trianglerighteq N(y,K) \text{ for all } y \in [K]\}.$$

Thus, for an option to be permissible, its set of normatively relevant properties must at least weakly outweigh or defeat the corresponding set for every available option. Call this the *defeating criterion* of permissibility. A different criterion is this: for any context K,

$$R(K) = \{x \in [K] : \text{not } N(y,K) \triangleright N(x,K) \text{ for any } y \in [K]\}.$$

Here, for an option to be permissible, its set of normatively relevant properties must not be strictly outweighed or defeated by the corresponding set for any available option. Call this the *undefeated criterion* of permissibility. This criterion is equivalent to the earlier one if the relation $\trianglerighteq$ is complete, because completeness implies that $N(x,K) \trianglerighteq N(y,K)$ if and only if not $N(y,K) \triangleright N(x,K)$. Generally, however, the undefeated criterion is less demanding, so that the corresponding set $R(K)$ may grow. A third criterion can be defined as follows. Let rank(x,K) be the number of available options y in context K such that $N(x,K) \trianglerighteq N(y,K)$. We can then consider the following definition: for any context K,

$$R(K) = \{x \in [K] : \text{rank}(x,K) \geq \text{rank}(y,K) \text{ for all } y \in [K]\}.$$

On this criterion, for an option to be permissible, its set of normatively relevant properties must weakly outweigh or defeat the corresponding set for as many available options as possible. Call this the *defeat-the-most criterion.* This last criterion has the feature of

[69] We are grateful to John Broome for helpful comments that have prompted this exploration.

ruling out moral dilemmas: in every context K, at least one option is permissible. The three criteria all have well-known counterparts in classical rational choice theory.[70]

Each criterion yields a corresponding definition of reason-based representation. Let us say that a rightness function R has a *D-representation*, a *U-representation*, and an *M-representation* if there is some reasons structure $\mathcal{R}$ that entails R according to the defeating criterion, the undefeated criterion, and the defeat-the-most criterion, respectively. While D-representation has been our default notion, our taxonomy of moral theories could be refined by recognizing the distinction between D-, U-, and M-representations and categorizing theories in terms of the permissibility criteria they employ.

Let us briefly illustrate the two additional permissibility criteria just introduced. To illustrate the undefeated criterion, consider a moral theory according to which some sets of properties are mutually incommensurable. Let K be a context in which *all* available options' sets of normatively relevant properties are mutually incommensurable. Suppose now that the given theory deems *every* available option permissible on grounds of incommensurability, i.e., $R(K) = [K]$. For a reasons structure to entail this rightness function under our default criterion (i.e., D-representation), we would have to define the relation $\trianglerighteq$ such that $N(x, K) \equiv N(y, K)$ for all available options x and y in context K. This would not capture the incommensurability of the relevant property sets, incorrectly suggesting that they are somehow equivalent. By contrast, if we defined the relation $\trianglerighteq$ so as to leave all those property sets mutually unrelated, our reasons structure would not entail R under the defeating criterion. However, it would entail R under the undefeated criterion, at least as far as context K is concerned. In that context, none of the options would be defeated, and so all would count as permissible according to the undefeated criterion. The example illustrates that U-representation may be more appropriate than D-representation for moral theories that recognize cases of incommensurable sets of properties and that deem all relevant options permissible in such cases.

Let us turn to the defeat-the-most criterion. Consider a moral theory which says that

- only x is permissible in a choice between x and y;
- only y is permissible in a choice between y and z,
- only z is permissible in a choice between z and x,
- all of x, y, and z are permissible in a choice between x, y, and z.

[70] They correspond to different *tournament solutions*: methods for choosing a set of "winners" among a set of options, based on some binary relation over these options. Our three criteria correspond to the "greatest element", "maximal element", and "Copeland" solutions, respectively. See Brandt, Brill, and Harrenstein (2016) and also Bossert and Suzumura (2010).

