Risk and Motivation:
When the Will Is Required
to Determine What to Do

Dylan Murray
Princeton University

Lara Buchak
University of California, Berkeley

© 2019 Dylan Murray & Lara Buchak
This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 3.0 License.
<www.philosophersimprint.org/019016/>

1. Introduction

Decision theory and philosophy of action both attempt to explain what it is for an ideally rational agent to answer the question “What to do?” From the agent’s point of view, the answer to that question is settled in practical deliberation and motivates her to act. The mental states that determine her answer are the sources of rationalizing explanations of the agent’s behavior. They explain why she performed a given action in terms of why it made sense, from her point of view, to so act. Rationalizing explanations should be contrastive, of the form “Agent S performed action A, rather than actions B, C, or D, because P, Q, and R” where B, C, and D are whatever S takes to be the possible alternatives to A, and P, Q, and R are whichever of S’s deliberative considerations and other factors yield a good explanation.

In philosophy of action, there are three broad views about what determines an answer to “What to do?” (Wallace 2006) — about what types of mental state(s), in addition to beliefs or credences, determine motivation. Humeans hold that desires alone play this role (Nagel 1970: 29; Smith 1987). Rationalists hold that evaluative or normative judgments, or beliefs about what’s good, valuable, or what one has most reason to do, can also motivate one to act. Volitionalists hold that answering “What to do?” requires, perhaps in addition to either or both of the above, some independent state(s) of the will, like intentions or plans, not fully reducible to any combination of desires or judgments.

Humeans and rationalists claim “What to do?” is completely settled by answering “What do I value?” or “What reasons for action are there?” They differ over whether the answer is provided only by desires, or at least in some cases by evaluative/normative judgments, but the positions are similar enough that both find classical expression in, for instance, Davidson (1980). To take the canonical example: Why did the agent open the refrigerator? She desired a beer (or judged she should

1. We borrow the “what to do” terminology from Gibbard (2003).

2. Humeanism in this sense differs from its usual sense in decision theory: the thesis that rationality places no constraints on preferences other than those requiring consistency.
have one), believed there was one in the refrigerator, and this belief-desire pair caused her to open it. Desires (or judgments) and beliefs alone are sufficient to determine rational action.3

Volitionalists, by contrast, hold that the will constitutes a third, separate factor in rational human motivation. Answering the question “What do I value?” or “What reasons are there?” is not enough to settle deliberation and move one to act (Raz 2002: 47–48; Scanlon 2014: 107). Answering this question may render some options eligible and others not, but settling the question “What to do?” requires, in addition, some contribution from an agent’s volitional states. There are other aspects of the will — like willpower and various executive capacities — and how these all relate is an open question (Holton 2009). But most contemporary volitionalists emphasize the functional role characteristic of intentions or plans, especially their role as inputs to further practical reasoning and deliberation, rather than as outputs or conclusions (Bratman 2014: 24–25). Here, we take this functional role and the consistency constraints characteristic of the attitudes that play it to be central to volitional states, which in addition are often directly arrived at or are at least revisable through choice and decision.4 We’ll specifically be concerned with Bratman’s (2007) notion of self-governing policies: plans to give desires or other conative attitudes particular weights in practical deliberation.5

Volitionalists typically invoke two types of case against rationalists and Humeans: (i) akrasia, and (ii) evaluative/normative underdetermination, like that faced by Buridan’s ass. (i) If one can act against one’s all-things-considered judgment — one’s answer to the question “What action do I have most reason to do?” — then some source of motivation must lie outside judgment. (ii) Buridan’s ass — so the story goes — was equidistant from two indistinguishable bales of hay that it took to be equally desirable and valuable, and took there to be equal reason to eat. So it starved to death. A human being would not starve: he would simply choose one or the other bale, and this choice would then explain his action. According to volitionalists, one of the central functional roles of volitional states is that they resolve such impasses, imbuing human agents with “an ability that is basic at the level of common-sense psychology: an ability to decide [or intend] in the face of equidesirability” (Bratman 1987: 11; see also Bratman 2007, 2014).

However, these two phenomena are somewhat unfortunate ways of motivating volitionalism. In akasia, volitional states seem to be more of a hindrance than a necessary component of rational action, since they constitute downstream interference with deliberation’s effect on behavior. In the case of Buridan’s ass and other tie-breaking cases of evaluative/normative underdetermination, volitional states look like mere randomization devices — mental analogues of flipping a coin. Hence, even if the will does play an independent role in determining behavior in these cases, it doesn’t appear to be of any central importance to rational agency: the will is at best a relatively peripheral contributor to rational agency, and at worst a hindrance to it.

