Metanormative Regress: An Escape Plan

Christian J. Tarsney*

Version 8, July 2022
(Latest version here.)

Comments welcome: christian.tarsney@philosophy.ox.ac.uk

Abstract

How should you decide what to do when you're uncertain about basic normative principles (e.g., Kantianism vs. utilitarianism)? A natural suggestion is to follow some "second-order" norm: e.g., *comply with the first-order norm you regard as most probable* or *maximize expected choiceworthiness*. But what if you're uncertain about second-order norms too—must you then invoke some *third*-order norm? If so, it seems that any norm-guided response to normative uncertainty is doomed to a vicious regress. In this paper, I aim to rescue second-order norms from this threat of regress. I first elaborate and defend the suggestion some philosophers have entertained that the regress problem forces us to accept *normative externalism*, the view that at least one norm is incumbent on agents regardless of their beliefs or evidence concerning that norm. But, I then argue, we need not accept externalism about first-order (e.g., moral) norms, thus closing off any question of what an agent should do in light of her normative beliefs. Rather, it is more plausible to ascribe external force to a single, second-order rational norm: the *enkratic principle*, correctly formulated. This modest form of externalism, I argue, is both intrinsically well-motivated and sufficient to head off the threat of regress.

1 Introduction

How should an agent decide what to do when she is uncertain about basic normative principles—for instance, when she is uncertain whether Kantianism or utilitarianism is the true moral theory and faces a choice for which those theories offer conflicting advice? Many philosophers have thought that such an agent should decide what to do by means of some *higher-order* normative principle. For instance, according to "My Favorite Theory" (MFT), she should act on the first-order normative theory she regards as most probably correct. According to "My Favorite Option" (MFO), she should choose the *option* that has the greatest total probability of being objectively right or permissible. According to still other views, she should weigh the reasons put forward by the various first-order normative theories against one another, perhaps by choosing the option that

*Global Priorities Institute, Faculty of Philosophy, University of Oxford

is best in expectation, given her credences over first-order theories.[1] The view that normatively uncertain agents should make deliberative use of some such higher-order norms has been dubbed *metanormativism* (MacAskill, 2014).

Though the need for higher-order norms may seem inescapable, all forms of metanormativism face a challenge: If an agent who is uncertain about first-order normative principles must decide what to do by means of some second-order principle, must not an agent who is uncertain about *second*-order principles decide what to do by means of some *third*-order principle—and likewise for every higher order of normative belief? Without some guarantee that, in the course of ascending to higher-order normative principles, a rational agent will eventually reach a point at which she experiences no further uncertainty (being certain that some nth-order normative principle is correct), won't the appeal to higher-order norms involve her in an infinite regress that prevents her from ever reaching a rationally guided decision?

This apparent regress is a threat to metanormativism, and to the philosophical project of identifying norms of choice under normative uncertainty.[2] The easiest way to avoid the regress problem, it might seem, is to eschew higher-order norms entirely and instead adopt the view I will call *first-order externalism*, according to which the true first-order (e.g., moral) norms are incumbent on all agents regardless of their beliefs or evidence. On this view, an agent who is uncertain about first-order norms simply ought to do what the *true* first-order norms require of her, even if she has no way of identifying those norms and even if her evidence leads her rationally to reject them and to place most of her credence in rival norms. Thus, for instance, if eating factory-farmed meat is in fact morally permissible, and slightly prudentially better than any alternative diet, then even an agent who believes on the basis of compelling arguments that it is almost certainly a serious moral wrong ought to eat meat anyway, in every interesting sense of "ought."[3]

In this paper, I propose a solution to the metanormative regress problem. My solution preserves metanormativism, and hence the idea that what we ought to do (in at least one important sense of "ought") depends on our normative beliefs. But it also concedes an important point to opponents of metanormativism: There must be *some* normative principle whose normative force does not depend on an agent's beliefs, and which therefore is incumbent on an agent even if she justifiably rejects that principle itself. Conceding this limited form of normative externalism is the price we must pay to avoid a vicious regress. But, I will argue, both externalism in general and my version of externalism in particular have strong independent motivations, and are not merely an *ad hoc* response to the threat of regress.

[1] Versions of MFT are defended by Gracely (1996) and Gustafsson and Torpman (2014). MFO is considered and rejected by Lockhart (2000) (under the name "PR2"), Gustafsson and Torpman (2014), and MacAskill and Ord (2020). Expectational views are defended by Lockhart (2000), Ross (2006), Sepielli (2009), MacAskill and Ord (2020), and Riedener (2020), among others. Other metanorms have been proposed by Guerrero (2007), Nissan-Rozen (2012), MacAskill (2016), Tarsney (2018, 2019), and Greaves and Cotton-Barratt, among others.

[2] This threat has been noted by Lockhart (2000, pp. 36–7), Sepielli (2010, pp. 267ff), MacAskill (2014, pp. 217–9), Bykvist (2013, pp. 132–4), Weatherson (2014, 2019) and Riedener (2015, pp. 25–31, 91–2), among others.

[3] Weatherson (2014, 2019) takes the regress problem, among other considerations, to support precisely this view. First-order externalism has also been defended on other grounds by Harman (2015) and Hedden (2016).

The paper therefore has two objectives: first, to explicate the regress problem and show how it supports externalism, but second, to defend a metanormative rather than a first-order version of externalism as the best response to the threat of regress.

In the next section, I introduce some conceptual scaffolding for the rest of the paper. In §3, I set out the regress problem as an argument for normative externalism, consider two internalist responses, and conclude that they are unsatisfactory. In §4, I introduce and motivate my own response to the regress problem, which posits a single belief-independent norm of practical rationality: the enkratic principle, correctly formulated. In §5, I consider a way in which enkratic externalism might fail to solve the regress problem and propose a tentative response to this worry. §6 is the conclusion.

2 Internalism, externalism, and metanormativism

2.1 Choice situations and norms

A *choice situation* is an ordered triple $S_i = \langle A_i, \mathbf{O}_i, Cr_i \rangle$, where A_i is an agent, $\mathbf{O}_i$ is a finite set of options $\{O^i_1, O^i_2, ..., O^i_n\}$ available to A_i in S_i, and Cr_i is A_i's credence function. Each option is understood as a vector of properties that completely specifies all its normatively relevant features.

A *norm* is a principle for making normative assessments of options in the context of particular choice situations. Formally, we can understand a norm as a set of propositions closed under logical consequence that includes such normative assessments. I leave it open exactly what form these assessments take (e.g., a preordering of options or an assignment of real numbers), except to stipulate that all norms have the purpose of identifying some options as *permissible* and others as *impermissible*.[4] That is, every norm must include at least some propositions to the effect that particular options are permissible or impermissible in particular choice situations. Thus, any norm is associated with two functions: one that maps some or all choice situations to *choice sets* of options that the norm designates as permissible in that choice situation, and another that maps choice situations to *prohibited sets* of options it designates as impermissible. (These functions are not redundant, as we will see, since a norm may be only partial: It may classify some options in a choice situation as permissible and others as impermissible, while leaving still others unclassified.)

2.2 Objective and subjective norms

Say that a norm N is *sensitive to* a given feature of a choice situation if, for some minimal pair of choice situations S_i and S_j that differ only with respect to that feature, containing a minimal pair of options $O^i_k \in \mathbf{O}_i$ and $O^j_l \in \mathbf{O}_j$ that differ only with respect to that feature, N asserts that O^i_k is permissible in S_i but O^j_l is impermissible in S_j. In particular, a norm N is sensitive to an agent's beliefs about some set of propositions Σ if there is some minimal

[4]These notions should be understood very thinly. To say that an option is "permissible" is just to say that it is possible for an agent who is in some relevant sense normatively ideal (fully rational and, in the case of objective norms, fully informed) to choose that option. To say that an option is "impermissible" is just to say that such an agent would not choose that option.

pair of choice situations S_i and S_j that differ only with respect to the agent's credences over propositions in Σ, such that N designates some minimal pair of options permissible in S_i but impermissible in S_j. In other words, a norm is sensitive to an agent's belief about a given subject matter if, according to that norm, those beliefs can make the difference between an otherwise identical option being permissible or impermissible.

A *subjective* norm is sensitive exclusively to facts about the agent's mental states, in particular her beliefs and/or evidence.[5] The output of a subjective norm is an assessment of options in terms of the degree of subjective reason the agent has to choose each option, and a designation of options as subjectively/rationally permissible or prohibited.

An *objective* norm is any norm that is not subjective, i.e., that is sensitive to features of choice situations other than the agent's beliefs. For instance, an objective norm may assess a given option as impermissible because it would harm some third party, even if the agent confidently believes that it would benefit them.[6] Just as subjective norms yield an assessment of options in terms of subjective reasons and designate options as rationally permissible/prohibited, so objective norms yield an assessment of options in terms of *choiceworthiness*, i.e., the degree of objective reason to choose a given option, and designate options as *objectively* permissible/prohibited.

Subjective norms tell an agent how to respond appropriately, in some sense, to her total belief state. There might be multiple distinct senses in which an option can be an appropriate response to one's beliefs (e.g., coherence, fittingness, instrumental optimality with respect to particular goals like moral or prudential value). But I will focus on one kind of subjective norm, namely, norms of *rationality*. My arguments and conclusions below are meant to require no commitments either regarding the existence of other kinds of subjective norm, or the exact character of rational norms (e.g., whether they are coherence requirements). I merely assume that rational requirements are *one* significant kind of subjective norm, describing one way in which an agent's choices can appropriately or inappropriately reflect her belief state.[7]

My own inclination is to treat rational requirements as synonymous with the "subjective *ought*"—i.e., an agent subjectively ought do something just in case she is rationally required to do it. So I would be equally happy to defend all the following arguments and conclusions if "subjective ought" (or "subjective rightness") were substituted for "rational requirement" throughout. But I do

[5] I will hereafter use "beliefs" to mean "beliefs and/or evidence," remaining neutral on whether the true subjective norms are sensitive to an agent's beliefs, her evidence, or some combination of the two. I will assume that an agent's beliefs come in the form of *credences* (i.e., degrees of belief that satisfy the axioms of probability) and that her evidence gives rise either to evidential probabilities or to evidential constraints on credences. Again for concision, I will use "credences" to mean "either subjective credences, or evidential probabilities, or some combination of the two (e.g., subjective credences constrained by evidence)".