Think of the theory as consequentialist and universalist. We are looking for a reason-based representation $\mathcal{R} = \langle N, \trianglerighteq \rangle$. Given consequentialism and universalism, each option must have a context-independent set of normatively relevant properties; we can therefore write $N(x)$, $N(y)$, and $N(z)$ for $N(x, K)$, $N(y, K)$, and $N(z, K)$, respectively. To represent the theory according to our original definition, we must have $N(x) \triangleright N(y)$, $N(y) \triangleright N(z)$, and $N(z) \triangleright N(x)$. Still, such a reasons structure fails to D-represent the theory: it fails to entail the verdict that all of x, y, and z are permissible when all three options are available, since none of the property sets $N(x)$, $N(y)$, and $N(z)$ weakly defeats the two others. Similarly, the reasons structure fails to U-represent the theory, since each of the sets $N(x)$, $N(y)$, and $N(z)$ is strictly defeated by one of the others. However, the reasons structure M-represents the given theory, because each of $N(x)$, $N(y)$, and $N(z)$ weakly defeats two out of the three sets, including itself, assuming $N(x) \equiv N(x)$, $N(y) \equiv N(y)$, and $N(z) \equiv N(z)$. This example shows that in some cases M-representations exist even when no D- or U-representations are available.

Finally, it is worth noting the following logical relationships between the three notions of reason-based representability:

Fact: Any U-representable rightness function is also D-representable, and any dilemma-free D-representable rightness function is also M-representable (relative to the same $\mathcal{P}$).

It is easy to see that any dilemma-free rightness function R that is D-representable is also M-representable: any reasons structure that D-represents R M-represents it as well.[71] Since $R(K)$ is non-empty in each context K, an option's normatively relevant properties defeat those of *all* available options in K if and only if they defeat those of *as many available options as possible*. The claim that U-representability implies D-representability is less trivial. Suppose a reasons structure $\mathcal{R} = \langle N, \trianglerighteq \rangle$ U-represents a rightness function R. We can then convert $\mathcal{R}$ into a reasons structure that D-represents R. To do so, we leave the normative relevance function N unchanged, but replace the relation $\trianglerighteq$ with a modified relation $\trianglerighteq^*$: for any sets of properties S and S',

- $S \equiv^* S'$, in case S and S' are incomparable under $\trianglerighteq$,
- $S \trianglerighteq^* S'$ if and only if $S \trianglerighteq S'$, in case S and S' are comparable under $\trianglerighteq$,

where *comparability* of S and S', as before, means that $S \trianglerighteq S'$ or $S' \trianglerighteq S$ (or both). It is easy to verify that the reasons structure $\langle N, \trianglerighteq^* \rangle$ D-represents R. The present argument also shows that U-representability implies D-representability by a reasons structure *with*

[71]To be precise, this claim requires the number of available options in each context to be finite.

a complete weighing relation. And since the defeating and undefeated criteria are equivalent when $\trianglerighteq$ is complete, it follows that U-representability *simpliciter* is deontically equivalent to D-representability by a reasons structure with a complete weighing relation.

The fact that all U-representable rightness functions are also D-representable does not make the notion of U-representability redundant, since a U-representation might be more faithful to the intended moral theory than a D-representation, as in our incommensurability example. Analogously, D-representations are not made redundant by the existence of deontically equivalent M-representations (assuming dilemma-freeness). The three permissibility criteria are interpretationally distinct and therefore non-interchangeable.

Appendix C: Two dimensions of relativism

We have recognized two possible dimensions of universalism and relativism, the *normative-relevance* and *weighing dimensions*, though our focus has been on the first. We now introduce a generalized definition of a reasons structure that allows us to formalize both dimensions. We then discuss a motivating example for relativism along the second dimension; and we finally show that many (though not all) instances of weighing relativism can be re-expressed in a weighing-universalist format (e.g., by shifting the relativism into the normative relevance function).[72]

C.1 A generalized reasons structure

A *generalized reasons structure* is a pair $\mathcal{R} = \langle N, (\trianglerighteq_K)_{K \in \mathcal{K}} \rangle$ consisting of:

- A *normative relevance function*, denoted N, which assigns to each context $K \in \mathcal{K}$ a set $N(K)$ of *normatively relevant* properties in that context.
- A *family of weighing relations* over sets of properties, denoted $(\trianglerighteq_K)_{K \in \mathcal{K}}$, consisting of one weighing relation $\trianglerighteq_K$ for each context K in $\mathcal{K}$, where $\trianglerighteq_K$, as before, is a binary relation whose relata are subsets of $\mathcal{P}$.