Furthermore, it may be argued that genuine underdetermination is infrequent: something typically tips the scales of values or reasons. In other words, underdetermination arguments for volitionalism are traditionally resisted on the grounds that such cases are (i) rare, or that whatever resolves them is (ii) arbitrary or random, or (iii) external to the agent’s perspective, in which case they fail to show us anything about what rational human motivation typically consists in. So Humeans and rationalists, at least, are apt to argue.6

3. Davidson (1980) extends his theory to include motivation by intention, but construes this to be a type of evaluative/normative judgment or belief.

4. By contrast, Hieronymi (2006) argues that not even intentions are under one’s direct voluntary control.

5. See also Nozick (1981: 294–316), who discusses decisions about how to weight reasons, and Chang (2009), who stresses the role of underdetermination.

6. To the extent that underdetermination has been discussed in decision theory, it has received similar treatment. Ullman-Margalit and Morgenbesser (1977: 773), for instance, claim that Buridan’s-ass-like cases involve mere picking, rather than choosing. Only the latter type of selection is determined by one’s preferences. In cases of picking, the agent herself may select an option, but only by being “transformed into a chance device that functions at random and effects arbitrary selections.”
Decision theory isn’t focused on questions about what kinds of mental states determine the sources of rationalizing explanations, but we can translate that philosophy of action debate into its framework. In decision theory, an agent’s answer to “What to do?” is determined by her preferences. The Humean is naturally interpreted as claiming that an agent’s preferences are completely determined by (i) her beliefs or credences, represented by a probability function of states of the world, \( p \), and (ii) her desires or the strengths thereof, represented by a utility function of outcomes, \( u \). The rationalist is naturally interpreted as claiming that \( u \) represents evaluative/normative judgments or beliefs (perhaps in addition to desires). The voluntalist, in contrast, claims that we need an additional, irreducible third component — one that represents the agent’s will.

Translation in hand, we’re in a position to state our thesis: that recent work in decision theory provides direct, independently motivated support for voluntalism. Specifically, Risk-Weighted Expected Utility (REU) Theory holds that \( u \) and \( p \) are not enough to determine preferences in cases of choice under uncertainty; in addition, an agent must also determine her attitude towards risk (represented by a risk function of probabilities, \( r \) (Buchak 2013). As reviewed in section II, Buchak shows that an agent’s risk function plays a crucial role in determining genuine preferences, a role that goes beyond the mere pickings or tie-breakings present in the cases normally employed by voluntalists. In section III, we argue that \( r \) represents mental states of precisely the kind voluntalists like Bratman appeal to. Thus, REU theory motivates underdetermination arguments for voluntalism that escape the three main objections to traditional versions. We discuss ways that Humeans and rationalists might resist this argument in section IV.

II. REU Theory

Traditional decision theory — Expected Utility (EU) Theory — holds that \( p \) and \( u \) completely determine a rational agent’s preference ordering: one must prefer the gamble with the highest expected utility, relative to one’s beliefs and desires. By contrast, REU theory holds that it is up to an individual agent how to aggregate the utilities of the different outcomes she might receive from a gamble in order to determine its total value: she need not average. Furthermore, two agents with the same probability and utility functions are not required to have the same preference ordering over gambles. In this sense, preferences and choices are underdetermined by beliefs and desires.

Consider a simple example. Suppose you’re asked to choose between a gamble in which you’ll receive a particular sum of money for sure and a gamble involving a coin flip. If the coin lands HEADS, you get $100; if TAILS, nothing. Assuming your utility function is linear in money (such that \( u(x) = x \)) and you believe the coin is fair (such that \( p(\text{HEADS}) = p(\text{TAILS}) = 1/2 \)), EU theory requires you to be indifferent between the coin flip and a gamble in which you get $50 for sure, since the gambles have equal expected utility. You are irrational if you prefer one gamble to the other. EU theory also requires of you, on pain of irrationality, to prefer $55 for sure to the coin flip and the coin flip to a sure thing of $45. These verdicts seem draconian. It seems rationally permissible to prefer a sure thing of $45 (or even, say, $40) over the gamble, since the gamble involves the risk of getting nothing; and it also seems permissible to instead prefer the coin flip to a sure thing of $55 (or even, say, $60), since the gamble gives one the chance at $100.

Or, consider a choice about your career: should you pursue becoming a rock star or becoming an accountant? Suppose you know the exact utilities and probabilities involved: the utility of becoming a rock star is much higher, but it is much more probable that you will succeed at becoming an accountant. Even after you know these values, it is still a rationally open question which career to pursue: “What to do: take a shot at becoming a rock star, or play it safe and pursue accountancy?”

EU theory claims this question is already answered: once \( u \) and \( p \) are not incommensurable (the existence of which REU theory does not rely on).