[6] An objective norm need not be entirely *in*sensitive to the agent's belief state. For instance, the fact that I am highly uncertain (have intermediate credence) about some important proposition might be an objective reason to seek out information concerning it.

[7] My arguments do depend to a significant extent, however, on the assumption that norms of rationality tell an agent how to respond appropriately to her *entire* belief state. In particular, there are not (merely) distinct norms of "first-order rationality" that tell agents how to respond to their empirical beliefs, norms of "second-order rationality" that tell agents how to respond to their first-order beliefs, and so on, without any overall rational norms to adjudicate between them (cf. the fourth point in footnote 22 below).

not intend to presuppose this identification.[8]

2.3 Higher-order subjective norms

All subjective norms, I assume, are sensitive to (at least some features of) an agent's non-normative beliefs—e.g., her beliefs about the consequences of her options or about what promises she has made. A *first-order* subjective norm is sensitive *only* to the agent's non-normative beliefs, and insensitive to her normative beliefs. A *second-order* subjective norm is sensitive to (i) the agent's non-normative beliefs as well as (ii) her beliefs about objective norms and/or first-order subjective norms, but insensitive to her beliefs about higher-order subjective norms. For finite $n > 2$, an nth-order subjective norm is sensitive to the agent's non-normative beliefs as well as her beliefs regarding norms of order $n - 1$ (and possibly lower-order norms as well), but not her beliefs regarding norms of order n or greater. And more generally, for any subjective norm N that is not first- or second-order, the order of N is the least ordinal greater than every order of normative belief to which N is sensitive. (Henceforth I will generally omit the word "subjective" and simply refer to "first-order norms," "second-order norms," etc.)[9,10]

[8] One might worry that, by framing the debate in terms of rationality, I am talking past opponents of metanormativism like Weatherson, Harman, and Hedden, who are often understood to be interested in properties like *moral rightness* rather than rational requirement. But I am talking past these philosophers only if they are prepared to concede metanormativism as a thesis about rationality. And it seems clear that they are not, because they deny the need for *any* kind of metanorms. For instance, Harman writes: "Because Uncertaintism [≈ metanormativism] is false, the puzzle we discussed above, about how to compare moral value between conflicting moral views, is not important. It may be interesting *as a puzzle*; but nothing normatively important hangs on solving it" (Harman, 2015, p. 58). And Hedden writes: "There is no normatively interesting sense of *ought* in which what you ought to do depends on your uncertainty about (fundamental) moral facts" (Hedden, 2016, p. 104).

[9] We could assign orders to subjective norms more elegantly by simply saying that for *any* subjective norm N, the order of N is the least ordinal greater than every order of subjective normative belief to which N is sensitive. But then we would lose the distinction between norms that are sensitive only to the agent's empirical beliefs and those that are sensitive to her objective normative beliefs, and would classify norms in a way that does not match the standard usage of "first-order" and "second-order" in the normative uncertainty literature. So I have adopted unnecessarily clunky definitions in order to interface better with the existing debate.

[10] Could there also be norms that are sensitive to an agent's normative beliefs at *every* order, or at an upwardly unbounded collection of orders? I'm not sure, but I think we ought to set this possibility aside. One the one hand, we might wish to rule out such norms as potential sources of paradox (just as, in standard set theory, upwardly-unbounded sets of ordinals are ruled out by the Burali-Forti paradox). On the other hand, if we allow such norms, then they themselves become potential objects of uncertainty, and that uncertainty will beget yet further metanorms. So we would only have succeeded in extending the metanormative hierarchy from ordinal to "super-ordinal" norms, without changing our situation in any fundamental way.

The essential point is that, like the ordinals themselves, the metanormative hierarchy appears to be "indefinitely extensible": We have a general procedure for going from any identified totality T of orders to a new order not contained in T—namely, the order of those norms that are sensitive to an agent's beliefs about all and only the norms in T. For discussion of the challenges raised by indefinite extensibility, see for instance Hellman (2006) and Shapiro and Wright (2006).

2.4 Internal and external subjective norms

Say that an option O in choice situation S is *in the domain* of norm N just in case N asserts either that O is permissible in S or that O is impermissible in S. Just as we said that N is "sensitive to" some feature of a choice situation if varying that feature alone can make an otherwise permissible option impermissible or vice versa, so we will say that N is *restricted by* some feature of a choice situation if varying that feature alone can affect whether an option O is in the domain of N.

Although an nth-order norm is not sensitive to an agent's beliefs about subjective norms of order n or higher, it can be restricted by them. In this case, we will call it an *internal* norm. These restrictions could in principle take many forms, but their natural motivation is the idea that an agent cannot be rationally required to obey a norm if she disbelieves either that norm itself, or its particular permissibility judgements. For instance, consider the competing second-order norms MFT (which tells an agent to follow the first-order norm in which she has greatest credence) and MFO (which tells her to choose an option for which her total credence in all first-order norms that permit it is maximal). If an agent is sufficiently confident that MFO is correct and MFT is incorrect, either in general or with respect to some particular option, she cannot be rationally required to obey MFT (the thought goes). This is so even if MFT is in some sense the *true* second-order norm—e.g., in the sense that sufficient rational deliberation would lead the agent to accept it.

Thus, an internal norm N paradigmatically applies only when the agent satisfies some threshold of belief with respect to either N itself, or particular rational permissibility judgements made by N. An internal nth-order norm N^n might assert that a sufficient condition for an option O to be rationally permissible is that (i) O satisfies some condition φ that makes reference only to the agent's non-normative beliefs and her normative beliefs of order less than n and (ii) the agent believes with sufficient confidence (e.g., with certainty, or credence greater than 0.5) either N^n itself or simply that O is rationally permissible. When only the first of these conditions is satisfied, N^n remains silent.

Because norms can be internal, it's possible for two apparently competing subjective norms to be true, despite the fact that all subjective norms assess options in terms of the same normative concepts. For instance, perhaps the true first-order norm says that agents rationally ought to maximize expected total welfare, while the true second-order norm says that agents rationally ought to follow the first-order norm in which they have greatest credence. On face, these two norms can disagree about what a given agent ought to do in a given choice situation, and so can't both be true at the same time. But if the first-order norm applies only to agents who believe it with certainty, or with credence greater than 0.5, then there is no conflict: Whenever both norms apply, they yield the same permissions and prohibitions.

An *external* nth-order norm, by contrast, has no such restrictions: Whether a given option is in its domain does not depend on the agent's normative beliefs of order n or greater. *Externalism*, then, is the thesis that there is at least one true external norm N, which applies to agents regardless of their belief in or agreement with N itself. *Internalism* is the thesis that all true norms are

internal.[11]

2.5 Metanormativism and first-order externalism

An nth-order norm N is *comprehensive* if (i) it is restricted, if at all, only by the agent's normative beliefs of order n or higher and (ii) it has in its domain, at a minimum, every option in any choice situation where the agent believes all the propositions in N with probability 1. Because external nth-order norms are not restricted by an agent's normative beliefs of order n or higher, a comprehensive external norm is completely unrestricted—that is, it classifies every option in every choice situation as either rationally permissible or rationally impermissible.

Now we can characterize metanormativism and its competitor, first-order externalism. *Metanormativsm* is the view that there is at least one true second- or higher-order subjective norm (i.e., a true nth-order subjective norm for some $n > 1$). Thus, metanormativism asserts that what an agent rationally ought to do sometimes depends on her purely normative beliefs, and is not determined solely by her empirical and other non-normative beliefs. The rival view, *first-order externalism*, asserts that there is a true comprehensive external first-order norm, N^1_*. This implies that there are no true higher-order norms: Since N^1_* is comprehensive and external, it determines the rationality of every option in every choice situation. Therefore, any true subjective norm can only yield assessments that agree with N^1_*, on pain of contradiction. But, since N^1_* is a first-order norm, any norm that always agrees with it is insensitive to the agent's normative beliefs, and therefore is also first-order.

The thesis I will defend in the coming sections, then, is that the regress problem forces us to accept *some* form of externalism but does not force us to accept *first-order* externalism—rather, the most plausible response to the regress problem is a form of externalist metanormativism.

3 The Regress Argument

We can now state the regress problem more precisely, and see why it supports externalism. My strategy here will be slightly indirect: I will give an argument for externalism, based on the threat of infinite regress, that applies to agents with unbounded capacities for theoretical and practical reasoning (of whom I will say more shortly). I will then argue that if externalism is true with respect to these unbounded agents, it is true of bounded agents as well.

[11] The internalist/externalist distinction is borrowed from Weatherson (2014, 2019), though I characterize it somewhat differently than he does. (For Weatherson's characterization, see in particular §1.3 of Weatherson (2019).) Other philosophers have recognized the same distinction in various terms. For instance, Broome endorses the view I am calling externalism when he says that some norms impose "strict liability" (e.g., in Broome (2013, pp. 91ff)). Bykvist (2013) endorses the same thesis when he writes: "[M]y tentative conclusion is that in cases of uncertainty of rational matters there is an answer to the question of what it is rational to prefer which is not sensitive to your own views about rationality" (p. 133). Lin (2014) endorses a different form of externalism based on the idea of "adaptive rationality." And I take Elga (2010) to endorse externalism in the epistemic domain when he says that certain epistemic norms "must be dogmatic with respect to their own correctness" (p. 185). (Although Elga's focus is on epistemic norms, he seems to endorse externalism regarding practical norms as well when he claims that this requirement of dogmatism applies to any "fundamental policy, rule, or method" (p. 185).)