Again, we introduce an invariance constraint: whenever two contexts K and K' have the same context properties, i.e., $\mathcal{P}(K) = \mathcal{P}(K')$, the normatively relevant properties and weighing relations are also the same, i.e., $N(K) = N(K')$ and $\trianglerighteq_K = \trianglerighteq_{K'}$.

A rightness function R has a *generalized reason-based representation* if there exists some $\mathcal{R} = \langle N, (\trianglerighteq_K)_{K \in \mathcal{K}} \rangle$ such that, for each context K,

$$R(K) = \{x \in [K] : N(x, K) \trianglerighteq_K N(y, K) \text{ for all } y \in [K]\}.$$

[72]We are grateful to John Broome for suggesting that we consider weighing relativism as an alternative to normative-relevance relativism and for suggesting the motivating example.

We call this representation *normative-relevance-universalist* if N is constant and *normative-relevance-relativist* if not. We call the representation *weighing-universalist* if $\trianglerighteq_K$ is the same for all K and *weighing-relativist* if not.

C.2 An example

Consider a moral theory that says: parents should promote the welfare of all children, but they should promote the welfare of their own children twice as much as the welfare of others' children. Specifically, suppose options are pairs of the form (w_a, w_b), where w_a and w_b are the welfare levels of little Alfie and little Bertie. Further, Alf is Alfie's parent, while Bert is Bertie's parent. Then our moral theory says that Alf should always choose an option (among the available ones) that maximizes $2w_a + w_b$, while Bert should always choose one that maximizes $w_a + 2w_b$. (Choice contexts are of the form $K = \langle i, Y \rangle$, where i is one of the parents and Y is the set of feasible options.)

This theory can be represented by a generalized reasons structure in which the weighing relation depends on the context. We invoke properties of the forms

"Alfie's welfare is w_a" (A_{w_a}), "Bertie's welfare is w_b" (B_{w_b}), and "the agent is i" (Ag_i).

Properties of the first two kinds, which we may call *children's welfare properties*, are option properties. Properties of the last kind, which we may call *agent properties*, are context properties. Now let $\mathcal{R} = \langle N, (\trianglerighteq_K)_{K \in \mathcal{K}} \rangle$, where:

- for every context K, $N(K)$ is the set of all children's welfare properties;
- for any context $K = \langle i, Y \rangle$,
  - if K satisfies Ag_{Alf}, the weighing relation $\trianglerighteq_K$ is such that

    $$\{A_{w_a}, B_{w_b}\} \trianglerighteq_K \{A_{w'_a}, B_{w'_b}\} \text{ if and only if } 2w_a + w_b \geq 2w'_a + w'_b;$$

  - if K satisfies Ag_{Bert}, the weighing relation $\trianglerighteq_K$ is such that

    $$\{A_{w_a}, B_{w_b}\} \trianglerighteq_K \{A_{w'_a}, B_{w'_b}\} \text{ if and only if } w_a + 2w_b \geq w'_a + 2w'_b.$$

This generalized reasons structure represents the given moral theory in a weighing-relativist way. It deems the same option properties normatively relevant in all contexts, but the weighing relation changes, depending on who the agent is. Thus represented, the theory is consequentialist, weighing-relativist, and normative-relevance-universalist.

