THE JOURNAL OF PHILOSOPHY

VOLUME CXVII, NO. 2, FEBRUARY 2020

STATISTICAL NORMALIZATION METHODS IN INTERPERSONAL AND INTERTHEORETIC COMPARISONS*

For a long time, economists and philosophers have puzzled over the problem of interpersonal comparisons of utility.[1] As economists typically use the term, utility is a numerical representation of an individual's preference ordering. If those preferences satisfy the von Neumann-Morgenstern axioms, then her preferences may be represented by an interval scale measurable utility function. However, as such the unit of utility for each individual is arbitrary: from individuals' utility functions alone, there is therefore no meaning to the claim that the difference in utility between coffee and tea for Alice is twice as great as the difference in utility between beer and vodka for Bob, or for any claim that makes comparisons of differences in utility between two or more individuals.

Yet it seems that we very often *can* make comparisons of preference-strength between people. And if we wish to endorse an aggregative theory like utilitarianism or prioritarianism or egalitarianism, combined with a preference-satisfaction account of well-being, then we need to be able to make such comparisons.

More recently, a formally analogous problem has appeared in discussions of normative uncertainty. The most widely suggested method for taking normative uncertainty into account in our decision-

* We wish to thank Stuart Armstrong, Hilary Greaves, Stefan Riedener, Bastian Stern, Christian Tarsney, and Aron Vallinder for helpful discussions and comments. We are especially grateful to Max Daniel for painstakingly checking the proofs and suggesting several important improvements. This work has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement No. 669751), and from the Leverhulme Trust (RPG-2014-064).

[1] See Ken Binmore, "Interpersonal Comparison of Utility," in Don Ross and Harold Kincaid, eds., *The Oxford Handbook of Philosophy of Economics* (Oxford: Oxford University Press, 2009), pp. 540–59, for an overview.

0022-362X/20/1702/61–95

making is *maximize expected choice-worthiness* (MEC), where the "choice-worthiness" of an option represents the strength of reasons one has to choose an option, according to an individual normative theory. This view has been defended by Ted Lockhart, William MacAskill, Toby Ord, Jacob Ross, Andrew Sepielli, and Ralph Wedgwood.[2]

However, MEC faces a serious problem. Maximizing expected choice-worthiness requires there to be a fact about how the difference in choice-worthiness between two options, according to one theory, compares with the difference in choice-worthiness between those two options, according to every other theory in which the decision-maker has credence. But how can this be done? For example, according to average utilitarianism, it is wrong to halve average well-being in order to quadruple population size; according to total utilitarianism, it is wrong not to do so. But is halving average well-being in order to quadruple population more wrong, according to average utilitarianism, than failing to do so is wrong according to total utilitarianism? And, in the absence of an obvious answer, how could we even begin to go about answering this question?[3] Several philosophers have suggested either that intertheoretic comparisons are never possible[4] or that they are *almost* never possible.[5] This has come to be known as the problem of intertheoretic comparisons.

[2] Ted Lockhart, *Moral Uncertainty and Its Consequences* (New York: Oxford University Press, 2000); Jacob Ross, "Rejecting Ethical Deflationism," *Ethics*, CXVI (July 2006): 742–68; Andrew Sepielli, "What to Do When You Don't Know What to Do," in Russ Shafer-Landau, ed., *Oxford Studies in Metaethics, Volume 4* (New York: Oxford University Press, 2009), pp. 5–28; Ralph Wedgwood, "Akrasia and Uncertainty," *Organon F*, XX, 4 (2013): 484–506; and William MacAskill and Toby Ord, "Why Maximize Expected Choice-Worthiness?," *Noûs* (forthcoming). It has also been referred to as "maximizing expected value," "maximizing expected rightness," and "minimizing expected wrongness," but it is clear that all authors are referring to approximately the same concept.

[3] For arguments that average and total utilitarianism must be incomparable, see John Broome, *Climate Matters: Ethics in a Warming World* (New York: W. W. Norton & Company, 2012), p. 185; and Brian Hedden, "Does MITE Make Right?," in Russ Shafer-Landau, ed., *Oxford Studies in Metaethics, Volume 11* (New York: Oxford University Press, 2016), pp. 102–28.

[4] James L. Hudson, "Subjectivization in Ethics," *American Philosophical Quarterly*, XXVI, 3 (July 1989): 221–29; Edward J. Gracely, "On the Noncomparability of Judgments Made by Different Ethical Theories," *Metaphilosophy*, XXVII, 3 (July 1996): 327–32; and Johan E. Gustafsson and Olle Torpman, "In Defence of My favorite Theory," *Pacific Philosophical Quarterly*, XCV, 2 (June 2014): 159–74.

[5] John Broome, "The Most Important Thing about Climate Change," in Jonathan Boston, Andrew Bradstock, and David Eng, eds., *Public Policy: Why Ethics Matters*, (Canberra: ANU E Press, 2010), pp. 101–16; and Broome, *Climate Matters*, *op. cit.*, p. 122.

In this article we introduce a class of potential solutions to both of these problems,[6] which we call *statistical normalization methods*. *Normalization methods* because they are ways of placing different utility functions or choice-worthiness functions on a common scale. *Statistical* because they normalize different utility functions or choice-worthiness functions based on their statistical properties, such as their range, or mean, or variance.

In this article we introduce some novel statistical normalization methods, and tentatively argue that one method, *variance normalization*, is superior to all other statistical normalization methods, including those that have been proposed in the literature.

Though we believe that the arguments we give in this article will be relevant to both interpersonal comparisons and to intertheoretic comparisons, for reasons of focus we will concentrate our attention on the application of statistical normalization methods to intertheoretic comparisons in the context of normative uncertainty.

The structure of this article is as follows. In section I, we introduce the framework within which we operate. In section II, we introduce the class of statistical normalization methods, including three novel accounts: variance normalization, mean absolute deviation from the median (MADAM) normalization, and mean absolute difference (MD) normalization. In section III, we consider specific examples and see how different statistical normalization methods fare, arguing that three accounts that have been proposed in the literature fail, whereas variance, MADAM, and MD do better. In section IV, we give the main argument of our paper: two approaches to formally specifying the "principle of equal say" and axiomatizations of various normalization methods given those specifications. On the first approach, we show that variance normalization, range normalization, and MADAM normalization can be axiomatized under reasonable assumptions. On the second approach, we show that variance normalization and MD normalization can be axiomatized under reasonable assumptions. Insofar as only variance normalization can be axiomatized using reasonable assumptions in both cases, and that in both

[6] More speculatively, it is possible that our account could also be used as a way to aggregate incommensurable goods. It is possible that some goods might be incommensurable in value, yet often we still need to make decisions even in the face of that incommensurability. Perhaps, for example, a government needs to decide whether to fund a number of economic development programs that would be positive in terms of improvements to people's well-being, but negative in terms of environmental impact. The accounts we suggest may be a way of normalizing the different value functions (such as welfarist value and environmental value) such that a decision can be made and each value is given equal consideration.

cases variance normalization results from the most natural choice of assumptions, we tentatively conclude that it is the best statistical normalization method.

I. OUR FRAMEWORK

We can now introduce the framework within which we work as follows. A *decision-situation* consists of a decision-maker, a time, and a set of options. A *decision-maker* is an actor who faces sets of options and must choose between them. An option is a proposition that the decision-maker has the power to make true at a time. The decision-maker has a credence function over propositions.[7] We note that the decision-maker's credence function also assigns credences to propositions concerning which choice situations she could face, and to which options she would choose in each choice situation.

Where we depart from standard decision theory is that we allow the decision-maker's credence function to roam over both empirical propositions and normative propositions, where normative propositions concern the choice-worthiness of options in decision situations.[8] Normative propositions can be specific normative claims, such as 'abortion is wrong' or 'eating meat is permissible', or claims about moral theories, such as 'utilitarianism is true'. In what follows, we will consider situations where the decision-maker has credence only in complete normative theories, which we define as propositions that give a choice-worthiness value for every possible option in every possible choice-situation.

We acknowledge that this is an unrealistic assumption for real-life decision-makers, who will typically think in terms of credences in specific normative claims rather than complete normative theories. However, in order to ultimately make sense of the choice-worthiness of

[7] Where a credence function is a function from propositions to real numbers in the interval [0, 1], which satisfies the Kolmogorov probability axioms, and is such that, for any set of disjoint and mutually exhaustive propositions, the sum of credences across those propositions equals 1. We assume precise credences for simplicity's sake in this article, but everything we say could be made to work if, as is plausible, the decision-maker's credences are imprecise. One might worry that, if non-cognitivism is true, then one cannot make sense of credences over normative theories. This has been a matter of some debate (see Michael Smith, "Evaluation, Uncertainty and Motivation," *Ethical Theory and Moral Practice*, v, 3 (September 2002): 305–20; Krister Bykvist and Jonas Olson, "Expressivism and Moral Certitude," *The Philosophical Quarterly*, lix, 235 (April 2009): 202–15; Andrew Sepielli, "Normative Uncertainty for Non-Cognitivists," *Philosophical Studies*, clx, 2 (September 2012): 191–207; and Krister Bykvist and Jonas Olson, "Against the Being For Account of Normative Certitude," *Journal of Ethics and Social Philosophy*, vi, 2 (July 2012): 1–8), so we shall sidestep this issue by assuming that cognitivism is true.

[8] We thank an anonymous referee for this journal for helping us to clarify this.

an action, conditional on some specific normative claim, we need to be able to make sense of the choice-worthiness of that option on all ways in which that specific normative claim could be true; that is, we need to be able to make sense of the average choice-worthiness of those complete theories on which the specific normative claim is true. So we regard the question of how choice-worthiness compares across complete theories as the bedrock question to address.