7. We take no stand on whether the utility function exists independently of preferences or is constructed from them, although for ease of exposition we will talk as if the utility function exists independently of preferences. See Buchak (2013: 16–20) for further discussion.

8. Suppose this is not a case of incommensurability (the existence of which REU theory does not rely on).
determined, your preferences are also determined — you must prefer whichever has the higher expected utility, and there is no further determination to make.

EU theory is inflexible in this way because it takes there to be only one rational way to aggregate the utilities of outcomes to determine the overall utility of a gamble: weighted averaging, where the weight of each outcome’s utility value is the probability of realizing that outcome. (Thus, the utility values in HEADS and TAILS get equal weight in determining your preferences, and the utility values of succeeding as a rock star and as an accountant get weight equal to their respective probabilities.) By contrast, REU theory holds that there is a range of rationally permissible ways to aggregate. Specifically, you might rationally place more weight on worse scenarios than better scenarios, or vice versa, even if you assign these states equal probability. Since the best-case and worst-case scenarios in the coin flip are respectively both better and worse than the sure thing, you don’t have to be indifferent between the coin flip and the sure thing: you can prefer one to the other. We turn to the formal details first.

To simplify the discussion, we employ the notion of an ordered gamble, represented as $g = (E_1, x_1; E_2, x_2; \ldots; E_n, x_n)$, where, for each $i$, $x_i$ is the outcome yielded by event $E_i$ (events being mutually exclusive and exhaustive), and $u(x_i) \leq u(x_{i+1})$ for all $i < n$. Every gamble can be represented as an ordered gamble — in the case above, e.g., the gambles are \{TAILS, $0;$ HEADS, $100\} and \{TAILS, $50;$ HEADS, $50\}.$\footnote{The sure thing can also be represented as \{HEADS, $50; TAILS, $50\} or as \{HEADS OR TAILS, $50\}.}

EU theory requires choosing the gamble with the maximum expected utility:

$$EU(g) = \sum_{i=1}^{n} p(E_i) u(x_i)$$

This latter equation ranks the utilities of outcomes and weights the differences between adjacent outcomes by the probability that an outcome at least as good as the better one will obtain (assigning a weight of 1 to the worst outcome). The value of a gamble is: its worst possible value; plus the interval difference between the worst value and the second worst value, weighted by the probability of getting at least the second worst value; plus the interval difference between the second worst and the third worst value, weighted by the probability of getting at least the third worst value; and so forth. In other words, the weight of each consideration of the form $I$ might obtain benefits of a certain size (in addition to whatever other benefits $I$ obtain) just is the probability of obtaining those (additional) benefits. For instance, suppose you can receive a C, B, or A on a paper with respective probability 1/2, 1/4, and 1/4. Then the expected utility of writing the paper is:

$$EU(paper) = u(C) + \frac{1}{2} (u(B) - u(C)) + \frac{1}{4} (u(A) - u(B))$$

One will receive a utility value at least equivalent to that of getting a C with probability 1; one might obtain additional utility, the difference between that of a B and a C, with probability 1/2 (the probability of getting at least a B); and one might obtain even more utility, the difference between an A and a B, with probability 1/4 (the probability of
getting an A); and each of these possible benefits is weighted by the probability of obtaining it. Again assuming a utility function linear in money, in our first example the EU of the coin flip is \((1/2)(0) + (1/2)(100)\) = 50. And the EU of a sure-thing $x is, trivially, \(x = (1/2)(x - x)\) = x. Hence, one must prefer the coin flip to any sure thing of less than $50, prefer any sure thing of greater than $50 to the coin flip, and be indifferent between the coin flip and a sure thing of exactly $50. Any other preferences are irrational.

In contrast, REU theory holds that the weights given to the benefits one might obtain — to utility differences between outcomes — do not have to be identical to the probabilities of obtaining them. While the fact that the coin flip could (with probability 1/2) yield $100 rather than $0 might be given half the weight in your evaluation of the gamble, it might instead be given more weight or less weight. That is, it’s up to each individual how to weight the benefits received in some states but not others, and so it’s up to each individual how to aggregate the utilities of outcomes to determine the utility of a whole gamble. REU theory requires choosing the gamble with the maximum risk-weighted expected utility:

\[
REU(g) = x_1 + \sum_{j=2}^{n} r\left(\sum_{i=j}^{n} p(E_i)\right)(u(x_j) - u(x_{j-1}))
\]

This equation weights the differences between adjacent outcomes by a function of the probability that an outcome at least as good as the better outcome will obtain. Thus, the weight of each consideration of the form I might obtain benefits of a certain size (in addition to whatever other benefits I obtain) is up to you.