3.1 Stating the argument

Here is an intuitive gloss of the argument: If the force of any norm N depends on the agent's beliefs about N, then agents (or at least unbounded agents) can't rationally act on norms of which they're uncertain, without somehow accounting for that uncertainty. When an agent has some credence in a conflicting norm of the same order that disagrees with N about which options are permissible, then the only way to account for her uncertainty is to invoke a higher-order norm. But if she finds—as seems likely—that she has credence in conflicting norms at *every* order, then she will not be able to make a rationally guided decision based on norms of *any* order. Thus internalism implies that, for agents who are generally uncertain about basic normative principles, rational choice is impossible. And this conclusion seems unacceptable.

Let's state the argument more carefully, so that we can assess it premise by premise. To avoid repeating a cumbersome locution, I will say that an nth-order norm N *authorizes* option O in situation S if either (i) N asserts that O is permissible in S or (ii) in any minimal variant of S that merely alters A's nth- or higher-order normative beliefs to place O in the domain of N, N asserts that O is permissible. Likewise, N *deauthorizes* O in S if either (i) N asserts that O is impermissible in S or (ii) in any minimal variant of S that alters A's nth- or higher-order normative beliefs to place O in the domain of N, N asserts that O is impermissible.

The Regress Argument for Externalism

P1. An agent A is rationally permitted to choose option O in situation S only if there is some true subjective norm N such that (i) N authorizes O in S and (ii) A's beliefs place O in the domain of N.

P2. Suppose A is an unbounded agent, N_i^n is a true nth-order norm that authorizes option O, and A assigns positive credence to some rival nth-order norm N_j^n that deauthorizes O. If internalism is true, then A's beliefs do not place O in the domain of N_i^n, unless there is some true higher-order norm N^p $(p > n)$ that authorizes O in light of A's nth-order normative beliefs, and A's normative beliefs place O in the domain of N^p.[12]

L1. If (i) internalism is true, (ii) A is an unbounded agent, and (iii) for all $n \leq m$, A has positive credence in some nth-order norm that deauthorizes O, then A is permitted to choose O only if there is some norm of order

[12]This premise is meant to allow that A is permitted to choose O on the basis of her nth-order normative beliefs alone, even under nth-order normative uncertainty, so long as all the nth-order norms in which she has positive credence authorize O. First, since norms are merely sets of propositions closed under logical consequence, any disjunction of norms is itself a norm (as long as it implies at least one permissibility or impermissility judgement). The disjunction $N_1 \vee N_2 \vee ... \vee N_n$ yields a norm (generally non-comprehensive) whose choice set in situation S is the intersection of the choice sets of norms N_1–N_n. Thus, if A is uncertain between various nth-order norms, but assigns positive credence to at least one true norm, and all the nth-order norms to which she assigns positive credence assert that O is permissible, then there is a true nth-order norm to which she assigns credence 1 that authorizes O (viz., the disjunction of all the nth-order norms in which she has positive credence). Because she assigns that norm credence 1, she presumably meets the belief conditions that place O its domain. But second, even if A is not certain of any nth-order norm that authorizes O, P2 asserts that she must resort to a higher-order norm in order to permissibly choose O only when she assigns positive credence to some nth-order norm that deauthorizes O.

greater than m that authorizes O and such that A assigns no credence to any norm of the same order that deauthorizes O. [from P1, P2]

P3. Necessarily, for any agent A facing an option O, and for any ordinal n, A is rationally required to have positive credence in some nth-order norm that deauthorizes O.

L2. If internalism is true, then, necessarily, an unbounded agent who satisfies all the requirements of epistemic rationality is never rationally permitted to choose any practical option. [from L1, P3]

P4. It's at least sometimes possible for unbounded agents to make choices in a way that satisfies all the requirements of both epistemic and practical rationality.

C. Internalism is false. [from L2, P4][13]

Several remarks are immediately in order. First, what is the intended target of the argument—an "unbounded agent"? Unbounded agents, in the sense I have in mind, represent a particular limited idealization of human agency. An unbounded agent (i) has conceptual resources at least as rich as our own, (ii) maintains probabilistically coherent beliefs about all the propositions she can construct from those conceptual resources, (iii) assigns probability 1 to all logical truths and probability 0 to all logical falsehoods, (iv) can instantaneously and costlessly update her beliefs in response to new evidence, and (v) has perfect introspective/recollective access to her own beliefs and evidence.[14]

Unbounded agents are, we might say, "computationally omniscient," in the sense that they face no purely computational constraints: For instance, any reasoning that a human being could carry out with unlimited time, pencils, and paper, an unbounded agent can carry out at no cost in time or resources. But unbounded agents are not "*a priori* omniscient"—they do not have perfect *a priori* insight that leads them to assign probability 1 to all *a priori* truths and probability 0 to all *a priori* falsehoods. Both these features play an important role in the Regress Argument: If we consider less idealized agents for whom deliberation is costly, we may have grounds to reject P2 (see Tarsney, ms). If we consider more idealized *a priori* omniscient agents, we could reject P3 (assuming that normative truths are *a priori*). But for this sort of agent, the question of internalism vs. externalism is moot anyway, since they are immune

[13]This presentation of the regress problem was originally inspired by remarks in Weatherson (2014), though I now take Weatherson to be making a slightly different argument, related to the "Argument from Fallibility" discussed in the next section. Sepielli seems to have something like the preceding argument in mind in this passage: "We can imagine someone who is...uncertain at *all* levels [of subjective normativity]. Indeed, one would suspect that this blanket uncertainty is typical. For who among us is certain about *morality*, let alone such esoterica as 8th-order, or 1,000th-order, normative uncertainty? But recall what animated our Divider [someone who recognizes both objective and subjective 'oughts'] in the first place: that we cannot guide our behavior by norms about which we are uncertain. It would seem to follow from this that someone who is uncertain 'all the way up' will be unable to guide her behavior by norms at all" (Sepielli, 2018b, p. 792).

[14]My primary focus in this paper is on the regress problem as it presents itself to agents who are unbounded in this sense. Though my primary conclusions apply to bounded agents as well (as I explain shortly), a distinct set of problems related to metanormative decision-making and the enkratic externalist view I propose in §4 arise in the context of bounded rationality (including a version of the well-known regress problems for boundedly rational deliberation discussed by Winter (1975), Elster (1977), Lipman (1991), Smith (1991), and Lin (2014), among others). I discuss these issues in a companion piece (Tarsney, ms).

from normative uncertainty or false belief.

If the Regress Argument concerns unbounded agents, what does it have to do with bounded agents like us? My claim is that, if externalism is true of unbounded agents, then it is true of bounded agents as well: (i) This conditional is intuitively plausible. It would be odd if internalism, which seems to place greater deliberative demands on agents (by requiring them to account for their uncertainty about norms that externalism lets them simply take for granted), were true of bounded agents but not of unbounded agents, when unbounded agents are, if anything, more capable of meeting the deliberative demands that internalism creates. (ii) There is plausibly a sort of limit relationship between bounded and unbounded rationality.[15] Boundedly rational agents like us are doing our best to approximate the choices we *would* make if we were unboundedly rational. Thus, if the bounds on our deliberative capacities are relaxed (e.g., as the cost of a unit of computation in time or other resources goes to zero), our rational choices should eventually tend toward those of an unbounded agent (except in some edge cases, e.g., where reaching the correct conclusion requires performing a computational supertask or where the goal is to truthfully answer the question "Are you an unbounded agent?"). But if internalism were true of us and externalism true of unbounded agents, then this limit relationship would be violated: There would be a qualitative divide between the requirements of rationality that apply to each type of agent that, in many choice situations, no finite augmentation of the boundedly rational agent could overcome. A bounded and an unbounded agent in the same choice situation could be required to choose different options, even when the difference in their deliberative capacities seems entirely irrelevant.

Now, to the premises. P1 is meant to be trivial. It says simply that an option is rationally permissible only if there's some true subjective norm that says it's permissible. A subjective norm is just a set of propositions saying that certain options are and aren't permissible in certain choice situations and describing the normative features of those choice situations in virtue of which particular options are or aren't permissible. So P1 is just an instance of the T-schema: If O is permissible in S, then it's true that O is permissible in S, so there's some true norm (indeed, infinitely many true norms) asserting that O is permissible in S.[16]

P4 should also be relatively uncontroversial. It simply asserts that the consequent of L2 is a *reductio*, i.e., we should not accept the conclusion that an unbounded and epistemically rational agent is never rationally permitted to do anything. I will take it for granted that this is correct.

The pressure points of the Regress Argument are P2 and P3. P2 is not a tautological consequence of the definition of internalism, but rests on a substantive claim about the *motivations* for internalism. What internalism necessarily asserts is that whether an nth-order norm applies to an agent A can depend on A's nth- or higher-order normative beliefs. What motivates this assertion, presumably, is the idea that rational choice must be guided by norms or principles, and that, as Sepielli puts it, "we cannot guide our behavior by norms

[15] Thanks to Owen Cotton-Barratt for this suggestion.

[16] Condition (ii) of P1, requiring that A's beliefs place her in the domain of N, is required for N to not merely authorize but assert the permissibility of O. If N is an external nth-order norm whose scope does not depend on A's beliefs about nth- or higher-order norms, then this condition is trivially satisfied.

about which we are uncertain" (Sepielli, 2018b, p. 792). P2 allows that, on the internalist conception of rationality, I can *sometimes* guide my behavior by nth-order norms of which I am uncertain, but only if I have taken account of that uncertainty—meaning, at minimum, that I accept some higher-order norm that authorizes me to act despite my nth-order uncertainty.