By "choice-worthiness" we mean the net strength of reasons that a decision-maker has in favor of choosing an option. Though "choice-worthiness" is a term of art, it clearly refers to a genuine concept: for example, it makes sense to say that, though one has reason against lying and reason against murdering, one has much stronger reason against murdering.[9] That is, if one can choose between a permissible option, lying, and murdering, the difference in choice-worthiness between a permissible option and murdering is much larger than that between the permissible option and lying. Though philosophers in this literature have sometimes chosen to focus on *moral* uncertainty, we understand choice-worthiness as representing the strength of reasons in favor of a certain option, *all things considered*. Prudential reasons, aesthetic reasons, and so forth are all taken into account in a normative theory's choice-worthiness ordering.[10]

We speak of *choice-worthiness functions*, which are numerical representations of a theory's choice-worthiness ordering. In this article we assume that choice-worthiness is cardinally measurable. That is, we assume that, whatever normative propositions are true, choice-worthiness can be represented using numbers such that statements like the following are meaningful: 'the difference in choice-worthiness between A and B is k times as great as the difference in choice-worthiness between C and D'.[11]

In this article we are interested in the idea that decision-makers should *maximize expected choice-worthiness* (MEC), where the expected choice-worthiness of an option A is the sum, across all theories T_i, of the decision-maker's credence in T_i multiplied by the choice-worthiness of A given T_i.

[9] This works even if you believe that all wrong acts are equally wrong, so long as you think that other people are at least coherent when they deny this.

[10] In what follows, we will assume that the decision-maker's credence in options is independent of their credence in theories. We thank an anonymous reviewer for this JOURNAL for noting that we need to make this assumption.

[11] Of course, real-life decision-makers will also have credence in the idea that choice-worthiness is merely ordinally measurable. For work exploring what to do in such a situation, see William MacAskill, "Normative Uncertainty as a Voting Problem," *Mind*, CXXV, 500 (October 2016): 967–1004.

MEC is sensitive to both the credences that the decision-maker has in different normative theories and the amounts of choice-worthiness at stake in the decision-situation according to the different theories. So in order to apply MEC, you need to be able to compare choice-worthiness across different first-order normative theories.[12] But it has been asserted by several philosophers that such intertheoretic comparisons are either always or nearly always impossible.[13] Those philosophers who believe intertheoretic comparisons to be always or nearly always impossible have taken this to be a strong reason to reject MEC.

II. STATISTICAL NORMALIZATION METHODS

A *normalization method* is an account of how to simultaneously choose choice-worthiness functions to represent different normative theories so that they all lie on a common scale—that is, just what is needed to apply MEC. In what follows, we restrict our attention to what we call *statistical* normalization methods: that is, those methods that place different normative theories on a common scale by treating some statistical property of a choice-worthiness function as being of the same magnitude across all normative theories. To our knowledge, only three statistical normalization methods have been proposed in the literature.

The first is known as the 'zero-one' rule in the literature on interpersonal utility comparisons: one normalizes different utility functions such that the maximal and minimal utility on all utility functions are the same.[14] In the normative uncertainty literature, the analogous rule was suggested by Ted Lockhart, who called it the Principle of Equity among Moral Theories (PEMT):

> The maximum degrees of moral rightness of all possible actions in a situation according to competing moral theories should be considered

[12] Technically, the thing that needs to be compared between theories is the difference in choice-worthiness between options, rather than absolute levels of choice-worthiness.

[13] Hudson, "Subjectivization in Ethics," *op. cit.*; Gracely, "On the Noncomparability of Judgments Made by Different Ethical Theories," *op. cit.*; and Gustafsson and Torpman, "In Defence of My Favorite Theory," *op. cit.* Broome, *Climate Matters*, *op. cit.*, p. 185, for example, says the following: "We then encounter the fundamental difficulty. Each different theory will value the change in population according to its own units of value, and those units may be incomparable with one another. . . . Most theories of value will be incomparable in this way. Expected value theory is therefore rarely able to help with uncertainty about value."

[14] Daniel M. Hausman, "The Impossibility of Interpersonal Utility Comparisons," *Mind*, CIV, 415 (July 1995): 473–90, argues that the zero-one rule is the *only* way that one can make interpersonal comparisons of utility when utility is understood as a numerical representation of preference strength. In what follows, we show that this is clearly not the case.

> equal. The minimum degrees of moral rightness of possible actions in a situation according to competing theories should be considered equal unless all possible actions are equally right according to one of the theories (in which case all of the actions should be considered to be maximally right according to that theory).[15]

The other two proposals we know of have been made in the context of arguing against the zero-one rule or the PEMT. Here is Amartya Sen arguing against the zero-one rule:

> It may be argued that some systems, e.g., assigning in each person's scale the value 0 to the worst alternative and the value 1 to his best alternative are interpersonally "fair" but such an argument is dubious. First, there are other systems with comparable symmetry, e.g., the system we discussed earlier of assigning 0 to the worst alternative and the value 1 to the sum of utilities from all alternatives.[16]

As long as the number of options under consideration is the same across all utility functions, Sen's proposal is formally identical to normalizing all utility functions at the distance between the mean utility and the minimum utility.

Here is Andrew Sepielli arguing against the PEMT:

> Lockhart's proposal seems arbitrary. Why equalize the maximum and minimum value, rather than, say, the mean value and the maximum value?[17]

Noting that they each normalize at some statistical properties of a utility or choice-worthiness function, we can refer to these three proposals as *range*, *max-mean*, and *mean-min* normalization.

In addition to these, we propose three novel normalization methods. (We introduce these and not others because we will ultimately show that there are natural sets of axioms that are only satisfied by these theories.) First, *variance normalization*, which treats the variance of the choice-worthiness of options as the same across all theories.[18] Second, *mean absolute deviation around the median normalization* (*MADAM*), which treats the mean absolute difference in choice-worthiness between every option and the median option as being

[15] Lockhart, *Moral Uncertainty and Its Consequences*, *op. cit.*, p. 84.

[16] Amartya K. Sen, *Collective Choice and Social Welfare* (San Francisco: Holden-Day, 1970), p. 98.

[17] Andrew Sepielli, "Moral Uncertainty and the Principle of Equity among Moral Theories," *Philosophy and Phenomenological Research,* LXXXVI, 3 (May 2013): 580–89.

[18] Variance is one of the standard measures of the spread of a distribution. It is defined by the sum of the squared differences in choice-worthiness of each option from the mean choice-worthiness. Variance is closely related to the standard deviation (which is simply the square root of the variance) and an account of normalization based on standard deviations would give exactly the same results as using variance.

the same across all theories.[19] Third, *mean absolute difference normalization* (*MD*), which treats the average absolute difference between any two options as the same across all theories. We note that, for all of these statistical accounts, if a theory ranks all options as exactly equally choice-worthy, then as a special case the normalization method leaves the choice-worthiness function alone: the normalized choice-worthiness function is just equal to the original one. (To not treat it as a special case would involve dividing by zero; and since the theory is indifferent between all options, how it is normalized does not matter.)

In order to fully describe specific statistical accounts, we need to make three clarifications.

First is the distinction between *broad* and *narrow* statistical accounts.[20] Narrow accounts normalize different theories within each decision-situation. So, for example, narrow range normalization would, for any decision-situation, treat the difference between the maximally and minimally choice-worthy options as being the same for all theories. This is what Lockhart proposes above. In contrast, broad accounts normalize different theories across all decision-situations. On our preferred way of making this precise, one would normalize at the expected range (or variance, and so on) across all option sets that the decision-maker might face, where the probabilities that go into the expectation are the decision-maker's fundamental prior probabilities of facing different option sets.[21]

We think that there are two good reasons for preferring the broad formulation. First, as we show in Appendix C, any narrow statistical normalization method generates cyclical recommendations across decision-situations.[22] That is, it can recommend doing *A* rather than *B* when the choice is between *A* and *B*, then recommend doing *B* rather than *C* when the choice is between *B* and *C*, but then recommend doing *C* rather than *A* in a third decision-situation when the

[19] This is a less common measure of the spread of a distribution than variance, but is still used in some contexts.

[20] The terminology of "broad" and "narrow" for this distinction comes from Amartya Sen, *Choice, Welfare and Measurement* (Cambridge, MA: Harvard University Press, 1997), p. 186.

[21] An alternative way of making this precise would be to say that the account should normalize different theories over the set of all conceivable options (as suggested by Sepielli, "Moral Uncertainty and the Principle of Equity among Moral Theories," *op. cit.*). However, this suggestion runs into numerous problems, including that, on many normative views, the choice-worthiness that a theory assigns to an option may depend on what other options are available in the decision-situation.

[22] A similar objection is made by Sepielli, but against the PEMT specifically; he does not show that this poses a problem for all statistical normalization methods (*ibid.*).

choice is between *A* and *C*. We take this to be a reason to reject the narrow formulation, though not necessarily a decisive reason. Second, and even more importantly, the broad formulation allows theories to regard some decision-situations as higher stakes than others. It seems clear that some decision-situations are higher stakes for some theories than for others: the decision about whether to tell a lie or tell the truth might be low stakes for utilitarianism but of high stakes for Kantianism. Narrow accounts do not allow for this.

Broad accounts have the problem that they would be significantly harder to use in practice than the narrow formulation. On broad accounts, one will often have very little idea how the choice-worthiness of an option in a choice-situation compares across theories. One would have to know at least approximately where a particular option lies within the distribution of the choice-worthiness of all possible options; but it seems that very often one will not know this. In contrast, on the narrow formulation, one only needs to know where in the distribution of choice-worthiness of all options within a decision-situation a particular option lies. This, presumably, would be much easier to know. So a decision-maker would more often be able to actually use narrow methods, at least approximately, than broad methods.

However, we ultimately do not think that this worry gives us sufficient reason to prefer narrow to broad accounts. We consider our project to be giving a "criterion of rightness" concerning what is correct to do under normative uncertainty, rather than a decision-procedure (something that is meant to be useful in guiding agents). So whether or not the criterion we give is practically useful for decision-makers is not of the first importance: what rules decision-makers should try to follow under normative uncertainty is a separate further question.