For example:

\[
REU(paper) = u(C) + r\left(\frac{1}{2}\right)\left(u(B) - u(C)\right) + r\left(\frac{1}{4}\right)\left(u(A) - u(B)\right)
\]

You might decide that as benefits are obtained in a smaller and smaller portion of states, they get proportionally less and less weight (for example, that benefits received in the top 1/2 of states only get weight 1/4 and benefits received in the top 1/4 of states only get weight 1/16). In this case, worse outcomes will count proportionally more, and better outcomes will count proportionally less: you will be risk-avoidant. Alternatively, you might decide that as benefits are obtained in a smaller and smaller portion of states, they get proportionally more and more weight (for example, that benefits received in the top 1/2 of states get weight 1/√2 and benefits received in the top 1/4 of states get weight 1/2). In this case, worse outcomes will count proportionally less and better outcomes will count proportionally more: you will be risk-inclined.

Equivalently — formulating the REU equation in terms of weightings of utility values (rather than differences between them) — the weight that each outcome’s value gets in an agent’s evaluation of the whole gamble depends both on its probability and its position in the gamble. On REU theory, for instance, you are free to weight the outcome in the worst-case scenario twice as heavily as the one in the best-case scenario. In that case, the REU of the coin flip would be \((2/3)(0) + (1/3)(100)\) = 33, such that you’d prefer a sure thing of $45 (or even $40). You are also free to do the opposite, and weight the outcome in the best-case scenario twice as heavily as that in the worst-case scenario, in which case you would prefer the coin flip to a sure thing of $55 (or even $60). Thus, on either formulation of REU theory, weights are up to the individual, and there are two ways to think about what you’re weighting: outcomes themselves or incremental benefits of outcomes.

---

10. There are some constraints on the risk function; for example that it output values in \([0, 1]\), that \(r(0) = 0\), that \(r(1) = 1\), and that \(r\) be non-decreasing (Buchak 2013: 60–70).
REU theory thus exonerates many preferences condemned to be irrational by EU theory.

These “weightings” are represented by a risk function of probabilities: \( r(p) \) represents the weight attached to the top \( p \)-portion of outcomes, or how much one is willing to trade off making a worse state less bad against making a better state even better—the extent to which one is willing to take risks. REU theory therefore maintains that an ideally rational agent’s preferences are determined by a utility function of outcomes, a probability function of states, and a risk function of probabilities. An individual need not explicitly determine these three functions \((u, p, \text{and } r)\) in the sense of consciously crunching the numbers. But Buchak’s (2013: Ch. 3) Representation Theorem shows that each of these entities corresponds to a feature that must be fixed in order to settle an agent’s preferences.

Important for the debate in philosophy of action is the particular way in which REU theory allows individuals to deviate from EU theory. Rather than introducing a new value that individuals might care about—the “riskiness” of a gamble—in addition to its expected utility, REU theory claims that rationality does not dictate how to weight the utilities of outcomes. Weighting utilities or incremental benefits by their probabilities (maximizing expected utility) is just one way among many that a rational agent could choose to aggregate utility values. Thus, REU (like EU) theory holds that there is only one source of value—outcomes, whose values are measured by utilities; but (unlike EU theory) it holds it to be an open question, even for ideally rational agents, how to move from the values of outcomes to a single value for a gamble as a whole.

Thus, \( u, p \) underdetermine preferences. Consider making the coin flip choice yourself—or even more starkly, the choice between careers. According to EU theory, once you know the utility of each outcome and its probability of occurring, your rational preference between the two gambles is fixed, completely determined: there’s nothing left for you to decide. But contra EU theory, there is still something you need to make up your mind about: how to structure your aims and balance the interests of your possible future selves (e.g., whether to ensure that your worst-off possible future self at least gets something, or whether to ensure that your best-off possible future self is even better off—gets $100 or becomes famous, but at the expense of other possible future selves ending up with nothing). What \( r \) represents is this balance (Buchak 2013: 55).

In every case of choice under uncertainty in which no gamble dominates (yields at least as good an outcome in every possible state of the world and a better outcome in some possible state than all other gambles; a rare occurrence), one must determine \( r \). Even if one makes up one’s mind to set \( r(p) = p \) (where EU is a special case of REU), one must still decide to do so: \( r(p) = p \) is not a default. Nor does the need to determine \( r \) only arise in cases where one has difficulty determining the precise values of \( u \) and \( p \). Again, Buchak shows that the need to fix \( r \) arises even in cases where one knows the exact utilities of the outcomes and their probabilities of obtaining.