P3 asserts a limited epistemic modesty requirement on normative beliefs—it claims that, when it comes to basic normative principles, there are few if any justified certainties. This claim could be defended by appeal to the common Bayesian regularity assumption that agents should not assign credence 1 or 0 to anything except logical truths and falsehoods. I find the arguments for regularity compelling (for a representative statement of these arguments, see Hájek (2003, pp. 31–2)), but you don't need to accept full-blown regularity to accept P3. First, P3 applies only to normative beliefs, not beliefs in general. And second, it does not require that an agent assign positive credence to *every* logically consistent norm, but merely that at every level of normativity, she should be at least a little uncertain about the permissibility of any given option. This seems plausible simply by reflection on the difficulty of normative theorizing, setting aside more general arguments for regularity. In assessing and assigning probabilities to norms, we have much less to go on than we do, say, in the physical sciences, which are taken to be a paradigm example of a domain in which certainty is unattainable.[17]

P3 can also be substantially weakened, at the cost of strengthening P4. For instance, it could merely claim that, for *most* agents in *most* choice situations, it's *permissible* to be uncertain at every order about the permissibility of each option. We would then have to strengthen P4 to assert that, for most agents in most choice situations, no *rationally permissible* set of credences should put the agent in a position where no option is rationally permissible. We could even allow that it is epistemically irrational to be uncertain at every level of subjective normativity (giving up P3 entirely), and simply hold that it should not be impossible for an agent who is *in fact* this uncertain to satisfy the demands of *practical* rationality—that is, the penalty for general normative uncertainty should not be total practical paralysis.

The premises of the Regress Argument, then, are at least *prima facie* plausible. But the internalist can still lodge objections. I will consider two, based on internalist responses to the regress problem in the recent literature.

3.2 Convergence results

The simplest way to avoid the problems posed by an endless regress is to end the regress, after some limited number of steps. The regress of higher-order norms might have such a happy ending, if the following hypothesis were true:

Convergence For any agent A in any situation S (perhaps excluding a few pathological cases), if A's credences are epistemically rational, then there is some n such that all nth- and higher-order norms in which A has positive credence authorize the same set of options in S.

Convergence would let the internalist escape the Regress Argument by denying P3. But why think that it's true? The most promising argument in this

[17] For more extended defense of epistemic modesty with respect to basic normative principles, see for instance Sepielli (2010, pp. 8–30) and Tarsney (2017, pp. 2-8).

direction comes from Trammell (2021), who shows that convergence is guaranteed under certain strong assumptions: in particular, when for every n, A has positive credence in only finitely many nth-order norms, all of which are complete, cardinal (assigning each option a degree of subjective choiceworthiness on a shared cardinal scale), and "compromising" (meaning that the nth-order subjective choiceworthiness of an option must be strictly between its minimum and maximum degrees of $(n-1)$-order subjective choiceworthiness, unless these are the same).[18] But various natural and widely discussed metanormative theories violate these conditions—e.g., My Favorite Theory is not compromising, and My Favorite Option is at least apparently non-cardinal. And of course many *first*-order normative theories are non-cardinal or incomplete. Finally, it seems plausible than an unbounded agent should have non-zero credence in *infinitely* many nth-order norms that assign an unbounded range of subjective choiceworthiness values to some or all of her options, which can also prevent convergence even when all these norms are cardinal, complete, and compromising.[19] So these results, while interesting and important, do not seem like a general solution to the regress problem.

3.3 Conscious vs. dispositional uncertainty

Andrew Sepielli suggests a different response to the threat of regress.[20] He starts by drawing a distinction between *conscious* and *dispositional* uncertainty. An agent, he claims, may be dispositionally but not consciously uncertain of a norm N. And if she acts on such a norm without considering alternative norms, her act is still in an important sense rationally guided, despite her dispositional uncertainty. If we interpret the Regress Argument as referring to dispositional uncertainty (as I will), this suggests a way of rejecting P2: We might hold that an agent can satisfy the belief conditions that place her options in the domain of an internal nth-order N_i^n, even if she has credence in rival nth-order norms that disagree with N_i^n about which options are permissible, so long as that nth-order uncertainty remains merely dispositional rather than conscious.

To assess this strategy, we need to know exactly what is meant by "conscious"

[18] Under these assumptions, convergence is guaranteed only at transfinite levels of the metanormative hierarchy. To guarantee convergence at finite levels, further substantial assumptions are needed.

[19] As a toy example: Suppose I face two options, O_1 and O_2. I am certain of the objective choiceworthiness of O_1, while my credences concerning the objective choiceworthiness of O_2 follow some distribution D that is unbounded both above and below. At every level n, I assign credence only to nth-order norms of the form $\mathrm{CW}_n(O) = f^{-1}[\mathbb{E}(f(\mathrm{CW}_{n-1}(O)))]$, where CW_n is the nth-other subjective choiceworthiness of O, $\mathbb{E}$ is the expectation operator, and $f : \mathbb{R} \mapsto \mathbb{R}$ is strictly increasing and bounded (both above and below). Any such norm is cardinal, complete, and compromising. But if $\mathrm{CW}_{n-1}(O)$ is unbounded above and below, then for any $x \in \mathbb{R}$, there is some f of this form that will yield $\mathrm{CW}_n(O) = x$. This means that, depending on the agent's credence distribution over subjective norms, the range of possible subjective choiceworthiness values for O_2 need not converge. It is possible, for instance, that the choiceworthiness distribution of O_2 simply stays fixed as we ascend the metanormative hierarchy, i.e., that for every n, my credences concerning the nth-order subjective choiceworthiness of O_2 have distribution D. In this case, of course, there is also no convergence with respect to the rationality of O_1 and O_2: at each order, some norms in which I have credence will claim that O_1 is more subjectively choiceworthy and therefore authorize O_1 but not O_2, while others will say the opposite.

[20] This approach is spelled out at greatest length in Sepielli (2014b), but see also Sepielli (2012, pp. 52ff) and Sepielli (2018b, p. 793).

and "dispositional" uncertainty. Here is how Sepielli explains the distinction.

> I think we need to distinguish between two types of uncertainty. The first is dispositional, not necessarily conscious, the sort of attitude I have towards any claim I wouldn't bet my life on. The second is conscious, *involving a feeling of directionlessness, the kind that appears when I deliberate, and disappears when I'm "in the zone"* [emphasis added]. I am uncertain in only the first sense about what the strings on a guitar are; I am uncertain in both senses about what the strings on a banjo are. That is why I can simply play an A7 on a guitar, but can play an A7 on a banjo only *by trying*. (Sepielli, 2014b, p. 91)

As Sepielli concedes, however, it is unclear why the absence of *this* sort of conscious uncertainty ("a feeling of directionlessness") should make it permissible to act on a norm N straightaway, without considering the possibility that N might be mistaken. He writes that "the waning of conscious uncertainty is only a solution to the *psychological* problem of how we can act without [taking unguided leaps of faith]. It's not a solution to the *normative* problem of how we can manage moral risks non-recklessly" (pp. 91-2).

This leads Sepielli to a moderately pessimistic conclusion:

> I think the right thing to say is that meta-rules offer us a normative advantage by *forestalling* moral recklessness, rather than by eliminating it entirely. More precisely, there is a sense in which it is better to leap [i.e., "take a leap of faith" by acting on a norm N the truth of which is uncertain] in the face of uncertainty about meta-rules than to leap in the face of uncertainty about ordinary moral rules, better still to leap in the face of uncertainty about meta-meta-rules, and so on. (Sepielli, 2014b, p. 92)

As I understand him, Sepielli's conclusion is that practical agents can never fully satisfy the demands of rationality. This has some *prima facie* plausibility with respect to bounded agents (which seems to be what Sepielli has in mind), but it is much less plausible with respect to unbounded agents. And in any case, it is the sort of conclusion we should adopt only if we are forced into it—which, as I will shortly argue, we are not.[21]

Neither convergence results nor the distinction between conscious and dispositional uncertainty seem to rescue internalism from the threat of regress. I conclude, therefore, that the Regress Argument gives us persuasive—though certainly still far from conclusive—reason to accept externalism.[22] In the next section, we will see where this leaves us vis-à-vis normative uncertainty.

[21] For further discussion of Sepielli's view, see Riedener (2015, pp. 25–30).

[22] Spelling out the regress problem as we have in this section helps us identify several escape routes for the internalist that are not obvious at first glance. In addition to the two we have considered—denying P3 in order to achieve convergence or denying P2 by allowing that agents may act in the face of merely dispositional uncertainty—there are at least four other possibilities: First, we could deny P2 by proposing some threshold less than certainty at which an agent may permissibly choose an option O based on her nth-order normative beliefs: e.g., a "Lockean threshold" for full belief or a requirement that the probability assigned to nth-order norms that deauthorize O be "*de minimis*" or "rationally negligible" (Smith, 2014). Second, we could deny P3 by holding that (i) agents are rationally required to assign probability 1 to all subjective normative truths and (ii) agents who violate this requirement of epistemic

4 Enkratic externalism

On pain of regress, it appears, we must conclude that at least one norm has belief-independent force, such that an unbounded agent is permitted (if not required) to act as that norm dictates, even if she assigns positive credence to conflicting norms. There are, of course, many norms of various orders to which we could attribute this external status. But I will propose that we should attribute belief-independent normative force to just a single norm of practical rationality: the *enkratic principle* (EP), correctly formulated. Let's call this view *enkratic externalism*. To make this view more determinate, I will focus on the hypothesis that the correct, general formulation of the enkratic principle is the principle of *maximizing expected choiceworthiness* (MEC).

By conceding the conclusion of the Regress Argument, enkratic externalism allows us to accept its premises without paradox. Whether this is enough to resolve the regress problem is a further question, to which we will return in §5. But on face it seems promising: MEC is a second-order norm—it is sensitive to an agent's objective normative beliefs (her beliefs about choiceworthiness), but not to her subjective normative beliefs (her beliefs about rational requirements).[23] If an agent is rationally permitted—indeed, required—to maximize expected choiceworthiness, even when she is uncertain of MEC, then the regress of higher-order norms simply stops at the second order. She can rationally act on MEC without considering conflicting second-order norms, so there is no need to resort to higher-order metanorms.