The second clarification is with respect to measure. In order for the accounts we develop to be well-specified, we must invoke a measure over all possible options. Because options are propositions, and we have already assumed that the decision-maker has a credence function, one might be inclined to simply use the decision-maker's credence over options. However, this would have the counterintuitive result that the measure over options changes as the decision-maker makes subsequent decisions and learns more about the world, thereby changing how different theories are normalized against each other. Whether this is a problem will depend on one's view on the purpose of statistical accounts (see the third clarification below). But it seems

to us that at least if normalization is making claims about how theories actually compare then the normalization between theories should stay the same across all decision-situations.[23]

As an alternative to using the decision-maker's credences over options, we propose we use the decision-maker's fundamental prior credence distribution over options. This will not change over time. If one is a subjective Bayesian, then it is simply a brute psychological fact what this measure is; if one is an objective Bayesian, then there are facts that constrain or specify what this measure should be. We do not wish, however, to get into this debate here: all we note is that the decision-maker will have *some* fundamental prior credence distribution over options, and this is sufficient for us to be able to define notions such as the variance of a choice-worthiness function.[24]

The third clarification is with respect to three distinct possible aims of statistical accounts. Statistical accounts could be understood as (i) making claims about how different theories actually compare; (ii) making claims about how different theories ought to be normalized for the purposes of maximizing expected choice-worthiness under moral uncertainty (even though their true normalization might be

[23] One might wonder whether this problem is really so bad: if the correct way to normalize theories will vary from decision-maker to decision-maker (because decision-makers have different priors), then why should we be concerned that the correct way to normalize theories will vary from decision-situation to decision-situation? We acknowledge that this could be a motivation for using the decision-maker's posterior credence function rather than their fundamental prior. However, we believe that the decision-maker would face problems of dynamic choice: situations where, for example, they should choose option A even though they know that, were they to do so, they would wish that they had chosen option B (because the measure and therefore the normalization has changed). These seem like additional significant problems for the posterior-credence version of our account, which make us disinclined to endorse it. However, if one wanted to use the posterior-credence version of our account, our arguments in the rest of the paper would go through *mutatis mutandis*. We thank an anonymous reviewer for this JOURNAL for raising this issue.

[24] One might have the following worry: in order for variance normalization to be applicable to the case of interpersonal comparisons, it would require that all people have the same fundamental prior. But unless we assume a strong form of objective Bayesianism, which we do not want to do, then different people will have different priors. Our response is that the issue that different people have different credence functions is a general problem for issues of interpersonal aggregation: if each person's preferences are coherent, and there is disagreement between people about the probabilities of different states of nature, then social preferences cannot be both coherent and Paretian (see John Broome, *Weighing Goods: Equality, Uncertainty and Time* (Oxford: Blackwell, 1991), p. 160). So it is already the case that, in order to engage in interpersonal aggregation, one need to assume a fixed credence function, rather than one that varies from person to person. The fact that we also have to rely on a fixed credence function in order to use variance normalization therefore creates no additional problem for us. We thank an anonymous reviewer for this JOURNAL for raising this issue.

different); or (iii) giving us a way to set different theories on a common scale even though they are not genuinely comparable.

To see the difference, consider the problem of interpersonal comparisons again. First, one could claim, for example, that different individuals' maximal and minimal utilities genuinely are the same. Second, one could claim that, even though some people's utility functions have a wider range than others, for reasons of fairness one should set all individuals' maximal and minimal utilities to be the same when aggregating different individuals' preferences. For example, perhaps one individual—a "utility monster"—has extremely strong preferences. One might think it unfair that such an individual could have such an enormous sway over the social ordering, and therefore wish to aggregate only a suitably dampened down normalization of that individual's utility function. Similarly, it might be the case that there is some fact of the matter about how different normative theories compare that is not given by statistical normalization methods. Perhaps, for example, all theories agree on the difference in choice-worthiness between two options in some specific uncontroversial decision. Even if so, statistical normalization methods may still be useful: we may conclude that maximizing expected choice-worthiness with respect to the true intertheoretic choice-worthiness comparisons is not the right approach, perhaps because doing so would allow those moral theories on which many decisions are extremely high stakes to have too much sway. Instead, one might think that the right way to act under normative uncertainty is to maximize *variance-renormalized* expected choice-worthiness.

Third, one could claim that, even though there is no real fact of the matter about how individuals' utility functions compare, we can still use range normalization to put those preferences on a common scale for the purpose of coming to an equitable agreement between different individuals. This is one understanding, for example, of range voting.

Our aim in this article is to assess the comparative merits of different statistical normalization methods. It is not to argue in favor of statistical normalization methods over other approaches. We therefore do not need to take a stand on which (if any) of the above three purposes of the statistical account is correct (though this would of course be a valuable further project).

III. EQUAL SAY

With these clarifications on board, we can turn to the methodology of assessing different statistical normalization methods. Sen and Sepielli both argue that there is no reason to choose between any of these

normalization methods, and implicitly suggest that there is no non-arbitrary normalization method. However, we do not think that this is the case.

We can judge different statistical normalization methods by how well they capture what we will call the principle of equal say: the idea (vague for now) that the aim of a statistical normalization method is to ensure that if different normative theories have equal credence, then in some sense they should get equal influence over the decisions of the decision-maker.

The motivation for the principle of equal say is as follows. In developing an account of decision-making under normative uncertainty, we want to remain neutral on what the correct normative theory is: we do not want to bias the outcome of the decision-making in favor of some theories over others. We mean this in the sense that if we have very high confidence in one normative theory T_i, then, no matter what theory T_i is (whether it is utilitarian or contractualist or a form of virtue ethics), our theory of decision-making under normative uncertainty should not generally recommend actions which are very bad under that normative theory.

Let us look at two specific cases of how this could go awry. First, consider average and total utilitarianism, and suppose that the decision-maker gives much higher credence to average than to total utilitarianism. Suppose that, in order to take an expectation over those theories, we choose to treat them as agreeing on the choice-worthiness of options concerning worlds with only one person in them. If so, then for almost all practical decisions involving variable populations, the option with the highest expected choice-worthiness will be the option that total utilitarianism regards as most choice-worthy because, for almost all real-life decisions (which involve a world with billions of people), the stakes would be large for total utilitarianism, but tiny for average utilitarianism. So even though the decision-maker has much higher credence in average utilitarianism than in total utilitarianism, she still almost always ought to act in accordance with total utilitarianism. So it is plausible that, if we treat the theories in this way, we are being partisan toward total utilitarianism.

In contrast, if we chose to treat the two theories as agreeing on the choice-worthiness differences between options with worlds involving some extremely large number of people (say 10^{100}), then for almost all real-life decisions, the option with the highest expected choice-worthiness will be the same as the option that average utilitarianism regards as most choice-worthy, even if the decision-maker had much higher credence in total utilitarianism than in average utilitarianism. This is because we are representing average utilitarianism as claiming

that, for almost all decisions, the stakes are much higher than for total utilitarianism. In which case, it seems that we are being partisan to average utilitarianism. What we really want is to have a way of normalizing such that each theory gets equal influence.

For a second way in which we could fail to give theories equal say, suppose that the decision-maker has credence in two moral views: utilitarianism, and a near-absolutist view on which one ought not to tell a lie unless one can provide a benefit as great as saving 10^{100} lives,[25] and that the decision-maker is 99.9999% sure in utilitarianism, but has 0.0001% credence in the near-absolutist view. The "natural" way of normalizing these two views is to suppose that they agree on the value of saving lives, but that the near-absolutist view also supposes that there are additional extremely strong reasons not to lie. If so, then the expected choice-worthiness of lying in order to save lives (even if that means saving billions of lives) is almost never greater than the choice-worthiness of refraining from lying. This conclusion seems perverse: it seems that "fanatical" moral views like the near-absolutist view should not be able to so unduly influence what it is rational for a decision-maker to do.

Against the idea that we should give different theories equal say, one could argue that some theories are simply higher stakes in general than other theories. Considerations of fairness, one might argue, are relevant to issues about how to treat *people*: one can be unfair to a person, but one cannot be unfair to a theory. Perhaps by saying that one was being "unfair" to Kantianism, one could mean that one's degree of belief was too low in it. But one cannot be unfair to it insofar as it "loses out" in the calculation of what it is appropriate to do. If a theory considers a situation to have low stakes, we should presumably represent it as such.

It may be the case that "equal say" has no bearing on how theories actually compare (though the authors of this paper are in disagreement on this issue). But it seems clearly to have bearing on the other two ways of understanding the purpose of statistical accounts that we described in the previous section. In order to avoid "fanatical" conclusions, where the expected choice-worthiness of one's options is almost entirely determined by the choice-worthiness function of a theory in which one has vanishingly small credence but which

[25] We consider a near-absolutist view rather than an absolutist view because there are difficulties in understanding absolutist views in terms of cardinal choice-worthiness. For work on modeling absolutist views in decision-theoretic terms, see Mark Colyvan, Damian Cox, and Katie Steele, "Modelling the Moral Dimension of Decisions," *Noûs*, XLIV, 3 (September 2010): 503–29.

claims that most decision-situations are enormously high stakes, one might wish instead to renormalize the moral views in which one has credence. Or, if one concludes that there is no ultimate fact of the matter about how to make choice-worthiness comparisons across two different theories, then one might conclude that, if we are to make any rational choices at all, we need some principled way of placing those theories on a common scale, and statistical normalization methods are the best account we have. For either of these two approaches, we think that "equal say" is a promising way of adjudicating between different statistical normalization methods.

In the rest of the article, we will use the principle of equal say to assess different statistical normalization methods. We will first do so informally, then develop two formal arguments in later sections.

IV. APPEAL TO CASES

To develop an intuitive sense of how different statistical normalization methods can differ in how they apportion "say" between theories, and why some accounts seem clearly inferior to others, we shall consider some examples.