Translating back from decision theory into philosophy of action, it seems motivation is typically underdetermined by beliefs and desires or judgments, contra Humeanism and rationalism. And this isn’t true of just a few isolated tie-breaking cases. Whenever one is choosing between gambles none of which dominates, rational action cannot be determined without determining \( r \). Even after the question of “What do I value?” or “What reasons are there?” has been answered (i.e., even after \( p \) and \( u \) have been fixed), there is still a question of “How to aggregate?” (“What is \( r \) to be?”) that must be answered before one has settled what to do. Thus, it seems that value and reason alone are not enough to provide contrastive explanations of most rational action.

III. Volitionalism

Buchak (2013) holds that the risk function represents risk attitudes, which are not beliefs, desires, or—we add—evaluative or normative judgments. The utility function already represents whatever attitudes one has toward the values of outcomes. And the role of risk attitudes cannot be re-characterized by construing them as contributing to what
Risk and Motivation

one takes to be valuable in the “first-order” sense that the utility function measures (the desirability or goodness of a state of the world, or the reasons for pursuing it), because outcomes are already individuated finely enough to capture everything an agent cares about (including, e.g., the history of how the agent obtained an outcome).11

Instead, risk attitudes determine how the values of outcomes contribute to preferences between gambles. They do not represent values or reasons themselves, but instead weight how much influence to give extant values in answering “What to do?” For example, given that $100 is desirable to a particular extent, or given the reasons I have to pursue a state of the world in which I have $100, how much should some action’s having a 50% chance of yielding $100 count in determining whether to perform that action? Buchak notes that risk attitudes correspond to familiar virtues or character traits: how one trades off being prudent (prioritizing a high minimum) against being venturesome (prioritizing a high maximum). Folk psychology arguably recognizes such attitudes as states of the will — as settled policies or higher-order plans — and they are exactly the kind of attitude that volitionalists have focused on.12

In particular, volitionalists tend to posit higher-order attitudes that share with first-order desires a “world-to-mind” direction of fit, in contrast to beliefs’ “mind-to-world” direction of fit, but where the relevant parts of the “world” are one’s other mental states: for example, a volition (or second-order desire) for a first-order desire to move one to action (Frankfurt 1988) or, more pertinently, a self-governing policy to treat

Risk and Motivation

a first-order desire, D, as having a particular weight, W, in practical deliberation (Bratman 2007: 240). Thus, these higher-order attitudes are plans about one’s own mental states. In contrast, the attitudes that both Humeans and rationalists tend to appeal to are first-order attitudes — responses to properties (values or reasons) of outcomes themselves, not higher-order weightings of such responses.

Risk attitudes are higher-order weightings of first-order desires. Risk attitudes are represented by a function, r, that takes probabilities of outcomes as input and, as output, gives particular weights to desires for outcomes (or their values). Therefore, r represents mental states that have the higher-order “world-to-mind” direction of fit characteristic of many volitional states. Specifically, risk attitudes fix how much a given desire for some outcome counts in practical deliberation, or how much a given practical reason counts in favor of a particular gamble. (For the risk-avoidant agent, e.g., a strong desire for an outcome with given probability might not count strongly in favor of taking a particular gamble if that outcome obtains only in the best possible state of the gamble, since such an agent cares much more about what happens in worse states of the gamble.) Hence, risk attitudes seem to be one type of self-governing policy in Bratman’s (2007) sense: a particular setting of r is a plan to give first-order desire D (the u portion of the REU equation) weight W (the r portion) in practical deliberation.

The downstream effect that risk attitudes have on further practical reasoning supports this conclusion. Consider again the choice between pursuing rock stardom and accountancy. There is some reason to try to become a rock star: the possibility of fame and fortune. And there is some reason against so doing: the possibility of failing and ending up unknown and destitute. Before the agent has determined how to weight these considerations against each other, it doesn’t seem irrational to hold intentions or preferences to perform additional actions that are incompatible with choosing to pursue one or the other career (so long as they are not incompatible with both). Suppose, though, that the agent comes to settle on being risk-inclined and prefers to pursue a career as a rock star. At that point, it seems to become irrational for

12. Note also the similarity to doxastic voluntarism. James (1896/1979) maintains that the will determines the tradeoff between two epistemic virtues: avoiding false beliefs and gaining true ones. (One might, e.g., weight gaining true beliefs more heavily than avoiding false ones by setting the threshold of evidence required for belief very low, making almost any evidence sufficient for belief. At the other extreme, one might weight avoiding false beliefs much more heavily than gaining true ones by setting a higher threshold, making one more skeptical.) Those who think the balance between the practical virtues of prudence and venturesomeness is struck by the will are, in our sense, volitionalists. See also Buchak (2013: 55–56).
the agent to hold intentions or preferences that are inconsistent with becoming a rock star, whereas these were not irrational prior to forming that risk attitude. This is one role that Bratman (1987: 23) identifies as characteristic of intentions or plans generally, *qua* states of the will independent of, and irreducible to, belief-desire pairs.