Because enkratic externalism makes rational requirements sensitive to an agent's beliefs about objective norms but not subjective norms, it may seem like an uncomfortable halfway house between internalism on the one hand and first-order externalism on the other. In this section, however, I will articulate what I take to be principled motivations for enkratic externalism—a simple and plausible picture of practical rationality underlying the view, which consistently

rationality will be unable to satisfy the requirements of practical rationality. (Claim (i) bears some resemble to the "Fixed Point Thesis" defended in Titelbaum (2015), although Titelbaum only claims that rationality prohibits *false belief* about the requirements of rationality, not that it prohibits *any positive credence* in false norms of rationality.) Third, we could deny P4 and hold that even unbounded agents cannot fully satisfy the demands of rationality, unless they are endowed with a degree of normative omniscience that lets them escape the uncertainty demanded by P3. Fourth, we could simply deny the assumption that there is a univocal notion of rational requirement (or subjective ought) that identifies appropriate responses to my belief state as a whole. This would probably require us to understand the debate between internalism and externalism very differently than I have in this paper. For instance, we might hold that when an agent is uncertain about norms of every order, the true nth-order norm nevertheless applies to her and determines what she nth-order-ought to do, but the true $(n+1)$-order norm *also* applies to her and determines what she $(n+1)$-order ought to do, and these norms can give conflicting prescriptions that are not resolved by any all-things-considered norm. (Sepielli (2014a) might be understood as describing a view of this kind, though one could also identify his notion of "global systemic rationality" with rationality simpliciter and interpret him as denying P4.) I don't find these responses particularly promising, but I won't try to evaluate them here.

[23]Second-order norms, as defined in §2.3, can be sensitive to an agent's beliefs about objective norms, her beliefs about first-order subjective norms, or both. But MEC in particular is sensitive only to an agent's beliefs about objective norms. As characterized in §2.2, objective norms assess options in terms of choiceworthiness (i.e., degree of objective reason) while subjective norms assess options in terms of rationality and subjective reasons. Thus, varying an agent's beliefs about objective norms in isolation (holding all other features of a choice situation fixed) can change the expected choiceworthiness of her options, but varying her beliefs about subjective norms cannot.

justifies its sensitivity to objective normative uncertainty and its insensitivity to subjective normative uncertainty. Secondarily, I will note some compelling considerations against first-order externalism on the one hand and internalism on the other (in addition to the regress problem). Enkratic externalism, I will claim, not only mitigates but wholly avoids these difficulties, meaning that the defects of each "extreme" view do not put any pressure on us to abandon enkratic externalism for the opposite extreme view. My goal will not be to conclusively justify enkratic externalism, but merely to convince you that it is a principled view with plausible motivations—and therefore that the regress problem does not force us to abandon metanormativism and embrace first-order externalism.

4.1 Core commitments

Enkratic externalism has two core commitments. First: We have said that practical rationality consists in responding appropriately to one's belief state. But enkratic externalism asserts more specifically that practical rationality consists in responding appropriately to one's beliefs *about one's objective reasons*—i.e., that these are the beliefs that determine the rational permissibility of one's options. An objective reason, the enkratic externalist asserts, is simply any consideration that counts for or against some option, from a normative point of view (from the viewpoint of an agent deliberating about what to do, as opposed to the viewpoint of an external evaluator). If you believe that an option O has a feature that makes it choiceworthy to some degree, the appropriate response to your beliefs (or, to the world as you believe it to be) is to be commensurately *pro tanto* motivated or disposed to choose O. On the other hand, if you believe that an option O has some feature that does not contribute positively or negatively to its choiceworthiness, it would be an inappropriate response to your beliefs for this to affect your motivations or choice dispositions. Let's say that an agent who responds appropriately to her beliefs about her objective reasons is *enkratic*, and call this first commitment *the enkratic conception of rationality.*

The second core commitment is that *rationality* is conceptually distinct from *choiceworthiness* (i.e., from degree of objective reason). This means that an agent's belief that O is rationally required (prohibited) does not *constitute* a belief that she has an objective reason for (against) choosing O, and that it is possible for an agent's beliefs about objective reasons to vary independently of her beliefs about rationality. Thus, even if it is in fact the case that agents always have objective reason (even, *decisive* objective reason) to be rational, it is possible for an agent to believe otherwise. And so the enkratic conception of rationality as responding appropriately to one's beliefs about objective reasons, which makes rational requirements sensitive to an agent's beliefs about objective reasons, does not make them (directly) sensitive to her beliefs about rationality. For instance, suppose I have false beliefs about the requirements of rationality, and believe with certainty that one option is rationally required and another is rationally prohibited; this need not mean that my *beliefs about my objective reasons* favor the first option over the second on balance, or that the first option is rationally preferable. This is what makes enkratic externalism a second-order view, and an externalist view.

There is a very extensive recent debate about "the normativity of rationality" (usually construed as the question of whether the rationality or irrationality of

an option constitutes a reason for or against it), which I won't wade into here.[24] I will simply note that the enkratic externalist's commitment is extremely modest. It does not, in particular, require us to deny that we *in fact* have objective reason to be rational, or even that facts about rationality *constitute* objective reasons—the claim is merely that rationality and objective reasons are conceptually distinct.[25]

4.2 The enkratic principle

Let's now consider what enkrasia (i.e., "responding appropriately" to one's objective reason beliefs) actually entails. That is, what particular second-order norm should the enkratic externalist endorse?

A fairly standard formulation of the enkratic principle is as follows:

EP1 It is rationally required of any agent A that, if she believes she objectively ought to choose option O, then she chooses O.[26]

If "belief" is understood as certainty, then this principle clearly expresses a minimal commitment of the enkratic conception of rationality: An agent who is certain that she objectively ought to choose O is certain that she has more objective reason to choose it than to choose any other alternative, and would clearly be responding inappropriately to her beliefs about her objective reasons if she chose otherwise.

But if we take this very restricted principle as a starting point for a theory of rational choice, what should we say about cases where an agent is uncertain what she objectively ought to do? If "belief" in EP1 is construed more liberally so that it does not require certainty, then EP1 is simply false: an agent may *believe* that she objectively ought to choose O for relatively weak reasons, but have positive credence that she has very strong reasons to choose some other option instead, such that on balance it is not rational for her to choose O.

As others have suggested (e.g. Wedgwood (2013)), a natural generalization of EP1 is the principle of *maximizing expected choiceworthiness*.

EP2 (MEC) For any agent A, choice situation S, and option O, A is rationally permitted to choose O in S if and only if no option in S has greater expected choiceworthiness than O.

[24] See Kolodny (2005), Broome (2007), Wedgwood (2017b), Kiesewetter (2017), and Lord (2018), among many others.

[25] For my part, I'm inclined to join Kolodny (2005) and others in thinking that facts about rationality don't constitute objective reasons. (We generally have instrumental objective reasons to *cultivate* rational dispositions in ourselves, and might have objective reason to choose rationally in particular cases for this or other instrumental reasons. But the fact that an option is rational or irrational is not *in itself* a reason for or against it.) On the other hand, I am happy to talk of *subjective* reasons (constituted by my credences in potential objective reasons), and to say that an action is rationally required just in case I have decisive *subjective* reason to choose it, and rationally permissible just in case there is no other option I have more subjective reason to choose. But none of these views or ways of speaking are commitments of enkratic externalism *per se*.

[26] Perhaps the most familiar formulation of EP is: "If A believes she ought to φ, then she is rationally required to intend to φ." I immediately substitute what I think is an improved formulation, to avoid distracting complications. I have no strong view on the debate between narrow- and wide-scope formulations of principles of practical rationality, but adopt the wide-scope formulation simply because it's weaker. I omit the usual reference to intentions in the consequent of EP for reasons described in Reisner (2013).

So that we have a fully specified version of enkratic externalism to evaluate as a response to the regress problem, I will assume that MEC is in fact the correct generalization of EP1.[27] Needless to say, there is a towering literature on the strengths and weaknesses of expectational decision rules. My aim here is not to advance the debate over whether expectational reasoning is the correct response to risk, but rather to see whether a second-order norm like MEC can satisfactorily resolve the metanormative regress problem. Thus, MEC serves as simply a plain-vanilla example of what a suitably general version of EP (and hence of enkratic externalism) might look like. Substituting a principle that, for instance, permits a wider range of risk attitudes or instructs agents to ignore *de minimis* probabilities would not substantially affect the following discussion.

Nevertheless, it's worth briefly considering some possible motivations for the claim that MEC correctly precisifies the idea of "responding appropriately to one's beliefs about objective reasons". In particular, we should confirm that it's possible to motivate a principle like MEC as a basic, belief-independent requirement of rationality, without simultaneously motivating either of the alternative pictures of subjective norms offered by internalism and first-order externalism.

First, then, we might appeal to a sort of motivational internalism about reason beliefs. On this view, part of what it is to *really, consistently* believe that you have an objective reason to choose option O (i.e., that O is choiceworthy to some degree) is to have some motivation or disposition toward choosing O. More precisely, part of what it is to really, consistently believe, with some probability p, that you have an objective reason of strength x to choose O is to be motivated to choose O, to a degree proportionate to both p and x. This would make enkratic rationality a kind of coherence: an agent who does not maximize expected choiceworthiness is incoherent in the sense that she holds certain beliefs incompletely or inconsistently. For instance, perhaps the part of her that engages in conscious verbal reasoning and produces sincere affirmations manifests a belief that O is choiceworthy to degree x, but her motivational system, the part of her that determines her choices and actions, does not. By contrast, suppose an agent believes that an option has some empirical feature (say, it will bring additional happy lives into existence) which she does not believe gives rise to objective reasons, although in fact it does. Or suppose she believes that O is

[27] For an extended defence of MEC in the spirit of what I have called "the enkratic conception of rationality", see Wedgwood (2017a). MEC is also defended in MacAskill (2014), Lazar (2017), and MacAskill and Ord (2020), though they don't associate it with EP. And I take Broome to endorse MEC, or something very much like it, e.g. in Broome (1991) and (2013). Broome (2013) defends a version of EP he calls *Enkrasia* that, apart from some complications that aren't relevant for our purposes, resembles a wide-scope version of the standard principle: Rationality requires that, if an agent believes she ought to φ, then she intends to φ (p. 170). But the "ought" Broome has in mind is "prospective" rather than objective, i.e., depends on the prospects of the options in a given choice situation (Ch. 3). And Broome says that "the value of a prospect is an expected value of some sort" (p. 41). As far as I can see, this makes Broome's Enkrasia a version of MEC.