To help see the implications of different normalization methods, we shall represent normative theories visually, where horizontal lines represent different options and are connected by a vertical line, representing the choice-worthiness function. The higher on the page the option, the more choice-worthy the option, and the greater the distance between two horizontal lines, the greater the difference in choice-worthiness between those two options. These diagrams are approximately to scale.

First, let us consider the normalization methods mean-min (as suggested by Sen) and max-mean (as suggested by Sepielli). We will consider how they normalize two types of normative theories. The first are *Top-Heavy* theories, according to which there are a small number of outliers in choice-worthiness, but they are only in one direction: there are just a small number of extremely un-choice-worthy possible options. Any consequentialist theory that has a low upper bound on value, but a very low lower bound on value, such that most options are close to the upper bound and far away from the lower bound, would count as a Top-Heavy moral theory. The second are *Bottom-Heavy* theories, which are the inverse of Top-Heavy theories.

Because Top-Heavy and Bottom-Heavy theories are simply inversions of each other, it seems very plausible, if we are to give theories equal say, that one should treat the magnitudes of choice-worthiness differences as the same according to both theories, just of opposite sign. But this is not what we find for Sen and Sepielli's suggestions.

First let us consider max-mean. Figure 1 represents the theories after normalizing.

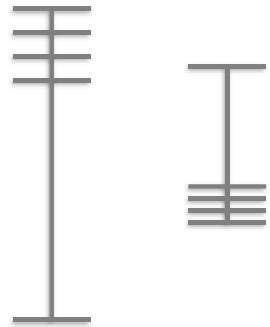

Figure 1. Top-Heavy and Bottom-Heavy, after normalizing by max-mean.

Max-mean favors Top-Heavy theories and punishes Bottom-Heavy theories. But these two theories are just inversions of each other, so presumably ought to be treated symmetrically. Absent any case that unlikely but extremely good outcomes should be treated differently than unlikely but extremely bad outcomes (and we do not see such a case), it appears that max-mean does not deal even-handedly between these two classes of theories.

When we consider mean-min, we get exactly the same problem, except that mean-min favors Bottom-Heavy over Top-Heavy (see Figure 2).

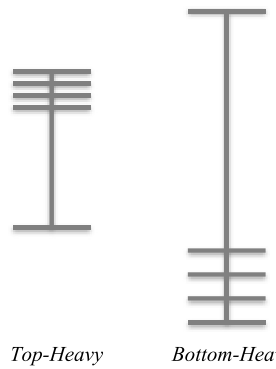

Figure 2. Top-Heavy and Bottom-Heavy, after normalizing by mean-min.

These examples therefore give us grounds for rejecting both max-mean and mean-min.

Next let us consider range normalization. To see the problems with this account, let us consider two new classes of theory. The first class is of *Bipolar* theories, which are theories where the choice-worthiness clusters around two different levels, such that the differences in choice-worthiness when comparing two highly choice-worthy options or two highly un-choice-worthy options are zero or tiny compared to the difference in choice-worthiness when comparing a highly choice-worthy option and a highly un-choice-worthy option. For example, a view according to which violating rights is impermissible, everything else is permissible, and where there is very little difference in choice-worthiness between different impermissible options and different permissible options, would be a Bipolar theory.

We will call the second type of theory *Outlier* theories. According to these theories, most options are roughly similar in choice-worthiness, but there are some options that are extremely choice-worthy, and some options that are extremely un-choice-worthy. A bounded consequentialist theory with very high and very low bounds on value might be like this: the differences in value between most options are about the same, but there are some possible worlds which, though unlikely, are very good indeed, and some other worlds which, though unlikely, are very bad indeed.

If we used range normalization, the normalized versions of examples from the four classes of theory would look as in Figure 3.

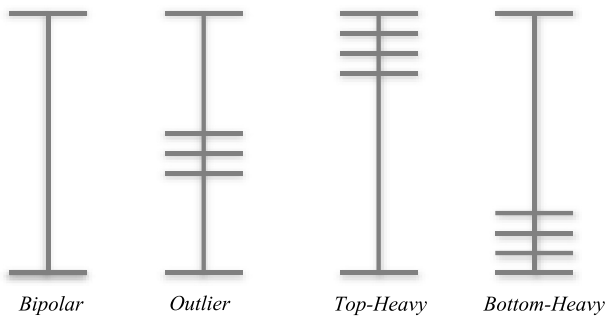

Figure 3. Bipolar, Outlier, Top-Heavy, and Bottom-Heavy, after normalizing by range.

For Top-Heavy and Bottom-Heavy, range normalization yields the intuitively right result. Top-Heavy and Bottom-Heavy are simply inversions of each other, so it seems very plausible that one should treat the magnitudes of choice-worthiness differences as the same according to both theories, just of opposite sign.

For Bipolar and Outlier, however, range normalization does not yield the right result. Because its scaling *only* cares about the maximal and minimal values of choice-worthiness, it is insensitive to how choice-worthiness is distributed among options that are not maximally or minimally choice-worthy. As we will now show, this means that Bipolar theories have much more power, relative to Outlier theories, than they should.

Let us consider a concrete case. Suppose that Sophie is uncertain between an absolutist moral theory and a form of utilitarianism that has an upper limit of value of ten billion happy lives, and a lower limit of ten billion lives of agony. She has 1% credence in the absolutist theory, and 99% credence in bounded utilitarianism. If range normalization is correct, then in almost every decision-situation she faces she ought to side with the absolutist theory. Let us suppose she is confronted with a murderer at her door, and she could lie in order to save her family: an action required by utilitarianism, but absolutely wrong according to the absolutist view. Given range normalization, it is as wrong to lie, according to the absolutist view, as it is to force ten billion people to live lives of agony, according to utilitarianism. So her 1% credence in the absolutist view means that she should not lie to the murderer at the door. In fact, she should not lie even if her credence in the absolutist theory was as low as 0.000001%. That seems incredible. Range normalization flagrantly fails to respect the principle of equal say in cases where some theories put almost all options into just two categories.[26] So this example gives us grounds to reject range normalization.

What, though, of variance normalization, MADAM, and MD? If we treat the variance of choice-worthiness as the same across all four theories, they would be represented as in Figure 4.

If we treat the mean absolute difference as the same across all theories, they would be represented approximately as in Figure 5.

If we treat the mean absolute distance from the median as the same across all four theories, they would be represented approximately as in Figure 6.

Variance, MADAM, and MD normalizations all do better than max-mean and max-min insofar as they normalize Top-Heavy and Bottom-Heavy in the same way. They also do better than range normalization insofar as they make Bipolar's range comparatively smaller than Outlier's range, which is the result we wanted. So the consideration of particular cases seems to motivate variance normalization, MADAM, and MD over their rivals.

[26] We thank Bastian Stern for initially suggesting this argument.

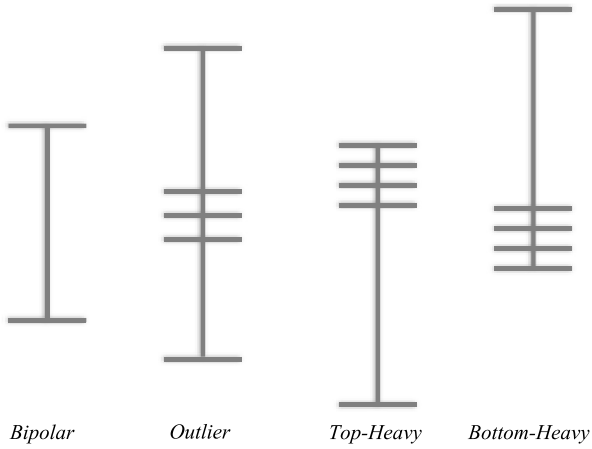

Figure 4. Bipolar, Outlier, Top-Heavy, and Bottom-Heavy, after normalizing by variance.

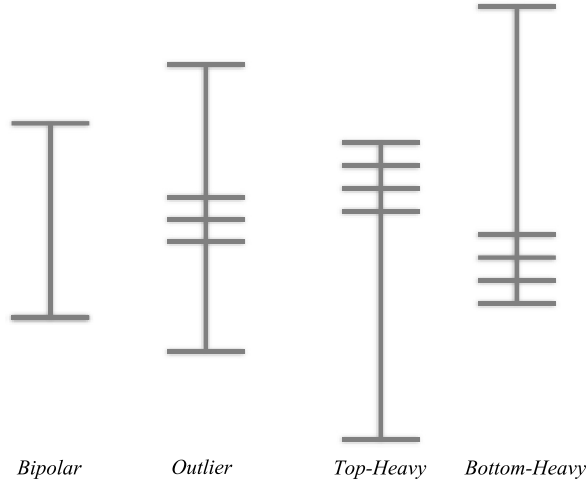

Figure 5. Bipolar, Outlier, Top-Heavy, and Bottom-Heavy, after normalizing by MD.

It is harder, however, to have clear intuitions about how to compare variance, MADAM, and MD with respect to these examples. In comparison to variance or MD normalizations, MADAM gives comparatively less weight to Bipolar than to the other three theories; it also gives slightly more weight to Outlier than to Top-Heavy and Bottom-Heavy. MD and variance give very similar results, though in comparison to variance normalization, MD gives slightly more weight to Top-

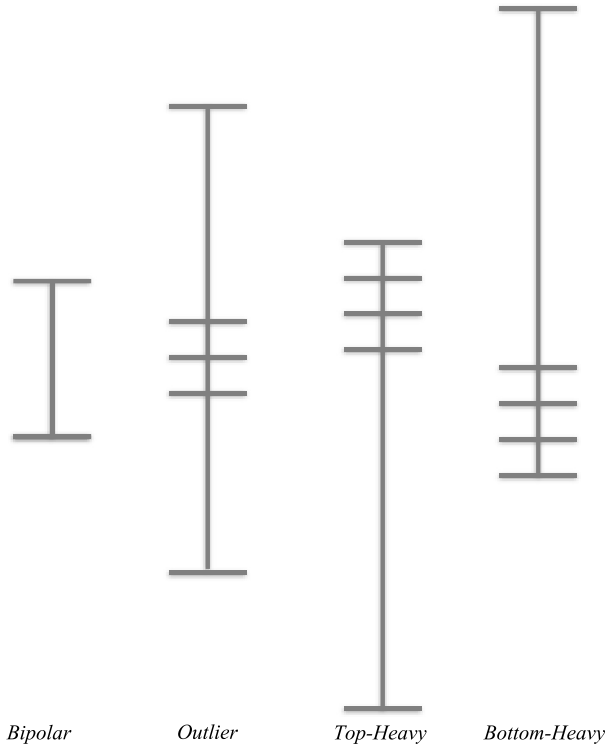

Figure 6. Bipolar, Outlier, Top-Heavy, and Bottom-Heavy, after normalizing by MADAM.