However, the given theory can also be represented by an ordinary reasons structure, without a variable weighing relation. For example, using the same properties as before, we may define a reasons structure $\mathcal{R} = \langle N, \trianglerighteq \rangle$ as follows:

- for every context K, $N(K)$ is the set of all children's welfare properties and all agent properties;
- the weighing relation $\trianglerighteq$ is such that
  - $\{A_{w_a}, B_{w_b}, Ag_{\text{Alf}}\} \trianglerighteq \{A_{w'_a}, B_{w'_b}, Ag_{\text{Alf}}\}$ if and only if $2w_a + w_b \geq 2w'_a + w'_b$;
  - $\{A_{w_a}, B_{w_b}, Ag_{\text{Bert}}\} \trianglerighteq \{A_{w'_a}, B_{w'_b}, Ag_{\text{Bert}}\}$ if and only if $w_a + 2w_b \geq w'_a + 2w'_b$.

This reasons structure represents our moral theory in a non-consequentialist agent-relativist format, as discussed in Section 5.3. However, we can make an interesting observation: there is a sense in which, *conditional on the agent*, the reasons structure becomes consequentialist. Consider the *restriction* of $\mathcal{R}$ to any domain $\mathcal{K}'$ of contexts in which the agent is held fixed, i.e., either all contexts in which Alf is the agent, or all contexts in which Bert is the agent.[73] Then this restricted reasons structure, denoted $\mathcal{R}|_{\mathcal{K}'}$, is "essentially" consequentialist, in the sense that all context-related properties that were normatively relevant in the original reasons structure $\mathcal{R}$ are redundant in $\mathcal{R}|_{\mathcal{K}'}$. (A property P is *redundant* in a reasons structure if it is contained in either all or none of the options' sets of normatively relevant properties and hence can be "deleted" from the reasons structure without changing the entailed deontic content.[74])

A third representation of the given moral theory can be obtained by invoking properties of the forms

"The welfare of the agent's child is w_1" (C_{w_1}), "The other child's welfare is w_2" (O_{w_2}),

where C_w and O_w are relational properties referring to Alfie's and Bertie's welfare when Alf is the agent, and to Bertie's and Alfie's welfare when Bert is the agent. Now we can define $\mathcal{R} = \langle N, \trianglerighteq \rangle$ as follows:

- for every context K, $N(K)$ is the set of all properties of the forms C_{w_1} and O_{w_2};
- the weighing relation $\trianglerighteq$ is such that $\{A_{w_1}, O_{w_2}\} \trianglerighteq \{A_{w'_1}, O_{w'_2}\}$ if and only if $2w_1 + w_2 \geq 2w'_1 + w'_2$.

[73] The *restriction* of a reasons structure $\mathcal{R} = \langle N, \trianglerighteq \rangle$ to a subset $\mathcal{K}'$ of $\mathcal{K}$ is the reasons structure $\mathcal{R}|_{\mathcal{K}'} = \langle N|_{\mathcal{K}'}, \trianglerighteq|_{\mathcal{K}'} \rangle$ where $N|_{\mathcal{K}'}$ is the restriction of the function N to the subdomain $\mathcal{K}'$ and $\trianglerighteq|_{\mathcal{K}'}$ is the restriction of the relation $\trianglerighteq$ to relata that are subsets of some $N(K)$ where K belongs $\mathcal{K}'$.

[74] In our example, $\mathcal{R}|_{\mathcal{K}'}$ is deontically equivalent to a reasons structure $\langle N', \trianglerighteq' \rangle$ on the domain $\mathcal{K}'$ arising from $\langle N|_{\mathcal{K}'}, \trianglerighteq|_{\mathcal{K}'} \rangle$ by deleting every occurrence of $P \in \{Ag_{\text{Alf}}, Ag_{\text{Bert}}\}$. Specifically, $N'(K) = N(K) \backslash \{Ag_{\text{Alf}}, Ag_{\text{Bert}}\}$ for every K in $\mathcal{K}'$; and for any property sets S and S' containing exactly one of Ag_{Alf} and Ag_{Bert}, we have $S \backslash \{Ag_{\text{Alf}}, Ag_{\text{Bert}}\} \trianglerighteq' S' \backslash \{Ag_{\text{Alf}}, Ag_{\text{Bert}}\}$ if and only if $S \trianglerighteq|_{\mathcal{K}'} S'$.