For my own part, I am inclined to favor not MEC but a formulation of EP in terms of *stochastic dominance*, holding that O is rationally prohibited iff there is another option P such that (i) for any degree of choiceworthiness, P is at least as likely as O to be at least that choiceworthy and (ii) for some degree of choiceworthiness, P is strictly more likely to be at least that choiceworthy. But I favor this principle largely because I believe that, under normal epistemic circumstances, it is in surprisingly close agreement with MEC (while better handling some standard problem cases for expectational decision theory). These arguments are too involved to reproduce here (but are laid out in Tarsney (2018)). So for simplicity, I will focus in this paper on the more familiar MEC.

uniquely rational, but does not believe that she has any objective reason to be rational. In either case, she does not manifest any incoherence—or at least, the same sort of incoherence—if she is unmotivated by these beliefs.

A second, perhaps more modest, picture holds that enkratic rationality is a matter of "fittingness". Specifically, while objective reason beliefs need not themselves *involve* any motivational element, the fitting motivational *response* to a belief of degree p that O is choiceworthy to degree x is a motivation to choose O that is proportionate to both p and x. Thus, an agent whose motivations are fitting with respect to her reason beliefs will maximize expected choiceworthiness. By contrast, again, consider an agent with some false belief about objective norms, e.g., one who falsely believes that she has no objective reason to bring additional happy lives into existence. While her lack of motivation to create more happy lives is, in some sense, an unfitting response to *the way things actually are, morally speaking*, it's not an unfitting response to the way she believes them to be. If she merits some negative evaluation, it's not for failing to respond appropriately to her belief state as a whole. And the same goes for an agent who believes that some option is rationally required or prohibited, but believes (rightly or wrongly) that she has no objective reason to be rational, and so is unmoved by the first belief.

Finally, one might try to justify MEC as a belief-independent requirement of rationality by axiomatic arguments, along the lines of the standard axiomatic justifications for expected utility theory. On this view, cardinal choiceworthiness is not a primitive feature of the strength of objective reasons, but rather a device for representing an ordinal ranking of options that satisfies certain axioms.[28] These axioms might be seen simply as coherence requirements. But another interpretation of the axiomatic approach is that rationality has a distinctive sort of *instrumental* value, namely, protecting agents against sure losses. Agents who violate the standard axioms of expected utility theory are, arguably, vulnerable to sure losses through devices like Dutch books and money pumps.[29] The further requirement that your utility function be an increasing function of the strength of your objective reasons can be justified on the same grounds: to choose an option that you're certain you have less objective reason to choose over one that you're certain you have more objective reason to choose, in violation of EP1, is the simplest possible form of sure loss. On the other hand, choosing an option

[28]There are at least two ways this sort of axiomatic rationale for MEC might go. One is via aggregation theorems, analogous to Harsanyi's famous aggregation theorem for social welfare, but with first-order norms taking the place of individual welfare subjects. Riedener (2020) gives an aggregation theorem of this kind, though his focus is on axiological uncertainty, and further difficult questions arise when we try to implement this approach in the broader context of normative uncertainty. Another approach is to stick with the simpler formalism of individual expected utility theory (e.g., Savage (1954)), while understanding states and outcomes as individuated not only by the physical state of the world but also by the objective normative facts, and imposing substantive "ethical" requirements on preferences (e.g., that for any two outcomes in which the same objective norm is true, the agent is rationally required to prefer whichever outcome is more choiceworthy according to that norm). But this approach also has its difficulties. For instance, Savage's formalism requires agents to consider and evaluate "constant acts" that yield the same outcome in every possible state of the world; but if both states and outcomes are individuated, among other things, by the moral facts, then this means evaluating apparently-impossible acts that "change the fundamental moral facts", mapping a state of the world where one moral theory holds to an outcome where a different moral theory holds.

[29]See for instance Vineberg (2022) on Dutch books and Gustafsson (forthcoming) on money pumps.

that you mistakenly believe is best supported by objective reasons (for instance, because you have false moral beliefs) is a loss, but not a sure loss from your own perspective. And similarly, choosing an option that you believe is irrational need not be a sure loss from your perspective, if you do not see irrationality as an objective reason against that option, or if you believe it to be outweighed by competing reasons.

I don't pretend that any of this is conclusive. But I have tried to show that enkratic externalism (and MEC in particular) represents at least a principled and stable picture of rationality, for which clear motivations can be given, and not merely an uncomfortable compromise between two more principled extremes.

4.3 Against internalism

In this subsection and the next, I will argue that enkratic externalism avoids the most important defects of internalism and first-order externalism respectively.

A central advantage of enkratic externalism over internalism for our purposes, of course, is its ability to avoid the regress problem, to which we will return in the next section. But there is another, simpler argument against internalism: Whatever the correct general theory of rationality turns out to be, it will yield *some* criterion of the form $\Box\forall x(\mathit{Rational}(x) \leftrightarrow \varphi(x))$ (where x might range over agents, attitudes, options, or something else). Simply by virtue of being a *general* criterion for rationality, this criterion will apply to all agents regardless of their normative beliefs—in particular, its application is not restricted by the agent's beliefs about subjective norms of any order, and therefore apparently it must be an external rather than an internal norm.[30]

A bit more carefully: The motivating idea of internalism is that agents cannot be rationally bound by norms or prescriptions that they disbelieve. But the true general theory of rationality, by virtue of applying to all agents in all choice situations, must apparently apply even to agents who disbelieve its prescriptions. The only way this could fail to be the case is if it's *impossible* for agents to disbelieve its prescriptions—that is, impossible for agents to have false beliefs about the rational requirements that apply to them.

It is not inconceivable that false belief about the rationality of one's options is impossible. This would be the case, for instance, if the true theory of rationality were: "It's rationally permissible for A to choose O if and only if A believes that it's rationally permissible to choose O." But, the argument claims, this sort of infallibility is implausible, and therefore internalism is implausible as well.

Here is one way of precisifying the argument:

[30] Arguments in this spirit are made by Broome (2013, p. 93), Bykvist (2013, p. 133), and Weatherson (2014, pp. 156–7). Weatherson puts the point as follows: "There is a worry that externalism is not sufficiently action guiding, and can't be a norm that agents can live by. But any philosophical theory whatsoever is going to have to say something about how to judge agents who ascribe some credence to a rival theory. That's true whether the theory is the first-order theory that Jeremy Bentham offers, or the second-order theory that Andrew Sepielli offers. Once you're in the business of theorising at all, you're going to impose an external standard on an agent, one that an agent may, in good faith and something like good conscience, sincerely reject. The externalist says that it's better to have that standard be one concerned with what is genuinely valuable in the world, rather than a technical standard about resolving moral uncertainty. But every theorist has to be a little bit externalist; the objector who searches for a thoroughly subjective standard is going to end up like Ponce de Leon."

The Argument from Fallibility

P1. If internalism is true, then there is some probability threshold t such that any true subjective norm authorizing an option O has O in its domain only when the agent's credence that O is permissible is at least t; and likewise, any true subjective norm deauthorizing O has O in its domain only when the agent's credence that O is impermissible is at least t.

P2. It is possible for agents to have false beliefs about the rationality of their options—in particular, to believe with probability greater than $1 - t$ that O is rationally permissible when it is in fact impermissible, or that O is rationally impermissible when it is in fact permissible.

L1. There is a true subjective norm that asserts the permissibility of an option O in a situation S (meaning that it both authorizes O and has O in its domain) even though the agent's credence that O is permissible is less than t—namely, norms describing the cases of false belief about rationality described in P2. And likewise, there is a true subjective norm that asserts the impermissibility of O (deauthorizes O and has O in its domain) when the agent's credence that O is impermissible is less than t. [from P2]

C. Internalism is false. [from P1, L1]

P1, it seems to me, expresses the core motivation of internalism—it is hard to see what would motivate a version of internalism that denied it. Why should we believe P2? Briefly, suppose for the sake of argument that $t = 0.5$—that is, subjective norms only apply to an agent when her credence in their prescriptions is at least 0.5. First, if I know this fact about rationality (that it's impossible to be required to do something while believing that I am probably not required), I should be able to reason from the fact that my credence that I am required to choose O is, say, 0.47 to *certainty* that I am not required to choose O. But this sort of reasoning seems suspect, to say the least.[31] Second, the impossibility of false belief about rationality seems to make questions of rationality vacuous: If the rationality of any option O is simply determined by my beliefs about the rationality of O (necessarily, O is required iff I believe it's required, and permitted iff I believe it's permitted), then what are beliefs about rational requirement actually *about*? They seem to have no truth conditions except themselves.[32]

The Argument from Fallibility, as much as the Regress Argument, convinces me that we must accept externalism. Note also that it applies directly to both bounded and unbounded agents, so its relevance to bounded agents like us does not depend on any conditional linkage between the requirements of bounded and unbounded rationality. (Indeed, it is more applicable to bounded agents, since it is more obviously possible for bounded agents to have false beliefs about rationality.)