Heavy and Bottom-Heavy theories compared to Outlier theories. In our view, it is just not clear, intuitively, which are the correct results.

What is more, by its nature an appeal to cases argument can be suggestive, but can hardly constitute a knockdown argument. Perhaps there are other normalization methods that do as well as variance, MADAM, and MD do on the cases above. Perhaps there are other cases in which variance or MADAM or MD do worse than the other methods we have mentioned. It would be nice to rely on more rigorous arguments.

In the next section we shall suggest two approaches to making the idea of equal say formally precise, though in each case with some leeway in how one exactly specifies the notion. On the first approach, we are led to conclude that either variance or range or MADAM is the normalization method that best captures the principle of equal say; on the second approach, we are led to conclude that either variance or MD is best. Let us now turn to these arguments.

V. EQUAL SAY AS EQUAL DISTANCE FROM THE UNIFORM THEORY

Consider a uniform choice-worthiness function—one that assigns the same degree of choice-worthiness to all options. If any theory's choice-worthiness function were normalized to be essentially uniform before applying MEC,[27] then that theory would not affect the final decision. Such a normalization would give that theory no "say." We could measure how much "say" a theory has by how far away its normalized choice-worthiness function is from the uniform choice-worthiness function. Remember that by "say" we are thinking of the degree to which the theory may influence the choice between options, for a fixed degree of credence in that theory.

Imagine starting each theory off with a uniform choice-worthiness function and an equal amount of credit, where this credit can be spent on moving the choice-worthiness function away from the uniform function. Every move away from the uniform choice-worthiness assignment increases the "say" of that theory, and uses up a proportionate amount of credit. On this account, giving every theory equal say means giving them an equal amount of starting credit. In this section we will spell out this suggestion, explain the motivation for it, and demonstrate that (for our normal notion of distance) variance normalization is the only normalization method that gives every theory equal say so understood.

To illustrate, let us begin by considering different theories that *are* intertheoretically comparable—they have already been normalized in some way, so there is a shared unit of choice-worthiness across them. We will say that a completely uniform theory, according to which all options are equally choice-worthy, gives all options choice-worthiness 0 (though we could have just as well have said it gives all options 17, or any other number). Next, consider a theory, T_1, which differs from the uniform theory only insofar as its choice-worthiness function gives one option, A, a different choice-worthiness, x. There are two ways in which a theory T_2 might have more say than T_1. First, it could have the same choice-worthiness ordering as T_1, but its choice-worthiness function could give A a higher numerical value (remembering that, because we are talking about theories that are intertheoretically comparable, this is a meaningful difference between these two theories). If it gave A a numerical value of $2x$, so that the choice-worthiness difference between A and any other option is twice as great according to

[27] If a theory is represented by a choice-worthiness function f, it is also represented by $0.1f$, $0.01f$, $0.001f$, and so on. These limit to a uniform choice-worthiness function, and if we go far enough down the sequence then the representative will be close enough to uniform as to make no difference.

T_2 than according to T_1, then T_2 would have twice as much "say" as T_1. A second way in which a theory could have more "say" than T_1 is if it assigned non-zero numerical values to another option in addition to A. Then it would have equal say with respect to A, but would have a greater say with respect to the other options.

But what does "moving away" from the uniform theory mean? We can take this idea beyond metaphor by thinking of choice-worthiness functions geometrically. To see this, suppose (to begin with) that there are only two possible options, A and B, and three theories, T_1, T_2, and T_3, whose choice-worthiness functions are represented by Table 1.

	T_1	T_2	T_3
A	−4	3	4
B	1	4	1

Table 1.

Using the choice-worthiness of A as the x-axis and the choice-worthiness of B as the y-axis, we may represent this geometrically as in Figure 7.

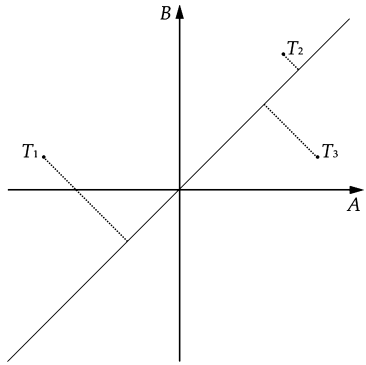

Figure 7.

Any point on this graph represents some choice-worthiness function and those corresponding to T_1, T_2, and T_3 are marked. The diagonal line represents all the uniform choice-worthiness functions. The dotted lines show the distance from each of T_1, T_2, and T_3 to their nearest uniform choice-worthiness function. These distances allow a way of precisely defining "equal say." Giving each theory equal say means choosing a (normalized) choice-worthiness function for each theory such that, for every choice-worthiness function, the distance

from that choice-worthiness function to the nearest uniform choice-worthiness function is the same.

It turns out that the distance from a choice-worthiness function to the nearest uniform function is always equal to the standard deviation of the distribution of choice-worthiness values it assigns to the available options (see Appendix A for a proof). So treating all choice-worthiness functions as having equal say means treating them as lying at the same distance from the uniform function, which means treating them such that they have the same standard deviation and thus the same variance. Variance normalization is thus the unique normalization method for preserving equal say on this understanding of equal say.

We can now look at the geometric interpretation of normalizing theories by their variance (see Figure 8).

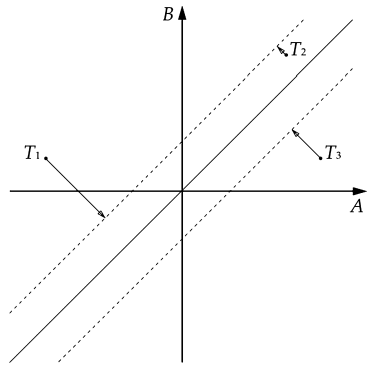

Figure 8.

The dashed lines in this diagram represent all the choice-worthiness functions that are a distance of 1 from the nearest uniform function.[28] This means that they also have a standard deviation of 1 and hence a variance of 1. In order to normalize each theory so that they have the same amount of "say," we move each theory to the closest point on one of the dashed lines (the arrows show these moves). This corresponds to linearly rescaling all of the theory's choice-worthiness values so that their variance is equal to 1, while keeping their means unchanged. This does not change the ordering of the options by that theory's lights; it just compresses it or stretches it so that it has the same variance as the others. One can then apply MEC to these normalized choice-worthiness functions.

[28] We could have chosen any non-zero value here; 1 is merely convenient.

This all works in the same way for any finite number of options.[29] A choice-worthiness function gives an assignment of a real number to each option, so if there are n options a choice-worthiness function can be represented as a collection of n real numbers. Just as pairs of real numbers give us Cartesian coordinates in the plane, and triples give us coordinates in three-dimensional space, so we can interpret this collection as the coordinates of a point in n-dimensional Euclidean space. We can then proceed the same way, looking at the distance in this n-dimensional space from a choice-worthiness function to the nearest uniform theory, equating this to "say," and normalizing to make the distances the same. Just as before, the distance corresponds to the standard deviation, and so normalizing to equalize variance is the unique way to provide equal say.[30]

While there is no need to normalize the means of the choice-worthiness functions (it does not affect the MEC calculation, as we are ultimately interested in comparing between options) it could be convenient to normalize them all to zero, by adding or subtracting a constant from each choice-worthiness function. If so, then the choice-worthiness functions are in the familiar form of "standard scores" or "z-scores" where the mean is zero and the unit is one standard deviation. These z-scores are commonly used in statistics as a way to compare quantities that are not directly comparable, so it is particularly interesting that our approach to intertheoretic choice-worthiness comparisons for non-comparable theories could be summerized as "compare them via their z-scores."

This argument made an implicit assumption that the appropriate way to measure distance between utility functions is the Euclidean distance (that is, the l^2 metric).[31] What happens if we instead use one of the other natural conceptions of distance, such as the l^1 or l^∞ metric? Under a l^1 metric, sizes are equal when theories are normalized so that the average distance in choice-worthiness between a random

[29] This argument applies only in the case where, in any decision-situation, there are finitely many options, and makes an assumption of symmetry in the weight we attach to each. This is the simplest case for intertheoretic value comparisons, and any method should at least behave well in this base case. Note, however, that our argument applies to all finite decision-situations, and so does *not* prevent us from considering what we have called "broad" normalization methods across infinitely many decision-situations—what is more, the aggregated set of options across all decision-situations may be infinite, as long as *within* each single decision situation there are only finitely many options to choose from.

[30] Again, see Appendix A for details.

[31] We thank an anonymous referee for this JOURNAL for pressing this issue.

option (according to the measure over options) and the median option is equalized. That is, the above argument using the l^1 metric supports the MADAM normalization. Under a l^∞ metric one sets as equal the distance between the best and worst option. That is, the above argument using the l^∞ metric supports range normalization. We show both of these results in Appendix A. (In both cases the broad formulation would equalize the expectations of the respective quantities across choice-situations.)