Bratman (1987, 2007) claims that one of the characteristic downstream roles of self-governing policies, specifically, on practical deliberation lies in resolving impasses of evaluative/normative underdetermination, sometimes in life-changing choices like whether to join the Free French or to care for one’s ailing mother (Sartre 1956/1975). However, the traditional worries about underdetermination arguments remain even for Bratman’s self-governing policies before more is said about their role in deliberation. Humeans and rationalists may claim that even in genuine cases of underdetermination, self-governing policies are external to the agent’s deliberation about “what to do” or that they still involve mere “picking” (even if of a life-changing sort).

Perhaps the most important upshot of REU theory for action theory lies in showing that there is a type of self-governing policy that doesn’t fall prey to these worries. REU theory establishes that a third type of attitude, beyond those represented by *u* and *p*, is required for far more than cases of tie-breaking (i.e., cases in which the options have equal EU). Crucially, *every choice under uncertainty* in which no gamble dominates requires settling one’s risk attitudes. In this sense, most choices are underdetermined. In giving contrastive rationalizing explanations of (ideally rational) agents’ behavior, we have to cite their risk attitudes *all the time*. Being a rational agent *characteristically* involves attitudes in addition to beliefs, desires, and judgments.

Making up one’s mind about how to trade off the fact that one gamble does better in one state of the world against the fact that another gamble does better in another state is a fundamental feature of determining one’s preferences and thus one’s choices about how to act. Risk attitudes are neither handicaps nor arbitrary selection gadgets.  

Folk psychology recognizes the character traits of being prudent or venturesome as important aspects of one’s identity *qua* practical agent. Settling what values *r* takes is a genuine constituent of one’s answer to “What to do?” Risk attitudes are not downstream from or outside the perspective an agent herself takes in practically deliberating.

Thus, REU theory motivates a new underdetermination argument for volitionality that escapes the traditional objections to such arguments: that the cases involved are (i) rare, or only involve states of the will in a way that is (ii) arbitrary or random, or (iii) external to the perspective taken by the agent herself in determining what to do. These objections do not apply to risk attitudes, so if risk attitudes are self-governing policies — states of the will — then REU theory vindicates volitionality.

There are of course new objections to the REU-motivated underdetermination argument, which we’ll argue below are less compelling. But what bearing does this conclusion have on the original Buridan-like cases? REU theory doesn’t speak directly to this issue (assuming risk is not a factor for the ass), and we leave it open what other types of self-governing policy might exist. However, REU theory does undermine any general presumption that self-governing policies are *not* at work in cases of underdetermination, and so leaves open that self-governing policies (of some non-risk-related kind) might be involved in impasses of the Buridan-like, tie-breaking variety.

### IV. The Functional Role of Risk Attitudes

Our main thesis is that the risk attitudes identified by REU theory are just the kind of mental state volitionals should appeal to in underdetermination arguments. Decision theory turns out to call for a picture of action that slots directly into the apparatus posited by volitionals, providing it support over rationalism and Humeanism. In particular, we’ve argued that the main functional role played by risk attitudes — weighting first-order desires or other conative attitudes — and their direction of fit matches those of volitional states much better than those of desires or judgments. Risk attitudes are able to break
impasses of evaluative and normative underdetermination in virtue of playing the weighting role characteristic of self-governing policies.

In this section, we first consider objections to the claim that risk attitudes fit the functional role that volitionalists posit, and we address whether risk attitudes can instead be subsumed by rationalism or Humeanism. Second, we discuss several additional reasons to think that risk attitudes are volitional states based on their further functional profile.

One might object that risk attitudes cannot be considered self-governing policies at all, but must instead be thought of as desires for outcomes. This objection can be developed in two ways. First, one might claim that we’ve misidentified risk attitudes — that they’re actually brute feelings or constitutional propensities, like gustatory tastes. Or, second, one might accept risk attitudes of the kind described by REU theory, but maintain that these attitudes are themselves determinants of values or reasons. We consider the misidentification reply first.

The misidentification reply holds that attitudes toward risk play the same role in practical deliberation as other attitudes already countenanced by EU theory: as further first-order values or reasons. Risk-related propensities, in this sense, are already included in one’s utility function — already taken into account by one’s desires or judgments — since they contribute to the value of a given outcome (construed broadly, so as to encompass any values that obtain in virtue of its occurrence, including any that do so in virtue of an agent’s mental states or their relation to the external state of affairs). Enjoying risk gives one reason to take risks, just like enjoying ice cream gives one reason to seek it. Folk psychology recognizes risk-related propensities, and one can try to preserve EU theory by subsuming such propensities in the utilities of outcomes.