[31] For a related observation in the context of morality rather than rationality, see the discussion of ProbWrong in Weatherson (2014, p. 146).

[32] The same worry arises, to only a slightly lesser degree, if the threshold t is lower. At the extreme, suppose the internalist only asserts that an agent cannot be rationally permitted to choose an option that she is certain is rationally impermissible, or prohibited from choosing an option that she is certain is permissible. Then certainty in rational permissions or prohibitions is self-fulfilling and self-justifying. If I become certain that an option O is rationally permissible (or impermissible), no matter how irrationally, my belief is guaranteed to be correct, and I am justified in maintaining it in the face of any new evidence or arguments.

But most importantly for present purposes, while the Argument from Fallibility favors externalist views over internalism, it does not favor first-order externalism over enkratic externalism. Enkratic externalism, though it is a metanormativist view that makes rational requirements sensitive to some of an agent's normative beliefs and uncertainties, is entirely compatible with agents having false beliefs about the rational requirements that apply to them.

4.4 Against first-order externalism

The debate between metanormativism and first-order externalism has generated a substantial recent literature, with many arguments on either side that I won't try to survey here.[33] But the most compelling argument against first-order externalism, it seems to me, is that it can require an agent to choose an option that she is *certain* is objectively worse than some available alternative. One obvious way in which this can happen is for an agent to be certain of a false normative theory. For instance, if subjective hedonistic utilitarianism is the true external first-order norm, then an agent who is justifiably certain that she has most objective reason to act in ways favoring the interests of her friends and family is rationally *prohibited* from doing what she is certain she has most objective reason to do. In this case, first-order externalism contradicts any version of the enkratic principle, including the extremely modest reading of EP1 that asserts only that agents are rationally required, if they are certain that O is more choiceworthy than any alternative, to choose O. So whatever you think of stronger principles like MEC, if you find *any* version of EP compelling, then you have reason to reject first-order externalism.

You might not find this objection particularly persuasive, though, since it concerns cases of *mistaken* certainty. Perhaps an agent can never be *justifiably* or *rationally* certain of a false normative theory, and it is not so implausible that we can be rationally required to act against our *irrational* certainties.

A stronger objection to first-order externalism is that it can require an agent to choose an option that she is *correctly* certain is objectively worse than some available alternative. Consider the following case, due to Podgorski (2020): You have equal credence in two first-order normative theories, T_1 and T_2 (one of which is in fact true), and face a choice between a "safe" option (which results in a sure payoff of 0) and an uncertain option that will have either outcome o_1 or outcome o_2, with equal probability. According to T_1, an option with outcome o_1 has a choiceworthiness of 1 and an option with outcome o_2 has a choiceworthiness of -2, while according to T_2, the values are reversed: o_1 yields a choiceworthiness of -2 and o_2 yields a choiceworthiness of 1. Evaluated by the lights of either first-order theory, therefore, the uncertain option has an expectation of -0.5. So whichever theory is correct, first-order externalism requires you to choose the safe option. But now suppose there is a probabilistic dependence between your normative and empirical credences: T_1 is true iff the uncertain option will have outcome o_1, and T_2 is true iff the uncertain option will have outcome o_2. This guarantees that the uncertain option has a choiceworthiness of 1, and is more choiceworthy than the safe option. So first-

[33]For defences of first-order externalism, see Weatherson (2014, 2019), Harman (2015), and Hedden (2016). For metanormativist replies, see Sepielli (2016, 2018a), Johnson-King (2018), and MacAskill and Ord (2020), among others. I give my own defense of metanormativism and reply to the first-order externalists in Tarsney (2017, Chs 2–3).

order externalism requires you to choose an option that you are *correctly* certain is objectively worse than the available alternative.

To my mind, this objection to first-order externalism is all but decisive. (See Podgorski (2020) for discussion of various possible first-order externalist replies.) But, crucially, it does not tell against externalism generally, or enkratic externalism in particular. In Podgorski's case, there are two possibilities with non-zero probability: T_1 is true and the uncertain option will have outcome o_1, or T_2 is true and the uncertain option will have outcome o_2. In either case, the uncertain option has a choiceworthiness of 1 and the safe option has a choiceworthiness of 0, so the uncertain option has an *expected* choiceworthiness of 1 and the safe option an expected choiceworthiness of 0. So MEC correctly recommends the uncertain option. More generally, MEC will always satisfy dominance with respect to choiceworthiness—if the choiceworthiness of O_1 is certainly at least as great as that of O_2, and possibility greater, then its expected choiceworthiness must be greater.[34]

Thus, as with internalism, there is a compelling objection to first-order externalism that does not seem to indict enkratic externalism or put any pressure on us to abandon it in favor of the opposite "extreme" view.

5 Normalization uncertainty

Although I have framed the regress problem as an argument for normative externalism, that doesn't mean that conceding any form of externalism is enough to make the problem go away. I suggested above that, since EP is a second-order norm, attributing belief-independent force to EP stops the regress at the second order: Since we are permitted to act on EP despite our second-order uncertainties, we are not required to consider third- or higher-order norms. But does EP alone allow us to set aside all our second-order uncertainties? That depends on whether EP is a *comprehensive* norm that settles all questions about rational permissibility, or a partial norm that simply *constrains the field* of subjective norms. If the latter, then the regress problem remains unresolved: for even if an agent is rationally required to comply with EP, this alone may not be enough to tell her what to do. And if all other norms besides EP have merely belief-dependent force, then she may have to go on the same fruitless quest up the hierarchy of metanorms that pure internalism condemns her to.

[34]Could enkratic externalism be vulnerable to a different but analogous objection, that it recommends options that are certainly *irrational*? This would require that, just as we can be uncertain about which of two options is more objectively choiceworthy according to a given objective norm (e.g., which of two acts would produce greater total happiness), we can be uncertain which of two options is rationally required *according to a given subjective norm*—and in particular, according to MEC. Clearly an ideal agent (as characterized in §3.1 above) cannot be uncertain about the expected choiceworthiness of her options, or therefore their rationality according to MEC, since their expected choiceworthiness just depends on her beliefs, to which she has perfect introspective access and from which she can effortlessly calculate expectations. Whether a bounded agent can be uncertain about the expected choiceworthiness of her options depends on how we understand MEC in the context of bounded rationality. In Tarsney, I argue that the "expected choiceworthiness" of an option for a bounded agent should be understood subjectively, as a sort of "best estimate" produced by the agent in deliberation and to which she has immediate introspective access, rather than a quantity determined by her entire belief state that can only be ascertained by elaborate calculation. If this is right, then even bounded agents can be certain which of their options has greatest expected choiceworthiness.

There is, moreover, an important way in which MEC might fail to be comprehensive: To determine the expected choiceworthiness of an option under normative uncertainty, we must first *normalize* the choiceworthiness scales of rival objective norms. That is, we need to say what increment on the choiceworthiness scale of one normative theory (e.g., how many units of total welfare) corresponds to a given increment on the choiceworthiness scale of another theory (e.g., a given unit of average welfare). And even if she is fully committed to maximizing expected choiceworthiness, an agent may be uncertain how to normalize rival theories. She might assign some credence to various statistical normalization methods (e.g., normalizing theories at the range (Lockhart, 2000) or variance (MacAskill, 2014) of their choiceworthiness assignments), some to "content-based" normalization methods (Tarsney, 2017), and some to the view that the correct normalization is simply brute (a possibility entertained, though not endorsed, by Riedener (2020)), conditional on which she distributes her credence widely over a range of plausible normalizations. This suggests that MEC is not truly comprehensive: MEC says that I should maximize *an* expectation of the various choiceworthiness hypotheses to which I assign positive credence, but *which* expectation—relative to which normalization? An agent committed to MEC might therefore be uncertain between several second-order norms: the combination of MEC with each normalization method in which she has positive credence. And if she is uncertain how to respond to that uncertainty, the regress is back in business.

One response for the enkratic externalist is simply to go externalist about the normalization method: All agents, regardless of their beliefs, are rationally required not just to maximize expected choiceworthiness but to normalize theories in a particular way (say, by variance normalization) and *then* maximize expected choiceworthiness. But while the requirement to maximize expected choiceworthiness can plausibly be teased out of the enkratic conception of rationality as responding appropriately to one's objective reason beliefs (as we saw in the last section), it is much less plausible that any particular intertheoretic normalization rule is somehow inherent to the very idea of rationality. Moreover, the question of how to make intertheoretic choiceworthiness comparisons looks like a question about the relative strength of objective reasons, and if we understand rationality as responding appropriately to your objective reason beliefs, then rational requirements should be sensitive to the agent's beliefs about this sort of question—that is, it's not the sort of question about which the enkratic externalist should go externalist.

Here is a more promising response: Normalization uncertainty is not a challenge to MEC, but just another instance of exactly the sort of uncertainty that MEC is meant to deal with. If an agent is committed to MEC but uncertain how to normalize the theories in which she has positive credence, she should simply take a probability-weighted average over the various possible normalizations—i.e., an expectation. For instance, suppose she is uncertain between classical utilitarianism and a simple sufficientarian theory that assigns twice as much weight to the interests of those below a certain welfare threshold as to those above it. She might be uncertain whether the sufficientarian theory cares *more* about the worse-off than classical utilitarianism, or cares *less* about the better-off, and hence be uncertain how to normalize the theories. But all she needs to do to apply MEC under these circumstances is to compute expectations using each normalization, and then take a probability-weighted average of the results.

On face, this response looks naive and inadequate: To "take an average" over several possible normalization methods, the agent must be able to *normalize the outputs* of those normalization methods—i.e., to normalize the expectations that she calculates using each normalization method. It seems that she might be uncertain how to do this. And so we are just still headed for a regress.