Euclidean distance is not obviously correct for this space, and we do not think that this argument alone is enough to conclude in favor of variance normalization. Rather, we think it suggests using a normalization method corresponding to some natural notion of distance. This rules out max-mean and mean-min. It does leave at least an infinite family of possibilities, based on the l^n distance norms. Among these, the three most naturally distinguished points are the three we have just mentioned: l^1 and l^∞ are the ends of the spectrum, and the l^2 norm is unique among the whole spectrum in giving an isotropic space, meaning that the geometry is particularly well-behaved by treating all directions equally. Mathematical taste might lead you to prefer one of these over the others, but it is at least unclear. Still, it narrows the space.

In the next section we shall look at a different style of argument that, depending on the precise assumptions used, motivates either variance normalization or MAD.

VI. EQUAL SAY AS EQUAL EXPECTED CHOICE-WORTHINESS OF VOTING

The previous argument cashed out the idea of "equal say" as "equal distance from a uniform choice-worthiness function." In our second argument, we shall borrow a concept from voting theory: *voting power*. An individual's voting power is the *a priori* likelihood of her vote being decisive in an election, given the assumption that all the possible ways for other people to vote are equally likely. It is normally used for elections with just two candidates, but the concept is perfectly general.

We could extend this concept to flesh out "equal say." A first challenge is that while voters all have just one vote, theories come with different credences. We want theories with the same credence to have the same voting power and for voting power to go up on average as the credence increases.[32] If we knew that all credences in theories

[32] The qualification "on average" is needed as it is possible for a theory to get its way all the time when it is given a credence that is slightly less than 1 and from that point increases in credence will not improve its power. This is analogous to how a voting block might have all the power with less than 100% of the votes.

were multiples of 10%, we could regard this as an electorate with 10 people, and ask that the voting power of each was the same. If we knew that the credences were all multiples of 1% we could similarly treat this as an electorate of 100 people. But these are special cases. For the general case, we propose instead to imagine asking that every individual's voting power is equal in the limit as we take increasingly large electorates. And in the limit with larger electorates we could perfectly approximate the split of credences between the different theories. Hence we will look at the voting power of the first small (and equally sized) amount of credence in each particular theory, and ask what would make those the same.

In Appendix B, we provide a proof that the only normalization aggregation method which gives equal voting power to all non-uniform theories in this sense is MD.

However, a second challenge is that by a theory's own lights it does not just matter that one's credence in it is decisive in determining which option gets chosen, it matters how much better this chosen option is than the option that would have been chosen otherwise. Getting its way in a decision about whether to prick someone with a pin matters a lot less, for utilitarianism, than getting its way in a decision about whether to let a million people die. If we are normalizing to provide "equal say," we should arguably take that into account as well. Since theories come with a measure of this difference between the options (the choice-worthiness difference), and they use its expectation when considering descriptive uncertainty, it is natural to use this here. This means we should speak not just of the likelihood of being decisive, but of the increase in expected choice-worthiness. This is not done in normal analysis of voting power since it is usually assumed that there is no access to information about how strong the preferences of the voters are. In our context, however, we have assumed that we do know strengths of choice-worthiness on different theories. We thus achieve "equal say" when from a position of complete uncertainty about how our credence will be divided over different choice-worthiness functions, an increase in our credence in a theory by a tiny amount will increase the expected choice-worthiness of the decision for that theory by the same degree regardless of which theory it was whose credence was increased.

There is one final challenge for this approach. If each theory had one canonical choice-worthiness function, this definition would work. But since each theory is described by infinitely many different choice-worthiness functions (positive affine transformations of each other), we do not yet know which choice-worthiness function to use to represent each theory and so cannot come up with a unique value for the "expected choice-worthiness."

However we can resolve this by considering that the normalization used to choose an option in a decision-situation should be the same normalization used to measure equal say in terms of this version of voting power. This does not sound like a strong constraint, but it is enough to let us prove that there is a unique normalization method that satisfies it and equalizes voting power. In Appendix B, we prove that variance normalization is the only normalization method that can coherently satisfy this interpretation of equal say as equal impact-adjusted voting power.

This second interpretation of voting power has given us a different preferred normalization method. We think that it is more natural to think of voting power in terms of both likelihood of success and value of success if success is achieved, and hence we prefer the set-up of the argument that gives variance normalization.

VII. CONCLUSION

In this article we considered methods of normalizing different choice-worthiness functions or utility functions with reference to statistical properties of those functions. We first argued, by appeal to cases, that those statistical normalization methods that have been proposed in the literature—range, max-mean, and mean-min—are unsatisfactory because they fail to give different theories equal say; in contrast, variance, MADAM, and MD seem to do well in the cases we considered. We then showed that, if we understand "equal say" as distance from the uniform theory, then—depending on one's choice of metric—range normalization, MADAM, or variance give all theories equal say. Finally, we showed that, if we understand equal say as equal voting power, then, depending on how one understands voting power, either variance normalization or MD gives all theories equal say.

We note that, for both of the arguments we considered, variance normalization is what results from what we think is the most natural formulation of the argument, and that variance normalization is the only account that has support from both forms of argument, rather than just from one.

Given that two distinct lines of argument, in their most plausible form, motivate variance normalization, we conclude that variance normalization is the uniquely best statistical normalization method.

APPENDIX A. EQUAL SAY AS EQUAL DISTANCE FROM THE UNIFORM THEORY

In this section, we will often consider the space of all choice-worthiness functions for a fixed set $\mathcal{D}$ of finitely many options, that is, the set $U = \{f : \mathcal{D} \to \mathbb{R}\}$. Note that by choosing an arbitrary numbering of

the finitely many options in $\mathcal{D}$ we can identify this set with $\mathbb{R}^n$, where n is the number of options in $\mathcal{D}$.[33] We will freely make use of this identification in our proofs.

Thus the propositions in this section are of the following type: "For all finite $\mathcal{D}$, the following statement is true about the associated space U of choice-worthiness functions on $\mathcal{D}$, where we tacitly identify U with $\mathbb{R}^n$."

Proposition A.1. If there are finitely many options, then Euclidean distance from the uniform theory in the space of choice-worthiness functions is equal to the standard deviation of the choice-worthiness function.

Proof. Let U denote the space of choice-worthiness functions on n options, equipped with Euclidean distance. The shortest path from any point in this space to the line of uniform theories runs perpendicular to that line. Given a point $p = (p_1, \ldots, p_n) \in U$, we may replace it by $p' = (p_1 - m, \ldots, p_n - m)$, where m is the mean of $p_1, \ldots, p_n$. Since p' is just a translation of p parallel to the line of uniform theories, it will lie at the same distance from that line. Moreover, since the mean satisfies $m = \frac{1}{n}\sum_{i=1}^{n} p_i$, we find that p' by construction lies in the plane through the origin $\{(x_1, \ldots, x_n) \mid \sum_i x_i = 0\}$ which is perpendicular to the line of uniform theories. The closest point to p' on this line thus is the origin $(0, \ldots, 0)$.

The Euclidean distance between these two points is $(\sum_i (p_i - m)^2)^{1/2}$, which is just the standard deviation $\sigma(p')$. Since the standard deviation is invariant under translation, $\sigma(p')$ equals $\sigma(p)$. □

Thus to normalize the Euclidean distance from the uniform theory one normalizes standard deviation, or, equivalently, the variance σ^2. (To see that normalizing the standard deviation σ is equivalent to normalizing the variance σ^2, note that both σ and σ^2 are non-negative real numbers, and that for such numbers x and y we have $x = y$ if and only if $x^2 = y^2$.)

Proposition A.2. If there are finitely many options, then the l^1-distance from the uniform theory is directly proportional to the mean absolute deviation around the median for its choice-worthiness function.

Proof. Let U denote the space of choice-worthiness functions on n options, equipped with l^1-distance. Given a point $p = (p_1, \ldots, p_n) \in U$, and a point on the line of uniform theories $q = (a, \ldots, a)$, then the l^1-distance between p and q is $\sum_i |p_i - a|$. This is minimized when an

[33] Formally, any numbering is just a bijection $\phi\colon \{1, \ldots, n\} \xrightarrow{\sim} \mathcal{D}$. It is routine to see that this yields an isomorphism of vector spaces $\psi\colon U \xrightarrow{\sim}, \mathbb{R}^n, f \mapsto (f(\phi(i)))_{i=1,\ldots n}$, where U is a vector space via point-wise operations.

equal number of p_i lie above a as lie below it[34]—which holds when a is the median of the p_i (and may also hold for some other values if the median does not coincide with one of the p_i—in this case there is not a unique closest point). The distance to this point is equal to the sum of the absolute deviation around the median. Since there are a fixed n options, this sum is directly proportional to the mean absolute distance around the median. □

Thus to normalize the l^1-distance from the uniform theory one normalizes MADAM (mean absolute deviation around median). (This follows immediately from $x = y$ if and only if $nx = ny$ for any real numbers x, y, and $n \neq 0$.)

Proposition A.3. If there are finitely many options, then the l^∞-distance from the uniform theory is directly proportional to the range of its choice-worthiness function.

Proof. Let U denote the space of choice-worthiness functions on n options, equipped with l^∞-distance. Given a point $p = (p_1, \ldots, p_n) \in U$, and a point on the line of uniform theories $q = (a, \ldots, a)$, then the l^∞-distance between p and q is $\max_i |p_i - a|$. This is minimized when a is half-way between the largest and smallest p_i. In that case the distance is equal to half the range of the p_i. This is directly proportional to the range. □

Thus to normalize the l^∞-distance from the uniform theory one normalizes range.

APPENDIX B. EQUAL SAY AS EQUAL EXPECTED CHOICE-WORTHINESS OF VOTING

We will work in the following setting.

- $\mathcal{O}$ will denote a countable set containing all *options* between which we might choose in any decision-situation. It will be treated as fixed throughout.
- $\mathcal{D} = \{\text{finite subsets of } \mathcal{O}\}$, which we will interpret as the set of admissible *decision-situations*. We interpret a member $D \in \mathcal{D}$ as representing the decision between precisely the options A with $A \in D \subset \mathcal{O}$. For simplicity, we only consider decisions between finitely many options.