Again, this yields a representation of our moral theory in a non-consequentialist agent-relativist format. Once more, the reasons structure becomes consequentialist, conditional on the agent. In fact, the restriction of $\mathcal{R}$ to contexts in which the agent is held fixed is consequentialist in the standard sense, even without any technical move of setting aside redundant properties. This is because, on any domain of contexts in which the agent does not vary, C_{w_1} and O_{w_2} will each be formally reclassified as option properties.[75]

C.3 An equivalence result

We now show that many instances of weighing relativism can be formally re-expressed in a weighing-universalist format, by suitably redefining the normative relevance function (but without changing the set $\mathcal{P}$ of admissible properties). We offer a construction in which the relativism shifts from the weighing dimension to the normative-relevance dimension (though it is also possible to eliminate relativism on both dimensions).

Call a generalized reasons structure $\mathcal{R} = \langle N, (\trianglerighteq_K)_{K \in \mathcal{K}} \rangle$ *non-entangled* if it never deems any context properties normatively relevant that also make a difference to the weighing relation. Formally, let Z be the set of those context properties that are not normatively relevant in any context; then $\mathcal{R}$ is *non-entangled* if $\trianglerighteq_K = \trianglerighteq_{K'}$ for any K and K' with $\mathcal{P}(K) \cap Z = \mathcal{P}(K') \cap Z$. Any consequentialist reasons structure is trivially non-entangled, because it deems no context properties normatively relevant at all.

Fact: Any non-entangled generalized reasons structure $\mathcal{R} = \langle N, (\trianglerighteq_K)_{K \in \mathcal{K}} \rangle$ is deontically equivalent to an ordinary reasons structure $\mathcal{R}' = \langle N', \trianglerighteq' \rangle$ where $N'(K) \supseteq N(K)$ for all K.

To establish this fact, consider any generalized reasons structure $\mathcal{R} = \langle N, (\trianglerighteq_K)_{K \in \mathcal{K}} \rangle$ with a non-constant weighing relation. Define an ordinary reasons structure $\mathcal{R}' = \langle N', \trianglerighteq' \rangle$ for the same set $\mathcal{P}$ as follows:

- for any context K, $N'(K) = N(K) \cup Z(K)$, where $Z(K)$ is the set of context properties in Z that are satisfied by K;
- for any sets of properties S and T, $S \trianglerighteq' T$ if and only if $S' \trianglerighteq_K T'$ for some context K where $S = S' \cup Z(K)$, $T = T' \cup Z(K)$, and S' and T' are disjoint from Z.

It is easy to verify that $\mathcal{R}'$ entails the same rightness function as $\mathcal{R}$. Note further that, while $\mathcal{R}$ is weighing-relativist, $\mathcal{R}'$ is normative-relevance relativist. By slightly amending the construction of $\mathcal{R}'$ or combining the present equivalence result with the earlier one from Section 6.2, one can also see that any non-entangled generalized reasons structure

[75] This follows from our definitions in footnotes 17-19, applied to the appropriate subdomain $\mathcal{K}'$.

that is weighing-relativist is deontically equivalent to an ordinary reasons structure that is universalist also with respect to the normative relevance function.

Appendix D: Moral choices under uncertainty

In this final appendix, we discuss two ways in which our framework can capture moral choices under uncertainty. The first is to take the options of choice to be *lotteries* (also known as "gambles" or "risky prospects"), i.e., probability distributions over outcomes. The second is to take the options to be *Savage acts*, i.e., functions from states of the world to outcomes, where there is uncertainty about the state of the world. These two approaches follow, respectively, von Neumann and Morgenstern (1944) and Savage (1954).