We grant there may be risk-related propensities. But Buchak (2013: Ch. 4) shows that these are not the only psychological responses that we have toward risk — that there are also risk attitudes of the sort represented by \( r \). Some cases simply cannot be captured by “locally re-individuating outcomes”. And even after taking into account all values or reasons that accrue in virtue of risk-related propensities (regret, surety, the thrill of gambling) and any of one’s other feelings, tastes, or mental states, one still faces the question of how to aggregate (Buchak 2013: Ch. 5–7).

Second, in a similar vein, one might grant the existence of risk attitudes but maintain that they are determinants of the values of outcomes. According to this reply, an outcome in which one acts in accordance with her risk attitudes is better, or more valuable, than one in which she doesn’t. The volitionalist need not deny this reply, either. However, the primary role that risk attitudes play in practical deliberation and determining one’s preferences is in weighting one’s first-order desires or values, even if such weightings then provide one with reasons to pursue a particular outcome.

Perhaps having certain risk attitudes — e.g., setting \( r(p) = p^r \) — gives one reason to act in ways that maximize EU according to that value, rather than others (e.g., \( p^{r'} \)). And perhaps once one sets \( r(p) = p^r \), one has reason to maintain that setting. But before making up one’s mind about which risk attitude to adopt, there’s no more reason to set \( r(p) \) to \( p^r \) than to (almost) anything else. You already have to weight outcomes in a given way in order for that weighting to provide you with reasons for action or deliberation. But the initial question of how to make up your mind about how to weight outcomes (“How to aggregate?”) is still evaluatively and normatively underdetermined, even including the values of having or acting on one’s risk attitudes, since prior to settling on which to have, all those (within a certain range, at least) are equally valuable.\(^{15}\)

\(^{14}\) Note, though, that these reasons cannot be captured in the utility function, or we would be “double-counting” (Buchak 2013: 137–138).

\(^{15}\) There may be settings of \( r \) at either extreme that are unreasonable (Buchak 2013: 232–234). That is, there may be substantive constraints on rationality that rule out being too risk-avoidant (prudent) or too risk-inclined (venturesome). Similarly, there may be substantive constraints on how extreme the weighting between the virtues of gaining true beliefs and avoiding false ones can be (see n. 12). Perhaps epistemic skepticism goes wrong not because of any mistaken assessment of the evidence, but because it involves a practically irrational plan about where to set the threshold of evidence required
This higher-order underdetermination also suggests that risk attitudes are not identical to higher-order desires or judgments concerning gambles, or concerning certain global properties of gambles (in contrast to first-order desires concerning particular outcomes). Again, prior to fixing \( r \), the question of what exact value to set \( r \) to — of which risk attitude to adopt — is underdetermined by values and reasons of any order. If evaluative/normative judgments and desires are necessarily representations of values or reasons, of whatever order, then risk attitudes cannot be evaluative/normative judgments or desires — prior to fixing \( r \), there are no values or reasons that resolve the underdetermination to be represented.

We’ve argued that the main functional role played by risk attitudes matches that of volitional states — in particular, of self-governing policies — much better than those of beliefs, desires, and judgments. In particular, risk attitudes play the same type of functional role in resolving impasses of underdetermination that is taken to characterize volitional states. Still, much depends on the exact nature of risk attitudes. We don’t propose to settle the debate with Humeans and rationalists here, but we close by discussing two further considerations that favor the volitionalist. First, risk attitudes are directly responsive to one’s decisions, and second, having conflicting risk attitudes seems to be practically irrational in the same way as harboring conflicting volitional states.

Recall that even after you know everything about the relevant values and reasons concerning what happens in each state of affairs if that state is realized — e.g., that you desire being a rock star more than being an accountant, and by how much — that doesn’t bring an end to your deliberations. Rather, you recognize that there’s a further choice you have to make before you’ll be motivated to act: you have to decide how considerations about what happens in each state of affairs will weigh against each other, how to balance prudence and venturesomeness. The phenomenology of making such a decision does not seem to be that of passively “looking inside oneself” to see what kind of person one already is (prudent or venturesome), and then applying the findings. Rather, what envisioning the career example yourself makes especially clear is that many times you, the agent, can only settle the question of “What to do?” by deliberately, reflectively making up your mind about how to make this trade-off and determining which risk attitude to adopt.\(^{16}\) This is precisely why career decisions are so agonizing, even when we know exactly how valuable each kind of career is.