[34] To see the claim about what values of a minimize $\sum_i |p_i - a|$, assume without loss of generality that $p_1 \leq \cdots \leq p_n$ and consider the continuous map $f\colon \mathbb{R} \to \mathbb{R}, a \mapsto \sum_i |p_i - a|$. Clearly, we have $\max_{p_1 \leq a \leq p_n} f(a) < \min_{a \notin [p_1, p_n]} f(a)$. A minimum of f in the interval $[p_1, p_n]$ thus is a global minimum of f. Since the image of a compact set under a continuous map is compact, the image of $[p_1, p_n]$ under f is a compact subset of $\mathbb{R}$, which thus has a minimal element. It therefore suffices to show that for any $a \in [p_1, p_n]$ such that there are not as many p_i below as above it the value $f(a)$ is not minimal, which is routine.

- A (moral) *theory* $\mathbf{T}$ is an equivalence class of families of *choice-worthiness functions* $T = (T_D)_{D \in \mathcal{D}}$, where $T_D \colon D \to \mathbb{R}$. Two families T and T' represent the same theory $\mathbf{T}$ if and only if there is a constant $k > 0$ such that $\forall D \in \mathcal{D} : T'_D = kT_D$. For a fixed decision situation D we will often identify the space of all choice-worthiness functions $T_D \colon D \to \mathbb{R}$ with $\mathbb{R}^{\#D}$, where $\#D$ denotes the number of options in D.
- A (statistical) *normalization method* N is a selection, for each theory $\mathbf{T}$, of one family of choice-worthiness functions $T = N(\mathbf{T})$ representing that theory. We interpret the functions in $N(\mathbf{T})$ for varying theories $\mathbf{T}$ as having values on a common scale, such that it makes sense to aggregate them by maximizing expected choice-worthiness (MEC).
- P will denote a prior over decision-situations, that is a probability measure on $\mathcal{D}$ (where we use the full power set as σ-algebra, which works as $\mathcal{D}$ is countable).
- Q will denote a prior over normalized (!) theories, that is a probability measure on the space $\mathcal{T} = \prod_{D \in \mathcal{D}} \mathbb{R}^{\#D}$ of families of choice-worthiness functions. Since $\mathcal{T}$ is a product of uncountably infinite sets this requires some elaboration. For each D, we equip $\mathbb{R}^{\#D}$ with the familiar σ-algebra of Lebesgue-measurable sets. We then get a σ-algebra on $\mathcal{T}$ by taking the product σ-algebra.

Remark B.1. We suspect that our results and proofs essentially remain valid for uncountable option sets $\mathcal{O}$. However, this would require modifications to our exposition. For instance, the set of decision situations $\mathcal{D}$ would then be uncountable as well, and so the choice of σ-algebra would matter for which priors P are feasible. We would also need to replace some sums with integrals.

We are interested in precise notions of

(i) the expected chance of voting being pivotal to the outcome, and
(ii) the expected choice-worthiness of voting (the choice-worthiness as regarded by the theory, per unit of credence held in that theory),

but as mentioned in the main text this is dependent on the credences held. To remove this dependence we look at the expected effect of adding a very small level of credence in the theory, using the machinery of derivatives to take the limit as the extra credence allocated goes to zero.

First, consider a fixed decision-situation D. Our prior Q determines a probability measure on the space $\mathbb{R}^{\#D}$ of choice-worthiness functions for this decision-situation. The expected value relative to that probability measure is a choice-worthiness function S_D that represents just how we would decide in this situation when using MEC together with the normalization method and credences implicit in Q.

Now consider if we were to make a small change p in credence for some family of normalized choice-worthiness functions T, adjusting credence in all other families uniformly. We would then replace S_D with $pT_D + (1-p)S_D$. We are interested in:

i. The probability (with respect to Q) that this replacement changes the set of options with maximal choice-worthiness. Call this $P_{T,D}(p)$. Regarding it as a function of p, we calculate the derivative at 0: call this $g_D(T)$.
ii. The expected improvement in choice-worthiness of the chosen option, according to T. Call this $E_{T,D}(p)$. We again look at the derivative at $p = 0$: call this $f_D(T)$. It is expressed in units of choice-worthiness according to T.

The fairness conditions we are interested in will be expressed in terms of the expected value of these derivatives across all decision-situations, assuming the latter are distributed according to P:

(1) $g(T) = \mathbb{E}_{D\sim P}[g_D(T)]$
(2) $f(T) = \mathbb{E}_{D\sim P}[f_D(T)]$

Remark B.2. Without assumptions on the priors P and Q, neither the derivative $g_D(T)$ nor the expectation $g(T)$ need exist, and similar for $f_D(T)$ and $f(T)$. For example:

(a) Consider the case that Q almost surely picks out a family of indifferent choice-worthiness functions. That is, with probability 1 we have $S_D(A) = S_D(B)$ for all decision-situations $D \in \mathcal{D}$ and all options $A, B \in D$. Pick some D, options $A, B \in D$, and choice-worthiness functions T such that $T_D(A) \neq T_D(B)$. Then almost surely any non-zero perturbation of S_D in the direction of T_D changes the set of top options. That is, we have $P_{T,D}(p) = 1$ for all $p \neq 0$, but of course $P_{T,D}(0) = 0$. Thus, $P_{T,D}(p)$ is not continuous in $p = 0$, and in particular has no derivative in $p = 0$.
(b) Even if $g_D(T)$ exists for all $D \in \mathcal{D}$, its expectation across decision-situations need not. To give an example, we will fix an arbitrary numbering $\mathcal{D} = \{D_n \mid n \in \mathbb{N}\}$. It is easy to see that there is a prior Q such that $g_{D_n}(T) = \frac{1}{nP(D_n)}$ for all n; the expectation $g(T)$ is then given by the harmonic series $\sum_n \frac{1}{n}$, which does not converge to a finite value.

We will later make assumptions that guarantee that all of $g_D(T), g(T), f_D(T)$, and $f(T)$ are well-defined and finite.

Definition B.3. A normalization method N (when used with MEC as aggregation method and relative to priors P and Q) is

i. *fair with respect to probabilities* (in giving equal say to all moral theories with respect to P and Q) if and only if, for any two moral theories $\mathbf{T}$ and $\mathbf{T}'$, we have $g(N(\mathbf{T})) = g(N(\mathbf{T}'))$;
ii. *self-consistently fair with respect to expected choice-worthiness* (in giving equal say to all moral theories with respect to P and Q) if and only if, for any two moral theories $\mathbf{T}$ and $\mathbf{T}'$, we have $f(N(\mathbf{T})) = f(N(\mathbf{T}'))$.

These definitions are understood to in particular require that $g(N(\mathbf{T}))$ and $f(N(\mathbf{T}))$, respectively, are well-defined and finite for all theories $\mathbf{T}$.

Remark B.4. In our definitions of $g(T)$ and $f(T)$ we have first defined a derivative capturing an intuitive notion from voting theory *within* each decision-situation $D \in \mathcal{D}$; then we took the expectation across decision-situations. We could also have proceeded the other way around, as follows:

1. Define $P_T(p) = \mathbb{E}_{D \in P}[P_{T,D}(p)]$, that is, *first* take the expectation of the *probability* that a slight perturbation will change the set of top options.
2. Define $\tilde{g}(T)$ to be the derivative of $P_T(p)$ in $p = 0$.

It is not clear if our intuitive notion of fairness with respect to probabilities is better captured by defining it in terms of $g(T)$ or $\tilde{g}(T)$. If we define $\tilde{f}(T)$ similarly, this applies mutatis mutandis to self-consistent fairness with respect to expected choice-worthiness. We will therefore make assumptions guaranteeing that $g(T) = \tilde{g}(T)$ and $f(T) = \tilde{f}(T)$.

In order to say anything further about which methods might be fair in either sense, we need to make some assumptions about the priors P and Q. We want to assume that they are essentially ignorant, analogous to the assumption for computing voting power that other people are equally likely to vote in all possible combinatorial permutations. But rather than assume a specified form for these priors, we will just make assumptions about some of their properties.

Smoothness assumption (on Q): For all $D \in \mathcal{D}$, the marginal distribution Q_D on $\mathbb{R}^{\#D}$ has a continuously differentiable cumulative distribution function.

Boundedness assumption (on P and Q): There is a non-negative random variable X on $\mathcal{D}$ that has a finite expectation relative to P and such that we have $|\frac{P_{T,D}(p)}{p}| < X(D)$ for all $D \in \mathcal{D}$ and $p \neq 0$. (Note that $P_{T,D}(p)$ depends on Q.)

First ignorance assumption (on Q): Using MEC on Q for each decision-situation results in a measure that is symmetric in options; that is, shuffling the labels of options in the description of an event does not change its probability according to any Q_D.

Second ignorance assumption (on Q): For all $D \in \mathcal{D}$, the derivative of the cumulative distribution function of Q_D vanishes almost nowhere; that is, its set of zeros has Lebesgue measure 0. (In other words, the Lebesgue density of Q_D is non-zero almost everywhere.)

Third ignorance assumption (on P): According to P, the probability of a given option appearing in a decision-situation is independent of which other options appear there.

Note that the boundedness assumption allows us to use the dominated convergence theorem to conclude that $g(T) = \tilde{g}(T)$ and $f(T) = \tilde{f}(T)$ (see Remark B.4 for notation and context).

We are now in a position to state the theorem.