D.1 Options as lotteries

We have not made any special assumptions about the nature of the options that are available in each context K. They are simply drawn from some underlying set X of possible options. This could be any non-empty set. Just as we can take the options to be actions with known consequences, so we can take them to be lotteries. To make this precise, let O be some non-empty set of outcomes, such as payoffs, allocations of goods or welfare, or states of affairs. A *lottery* is a probability distribution over O, i.e., a function that assigns to every outcome a non-negative number (its probability), with a sum-total of 1 across all outcomes.[76] Let X be the set of all lotteries over O. Then, in each choice context K, the set $[K]$ is the set of available lotteries. A rightness function R, as before, assigns to each context the set of those options that are "right" or "permissible" in that context, i.e., it selects the lotteries that may be permissibly chosen.

To illustrate how we can represent some familiar normative theories of choice under uncertainty, consider two examples, each applied to the simple case where each element of the outcome set O is a particular welfare level that is attained under the given outcome.

- **Expected-welfare maximization:** Choose a lottery which maximizes the *expected welfare* (defined as the lottery's probability-weighted average of the welfare levels in the outcome set).
- **Maximin:** Choose a lottery for which the minimum welfare level whose probability is non-zero is maximal.

[76] We here consider the simplest case, in which each lottery assigns non-zero probability to only finitely many outcomes. Of course, our analysis can be extended to measurable sets O.

Each of these theories admits a straightforward reason-based representation. To represent expected-welfare maximization, we must invoke properties of the form

$$P_{\text{exp}=w}: \text{ “The expected welfare is } w\text{”},$$

where w is some real number. Call such properties *expected-welfare properties.* Now the reasons structure is $\mathcal{R} = \langle N, \trianglerighteq \rangle$, where:

- for every context K, $N(K)$ is the set of all expected-welfare properties;
- the weighing relation ranks singleton sets consisting of expected-welfare properties such that $\{P_{\text{exp}=w}\} \trianglerighteq \{P_{\text{exp}=w'}\}$ if and only if $w \geq w'$.

To represent maximin, we must invoke properties of the form

$$P_{\text{min}=w}: \text{ “The minimum welfare whose probability is non-zero is } w\text{”},$$

where w is some real number. Call such properties *minimum-welfare properties.* Now the reasons structure is $\mathcal{R} = \langle N, \trianglerighteq \rangle$, where:

- for every context K, $N(K)$ is the set of all minimum-welfare properties;
- the weighing relation ranks singleton sets consisting of minimum-welfare properties such that $\{P_{\text{min}=w}\} \trianglerighteq \{P_{\text{min}=w'}\}$ if and only if $w \geq w'$.

Both theories count as consequentialist in our taxonomy, since the properties they deem relevant, expected-welfare or minimum-welfare properties, are option properties. Whether a lottery has a property such as $P_{\text{exp}=w}$ or $P_{\text{min}=w}$ does not depend on the context.

We can define many other properties of lotteries, such as their variance, the difference between the best and the worst outcomes they may yield with non-zero probability, or the difference between their expected value and the average expected value across all the available lotteries. While the first two of these examples are still option properties, the last is a relational property; it depends not only on the lottery under consideration but also on the other available lotteries. Formally, the extension of any property in the present case is always the set of those lottery-context pairs that have the property.

Once we understand that options can be lotteries, and that those lotteries can have a variety of properties in each context, it should be clear that normative theories that deal with lotteries are just as easily representable as are theories for which options do not take this form. Our taxonomy continues to apply and can even be extended. For example, one might distinguish between theories according to which only “modal” properties matter – i.e., properties that depend only on what can or cannot happen with non-zero probability

– and theories according to which "probabilistic" properties matter too – i.e., properties which depend also on the numerical probabilities of possible outcomes. The maximin theory is purely modal; expected-welfare maximization is probabilistic.