Of course, one need not determine \( r \) anew for every relevant decision: making up one’s mind about which risk attitude to adopt in a given case plausibly carries forward as a standing policy that acts as input to future deliberation and decisions (at least within the same domain). But as the coin flip and career examples illustrate, we sometimes have not considered the question yet — we have no relevant standing policies — and so must actively make up our minds about which risk attitudes to adopt in a way that’s evaluatively and normatively underdetermined. Even where we have standing risk attitudes, we seem to be able to change our minds about them voluntarily, where this type of direct responsiveness to choice is typically associated with volitional states.\(^{17}\)

Second, having inconsistent risk attitudes seems to be irrational in the same way that having inconsistent attitudes with volitional functional roles is irrational. One can desire to go to Cuba and desire to go to Africa for vacation, even though one cannot do both, without being irrational. One is irrational, however, if one intends to do both. Having inconsistent beliefs can also be genuinely irrational, but in a different way. Because beliefs function to represent the world, and the same

---

\(^{16}\) Alternatively, you might simply choose an action you know will maximize REU according to one risk function rather than others. In that case, you do not directly choose between different risk attitudes, but (understand that) your choice will determine which risk attitudes you thereby adopt.

\(^{17}\) However, see n. 4.
world cannot be two incompatible ways, two beliefs are inconsistent in a way amounting to irrationality if they represent the world as being two incompatible ways; at least one such belief must fail to fulfill its function. Bratman (2009a, 2009b, 2009c) traces the consistency constraint on intentions, in contrast, back to an analogous function but with reversed direction of fit: to the role of intentions in constituting the unity of the practical agent.\(^{18}\) One and the same agent cannot be radically disunified: there is pressure to “take a stand” that is coherent and non-self-defeating in all of one’s actions. For human agents, at least, self-governing policies are necessary for taking such a unified stand, and so must be consistent or else fail to fulfill their functional role.  

Plausibly, having one risk attitude that gives a first-order desire, \(D\), weight \(W\) in practical deliberation and another risk attitude that gives \(D\) a different weight not only fails to settle where the agent stands with respect to risk; it seems to be trying to stand in two incompatible places at once—trying to deliberate in impossible or at least unity-attenuating ways. Setting \(r(p)\) equal to \(p^2\) and equal to \(p^{3/2}\) for one and the same domain or decision, for instance, seems strongly self-defeating—one might derail all of one’s possible careers. Thus, having inconsistent risk attitudes, like inconsistent intentions, undermines the practical unity of the agent.

The Humean and rationalist may have replies.\(^{19}\) But the fact that risk attitudes are under the direct influence of decisions and subject to unity-based consistency constraints strengthens the case for their being volitional states, rather than desires or judgments.

The main upshot of REU theory for the debate in philosophy of action about what types of mental state(s) determine motivation lies

---

\(^{18}\) See Kolodny (2005) for skepticism about both types of consistency constraint.

\(^{19}\) Humeans and rationalists, e.g., have worried that the normative consistency constraints on intentions are actually identical to, or derived from, the normative consistency constraints on beliefs (Setiya 2007; Wallace 2006) and others allege that the structuring roles Bratman attributes to self-governing policies can also be played by other types of mental state, including evaluative beliefs and judgments (Wallace 2014).

in showing that there are far more cases of underdetermination than commonly recognized. There are open questions about the further functional profile of risk attitudes, but insofar as it is distinctive of certain volitional states that they resolve cases of underdetermination by weighting one’s first-order desires, and in a way directly subject to choice and unity-based consistency constraints, REU theory vindicates volitionalism.

### V. Conclusion

Volitionalists claim that more goes into motivating rational agents than beliefs and desires or judgments—in particular, that self-governing policies or intentions play an independent role in determining what to do, and that evaluative/normative underdetermination demonstrates as much. We have seen that risk attitudes closely fit this characterization. In the first instance, the contribution that risk attitudes make in determining one’s preferences is in their volitional capacity—qua plans or weightings of first-order states. Risk attitudes are also genuine constituents of one’s preferences, and so underdetermination arguments based on them escape traditional objections to such arguments. In particular, the role of risk attitudes extends far beyond mere cases of tie-breaking. Whenever faced with a choice in which no one gamble dominates, in addition to settling the question “What do I value?” or “What is there most reason to do?” one still has to make up one’s mind about which risk attitude to adopt (one must settle “How to aggregate these values or reasons?”) in order to answer “What to do?” If volitionalists are right that the mental states that resolve such cases of underdetermination are states of the will, then their theory has been starved on a restricted diet of examples. REU theory proffers a cornucopia: the will is required for most actions.\(^{20}\)

\(^{20}\) For helpful comments on and conversations about earlier drafts of this paper, we thank Michael Bratman, Eugene Chislenko, Nick French, Quinn Gibson, Zac Irving, Eddy Nahmias, Reuben Stern, R. Jay Wallace, our audience at the 2014 Northwestern University Society for the Theory of Ethics and Politics (NUSTEP), and our commentator at that conference, Deborah Goldgaber.
References