Theorem B.5. Suppose that P and Q satisfy the above assumptions. Then:

1. The normalization method N is fair with respect to probabilities if and only if N normalizes the mean absolute difference of the choice-worthiness functions—that is,

$$\mathrm{MAD}_P(\mathbf{T}) := \sum_{A,B \in \mathcal{O}} \mathbb{P}_{D\sim P}(A \in D)\mathbb{P}_{D\sim P}(B \in D)|N(\mathbf{T})_D(A) - N(\mathbf{T})_D(B)|$$

   does not depend on the theory $\mathbf{T}$; and
2. N is self-consistently fair with respect to expected choice-worthiness if and only if N normalizes the variance of the choice-worthiness functions—that is,

$$\mathrm{Var}_P(\mathbf{T}) := \sum_{A,B \in \mathcal{O}} \mathbb{P}_{D\sim P}(A \in D)\mathbb{P}_{D\sim P}(B \in D)(N(\mathbf{T})_D(A) - N(\mathbf{T})_D(B))^2$$

   does not depend on the theory $\mathbf{T}$.

In the proof of the theorem, we will use the following notation for fixed $T \in \mathcal{T}$, $D \in \mathcal{D}$, $A, B \in D$, and $p < 1$.

i. $P_{T,D,A,B}(p)$ denotes the probability (according to Q) that S_D and $pT_D + (1-p)S_D$ differ in their ranking of options A and B; and
ii. $E_{T,D,A,B}(p)$ denotes the expected choice-worthiness (in units according to T_D) of that flip in ranking—that is, $E_{T,D,A,B}(p) = P_{T,D,A,B}(p)(T_D(A) - T_D(B))$.

We will also use the following:

Lemma B.6. Suppose that the above assumptions are satisfied. Then, for each decision-situation D, the derivatives $g_D(T)$ and $f_D(T)$ depend only

on the event that S_D and $pT_D + (1-p)S_D$ differ in their ranking of the top and a *single* other option. That is,

$$g_D(T) = \lim_{p\to 0} \frac{1}{p} \sum_{B\in D} \mathbb{P}_{S\sim Q}(S_D \text{ has top option } B) \sum_{A\in D: T_D(A)>T_D(B)} P_{T,D,A,B}(p)$$

$$f_D(T) = \lim_{p\to 0} \frac{1}{p} \sum_{B\in D} \mathbb{P}_{S\sim Q}(S_D \text{ has top option } B) \sum_{A\in D: T_D(A)>T_D(B)} E_{T,D,A,B}(p)$$

Proof (of the Lemma). Given that we are only adding a small amount of credence to the theory, it is unlikely that are we are able to affect the decision at all. But it is vanishingly unlikely that we are able to affect a choice between three or more outcomes, so it is enough to consider the chance of moving it between each pair of outcomes (formally speaking the chance of being able to affect it between two outcomes is $O(p)$, and the chance of being able to affect it between three or more outcomes is $O(p^2)$ and thus vanishes as we take the derivative in $p = 0$). □

Proof (of the Theorem). Let $\mathbf{T}$ be a moral theory; to avoid clutter, set $T = N(\mathbf{T})$. The proof will proceed by showing that $g(T)$ and $f(T)$ are proportional to the mean absolute difference and variance of T, respectively. We will only give a full proof of the first statement; a proof of the second statement can then be obtained by multiplying in every step by $T_D(A) - T_D(B)$.

Step 1: We calculate $P_{T,D,A,B}(p)$ and its derivative in $p = 0$.

Two choice-worthiness functions S_D and S'_D rank A and B differently if and only if $S_D(A) - S_D(B)$ and $S'_D(A) - S'_D(B)$ have different signs. Therefore,

$$\begin{aligned}
&P_{T,D,A,B}(p) \\
&= \mathbb{P}_{S\sim Q}\Big((S_D(A) - S_D(B))\big((1-p)(S_D(A) - S_D(B) + p(T_D(A) - T_D(B))\big) < 0\Big) \\
&= \mathbb{P}_{S\sim Q}\left(\frac{p}{p-1}(T_D(A) - T_D(B)) < S_D(A) - S_D(B) < 0\right),
\end{aligned}$$

where for the second equality we have without loss of generality assumed that $T_D(A) > T_D(B)$. (Note that $P_{T,D,A,B}(p) = 0$ if $T_D(A) = T_D(B)$.)

The smoothness assumption implies that $S_D(A) - S_D(B)$ is a real-valued random variable (where S is distributed according to Q) with a Lebesgue density and a continuously differentiable cumulative distribution function $F_{D,A,B}$. We thus have:

$$P_{T,D,A,B}(p) = F_{D,A,B}(0) - F_{D,A,B}\left(\frac{p}{p-1}(T_D(A) - T_D(B))\right).$$

By the chain rule, since $F_{D,A,B}$ is differentiable in $\frac{p}{p-1}(T_D(A) - T_D(B))$, we can calculate the derivative in p as

$$P'_{T,D,A,B}(p) = F'_{D,A,B}\left(\frac{p}{p-1}(T_D(A) - T_D(B))\right)\frac{1}{(p-1)^2}(T_D(A) - T_D(B)).$$

In particular, for $p = 0$ we get

$$(3)\quad P'_{T,D,A,B}(0) = F'_{D,A,B}(0)(T_D(A) - T_D(B))\ .$$

Step 2: We calculate $g_D(T)$ for fixed D.

By the first ignorance assumption, the cumulative distribution function $F_{D,A,B}$ introduced in the first step depends only on the number of options $\#D$ in D, and in particular is independent of A and B—going forward, we will denote it by $F_{\#D}$. Using this notation, Lemma B.6 and the first step imply that

$$\begin{aligned} g_D(T) &= \lim_{p\to 0} \frac{1}{p} \sum_{B\in D} \mathbb{P}_{S\sim Q}(S_D \text{ has top option } B) \sum_{A\in D:\, T_D(A)>T_D(B)} P_{T,D,A,B}(p) \\ &= \frac{1}{\#D} \sum_{A,B\in D:\, T_D(A)>T_D(B)} P'_{T,D,A,B}(0) \\ &= \frac{F'_{\#D}(0)}{2\#D} \sum_{A,B\in D} |T_D(A) - T_D(B)|\ . \end{aligned}$$

Step 3: We take the expectation over decision-situations $D \in \mathcal{D}$ according to P to obtain $g(T)$ from the $g_D(T)$.

By the second step and the ignorance assumptions, we have

$$\begin{aligned} g(T) &= \mathbb{E}_{D\sim P}[g_D(T)] \\ &= \sum_{D\in\mathcal{D}} P(D) g_D(T) = \sum_{D\in\mathcal{D}} P(D) \frac{F'_{\#D}(0)}{2\#D} \sum_{A,B\in D} |T_D(A) - T_D(B)| \\ &= \frac{1}{2} \sum_{n=2}^{\infty} \frac{F'_n(0)}{n} \sum_{D\in\mathcal{D}:\#D=n} P(D) \sum_{A,B\in D} |T_D(A) - T_D(B)| \\ &= \frac{1}{2} \sum_{A,B\in\mathcal{O}} \sum_{n=2}^{\infty} \mathbb{P}_{D\sim P}(\#D = n) \frac{F'_n(0)}{n} \\ &\qquad\qquad \mathbb{P}_{D\sim P}(A \in D)\mathbb{P}_{D\sim P}(B \in D)|T_D(A) - T_D(B)|. \end{aligned}$$

Now consider $k := \sum_{n=2}^{\infty} \mathbb{P}_{D\sim P}(\#D = n)\frac{F'_n(0)}{n}$, which does not depend on any of T, D, A, B. The boundedness assumption implies that k is finite, for else we could express $P_{T,D}(p)$ in terms of the probabilities $P_{T,D,A,B}(p)$ calculated in step 1 and use equation (3) to derive that $|\frac{P_{T,D}(p)}{p}|$ has infinite expectation across decision-situations D. Also, $k \neq 0$ by the second ignorance assumption. We thus have seen that $g(T) = \frac{k}{2}\,\mathrm{MAD}_P(T)$ is proportional to $\mathrm{MAD}_P(T)$, as desired. □

APPENDIX C. NARROW STATISTICAL NORMALIZATION METHODS MAKE CYCLICAL RECOMMENDATIONS ACROSS CHOICE SITUATIONS

Proposition C.1. There is a decision-maker with fixed credences in fixed moral theories, so that applying any narrow statistical normalization method will result in cyclical recommendations of options over one another, in varying choice-situations.

Proof. The proof is straightforward, based on the fact that when there are only two options in a choice situation, all narrow statistical normalization methods must make the same recommendation. Let A, B, and C be options, and consider three moral theories with choice-worthiness across these options as indicated by R, S, and T in Table 2.

	R	S	T
A	0	2	1
B	1	0	2
C	2	1	0

Table 2. Choice-worthiness functions generating cyclic preferences.

Suppose the decision-maker has credence 0.4 in R and 0.3 in each of S and T, and that she faces a choice between A and B only. All that a narrow and statistical normalization method can see of each theory is whether it prefers A or B, and it must treat each of these in the same way. Since it is a normalization method, they are all normalized to the same thing—without loss of generality, the preferred option at 1 and the less preferred option at 0. Then the expected choice-worthiness of A is $0.4 \cdot 0 + 0.3 \cdot 1 + 0.3 \cdot 0 = 0.3$. The expected choice-worthiness of B is $0.4 \cdot 1 + 0.3 \cdot 0 + 0.3 \cdot 1 = 0.7$. In effect the procedure has reduced to asking whether there is more credence on theories preferring A or B. In this case credence 0.7 lay with theories preferring B, so the decision-maker will choose B over A.

Suppose now that the decision-maker faces instead a decision between B and C. Again the theories will all be normalized, so we need know only the total credences preferring each option. Now R and S (with total credence 0.7) prefer C to B, so the decision-maker will choose C over B.

Finally suppose the decision-maker faces a decision between C and A. Here R prefers C to A, but the other two theories prefer A to C. Since there is credence 0.6 in these theories the decision-maker will choose A over C. □

WILLIAM MACASKILL
OWEN COTTON-BARRATT
TOBY ORD

University of Oxford