D.2 Options as Savage acts

While lotteries explicitly encode the probabilities of outcomes, our second approach to capturing uncertainty is different. A *Savage act* is a specification of what will happen in each one of several possible states of the world, given that one performs the act. No assignment of probabilities is built into a Savage act. Formally, we assume that there is a non-empty set S of possible states of the world; these are beyond a given agent's control. The different states could be different weather conditions (e.g., whether there is a drought or not), different economic circumstances (e.g., the level of growth), different health conditions (e.g., who does or does not get ill), or any other external circumstances on which the outcome of an agent's actions may depend. As before, we assume that there is a set O of possible outcomes that might result from taking action. A *Savage act* is a function from the set S into the set O. It assigns to each state of the world the outcome that would result from performing the act in that state. The set X of possible options then consists of all Savage acts, and as before, $[K]$ will always be some subset of X.

For example, suppose there are two states of the world: "flood" and "no flood". The Savage act of "building a flood protection barrier" would yield the outcome "moderate cost" in the state "no flood" (say, a payoff of -100), and the outcome "moderate cost and lucky escape" in the state "flood" (which, say, also generates a payoff of -100). The Savage act of "building no flood protection barrier" would yield the outcome "no cost" in the state "no flood" (say, a payoff of 0), and the outcome "disaster" in the state "flood" (say, a payoff of -1000). Our assessment of those Savage acts may focus on a variety of properties, such as their worst or best possible outcomes, or the amount of "regret" we would experience if our choice led to a bad outcome. Here are two illustrative theories:

- **Maximax:** Choose a Savage act with the greatest best-case scenario. In our example, this theory would recommend building no flood protection barrier.
- **Regret minimization:** Choose a Savage act which minimizes maximal regret, where the *regret* in each state of the world is the discrepancy between the actual outcome in that state and the best possible outcome one could have attained through some available act. In our example, the maximal regret for "building a flood protection barrier" would be 100. This is the regret we would experience in case no flood occurs. By contrast, the maximal regret for "building no flood

protection barrier" would be 900. This is the regret we would experience in case a flood occurs. So, the theory would recommend building a flood protection barrier.

Both of these theories can be represented in our framework. To represent maximax, we simply need to invoke properties of the form

$$P_{\text{max}=x}: \text{ "The maximum payoff is } x\text{"},$$

where x is some real number. Call such properties *best-case properties*. The reasons structure of the maximax theory is $\mathcal{R} = \langle N, \trianglerighteq \rangle$, where:

- for every context K, $N(K)$ is the set of all best-case properties;
- the weighing relation ranks singleton sets consisting of best-case properties such that $\{P_{\text{max}=x}\} \trianglerighteq \{P_{\text{max}=x'}\}$ if and only if $x \geq x'$.

To represent the regret-minimization theory, we need to invoke properties of the form

$$P_{\text{reg}(s)=r}: \text{ "The regret in state } s \text{ is } r\text{"},$$

where r is some non-negative real number and s is some state of the world in S. Call such properties *regret properties*. Each act-context pair has as many regret properties as there are states in S: one property specifying what the regret would be in each state, if the act were performed. Now the reasons structure is $\mathcal{R} = \langle N, \trianglerighteq \rangle$, where:

- for every context K, $N(K)$ is the set of all regret properties;
- the weighing relation ranks sets of regret properties such that, for any two sets A and B, we have $A \trianglerighteq B$ if and only if $\max(r : P_{\text{reg}(s)=r} \in A \text{ for some } s \in S) \leq \max(r : P_{\text{reg}(s)=r} \in B \text{ for some } s \in S)$.

Again, the two theories can be readily categorized in our framework. Maximax is a consequentialist theory, since the best-case properties are option properties. An act's best-case outcome does not depend on the other available acts. Regret minimization is a non-consequentialist theory. The regret properties of each act depend on the other available acts and are thus relational. However, like maximax, the regret-minimization theory is universalist: it deems the same properties normatively relevant in all contexts.

As should be clear, there is no barrier to representing normative theories of choice under uncertainty in our framework.[77] Our point has simply been to give a "proof of concept". A further analysis of reason-based choice under uncertainty is left for future work.

[77] Another example of such a theory is Lara Buchak's risk-weighted expected utility theory (2013).