But I think—albeit tentatively—that this problem can be overcome, and that the "average the normalizations" response is ultimately on the right track. The key to making this response work is a particular way of thinking about the problem of intertheoretic normalization, which MacAskill (2014) has dubbed the "universal scale" approach. On this view, there is a determinate set of choiceworthiness properties, which all normative theories are in the business of assigning to options. Theories differ in *which* properties they assign to which options, but their assignments share the same codomain. On this view, uncertainty about how to normalize two theories—say, classical utilitarianism and sufficientarianism—is to be understood as uncertainty between various "amplifications" (i.e., rescalings) of one or both theories. For instance, I might assign credence only to a single version of classical utilitarianism, while distributing my credence between two versions of the sufficientarian view, one that assigns the same weight as classical utilitarianism to the worse-off and another that assigns the same weight as classical utilitarianism to the better-off (MacAskill, 2014, pp. 134–142). If my uncertainty about how to normalize the two theories can be understood in this way, then there is no further obstacle to applying MEC. In effect, we simply distinguish theories at a more fine-grained level so that instead of two theories that I'm uncertain how to normalize, I have *three* theories that I'm *certain* how to normalize.

I admit that I am not entirely sure of this response. In particular, I don't find MacAskill's metaphysical defense of the universal scale approach entirely convincing.[35] Still, I suspect that something like MacAskill's story is ultimately correct. Roughly, all objective normative theories are theories of the same concept (viz., choiceworthiness), and as such it is more plausible to think of them as drawing from the same set of choiceworthiness properties and simply disagreeing about which properties to assign to which option than to think of each theory as inventing its own set of choiceworthiness properties out of whole cloth, entirely disjoint from the properties mentioned by any other theory. This allows for uncertainty about how to normalize pairs of theories like classical utilitarianism and sufficientarianism—it simply casts that uncertainty in a way that is tractable to expectational reasoning. A fully worked-out defense of this view would be (at least) a paper in its own right. But the idea that normalization uncertainty can be understood in this way is *prima facie* plausible and, to my knowledge, there is no compelling argument *against* it. So I conclude that, although there is an important residual worry here, the expectational version of enkratic externalism has reasonable prospects for avoiding a vicious regress brought on by normalization uncertainty.

[35]For MacAskill's defense, see MacAskill (2014, pp. 149–157). For my reservations, see Tarsney (2017, pp. 209–212). See also Riedener (2015, Ch. 3) and Carr (2020), who defend theses in a similar spirit to MacAskill's that would also serve my purposes in this section.

6 Conclusion

I had two objectives in this paper: first, to develop and defend the regress argument for normative externalism, but second, to defend a metanormativist rather than a first-order version of externalism as the best response to the threat of regress. Specifically, I have argued that we need attribute belief-independent force to only one norm: the enkratic principle, in some comprehensive formulation that covers choices under uncertainty. Enkratic externalism solves the metanormative regress problem by providing a comprehensive second-order norm that obviates any need to consider the full hierarchy of higher-order norms.

With respect to both objectives, it seems to me that the preceding arguments are persuasive but hardly conclusive, and leave a great deal more to be said. Future research might usefully explore (i) alternative characterizations of basic concepts like normative internalism, externalism and metanorms, (ii) internalist responses to the regress problem (e.g. generalizing Trammell's convergence results, or exploring one of the escape routes noted in footnote 22), (iii) other forms of metanormativist externalism (based on different versions of the enkratic principle or entirely different second- or higher-order metanorms), or (iv) the challenge posed by normalization uncertainty for expectational metanorms like MEC.

References

J. Broome. Desire, belief and expectation. *Mind*, 100(2):265–267, 1991.

J. Broome. Is rationality normative? *Disputatio*, 2(23):161–178, 2007. doi: 10.2478/disp-2007-0008.

J. Broome. *Rationality Through Reasoning*. Wiley-Blackwell, 2013.

K. Bykvist. Evaluative uncertainty, environmental ethics, and consequentialism. In A. Hiller, R. Ilea, and L. Kahn, editors, *Consequentialism and Environmental Ethics*. Routledge, 2013.

J. R. Carr. Normative uncertainty without theories. 98(4):747–762, 2020. doi: 10.1080/00048402.2019.1697710.

A. Elga. How to disagree about how to disagree. In R. Feldman and T. A. Warfield, editors, *Disagreement*, pages 175–186. Oxford University Press, Oxford, 2010.

J. Elster. Ulysses and the Sirens: A theory of imperfect rationality. *Information (International Social Science Council)*, 16(5):469–526, 1977.

E. J. Gracely. On the noncomparability of judgments made by different ethical theories. *Metaphilosophy*, 27(3):327–332, 1996.

H. Greaves and O. Cotton-Barratt. A bargaining-theoretic approach to moral uncertainty. Unpublished ms, August 2019.

A. A. Guerrero. Don't know, don't kill: Moral ignorance, culpability, and caution. *Philosophical Studies*, 136(1):59–97, 2007.

J. E. Gustafsson. *Money-Pump Arguments*. Cambridge, UK: Cambridge University Press, forthcoming. URL `https://philpapers.org/rec/GUSMA`.

J. E. Gustafsson and O. Torpman. In defence of My Favourite Theory. *Pacific Philosophical Quarterly*, 95(2):159–174, 2014.

A. Hájek. Waging war on Pascal's wager. *Philosophical Review*, 112(1):27–56, 2003.

E. Harman. The irrelevance of moral uncertainty. In R. Shafer-Landau, editor, *Oxford Studies in Metaethics*, volume 10. Oxford University Press, 2015.

B. Hedden. Does MITE make right? On decision-making under normative uncertainty. In R. Shafer-Landau, editor, *Oxford Studies in Metaethics*, volume 11. Oxford University Press, 2016.

G. Hellman. Against 'absolutely everything'! In A. Rayo and G. Uzquiano, editors, *Absolute Generality*, pages 75–97. Oxford University Press, Oxford, 2006.

Z. A. Johnson-King. *Trying to Act Rightly*. PhD thesis, University of Michigan, 2018. URL: https://zoe-johnsonking-blhb.squarespace.com/s/Trying-to-Act-Rightly-Doctoral-Dissertation-Zoe-Johnson-King-Rackham-formatting.pdf.

B. Kiesewetter. *The Normativity of Rationality*. Oxford, UK: Oxford University Press, 2017.

N. Kolodny. Why be rational? *Mind*, 114(455):509–563, 2005.

S. Lazar. Deontological decision theory and agent-centered options. *Ethics*, 127 (3):579–609, 2017.

H. Lin. On the regress problem of deciding how to decide. *Synthese*, 191(4): 661–670, 2014.

B. L. Lipman. How to decide how to decide how to...: Modeling limited rationality. *Econometrica*, 59(4):1105–1125, 1991.

T. Lockhart. *Moral Uncertainty and Its Consequences*. Oxford University Press, 2000.

E. Lord. *The Importance of Being Rational*. Oxford, UK: Oxford University Press, 2018.

W. MacAskill. *Normative Uncertainty*. PhD thesis, University of Oxford, 2014. URL: https://philpapers.org/rec/MACNU.

W. MacAskill. Normative uncertainty as a voting problem. *Mind*, 125(500): 967–1004, 2016.

W. MacAskill and T. Ord. Why maximize expected choice-worthiness? *Noûs*, 54(2):327–353, 2020. doi: 10.1111/nous.12264.

I. Nissan-Rozen. Doing the best one can: A new justification for the use of lotteries. *Erasmus Journal for Philosophy and Economics*, 5(1):45–72, 2012.

A. Podgorski. Normative uncertainty and the dependence problem. *Mind*, 129 (513):43–70, 2020. doi: 10.1093/mind/fzz048.

A. Reisner. Is the enkratic principle a requirement of rationality? *Organon F*, 20(4):436–462, 2013.

S. Riedener. *Maximising Expected Value under Axiological Uncertainty: An Axiomatic Approach*. PhD thesis, University of Oxford, 2015. URL: https://uzh.academia.edu/StefanRiedener.

S. Riedener. An axiomatic approach to axiological uncertainty. *Philosophical Studies*, 177(2):483–504, 2020. doi: 10.1007/s11098-018-1191-7.

J. Ross. *Acceptance and Practical Reason*. PhD thesis, Rutgers University – New Brunswick, 2006. URL: https://philpapers.org/rec/ROSAAP-2.

L. J. Savage. *The Foundations of Statistics*. John Wiley & Sons, New York, 1st edition, 1954.

A. Sepielli. What to do when you don't know what to do. *Oxford Studies in Metaethics*, 4:5–28, 2009.

A. Sepielli. *'Along an Imperfectly-Lighted Path': Practical Rationality and Normative Uncertainty*. PhD thesis, Rutgers University Graduate School - New Brunswick, 2010. URL: https://philpapers.org/rec/SEPAAI.

A. Sepielli. Subjective normativity and action guidance. In M. Timmons, editor, *Oxford Studies in Normative Ethics, Vol. II*. Oxford University Press, 2012.

A. Sepielli. What to do when you don't know what to do when you don't know what to do.... *Noûs*, 48(3):521–544, 2014a.

A. Sepielli. Should you look before you leap? *The Philosophers' Magazine*, 66: 89–93, 2014b.

A. Sepielli. Moral uncertainty and fetishistic motivation. *Philosophical Studies*, 173(11):2951–2968, 2016.

A. Sepielli. How moral uncertaintism can be both true and interesting. In M. Timmons, editor, *Oxford Studies in Normative Ethics*, volume 7, pages 98–116. Oxford University Press, Oxford, 2018a.

A. Sepielli. Subjective and objective reasons. In D. Star, editor, *The Oxford Handbook of Reasons and Normativity*, pages 784–799. Oxford University Press, Oxford, 2018b.

S. Shapiro and C. Wright. All things indefinitely extensible. In A. Rayo and G. Uzquiano, editors, *Absolute Generality*, pages 255–304. Oxford University Press, Oxford, 2006.

H. Smith. Deciding how to decide: Is there a regress problem? In M. Bacharach and S. Hurley, editors, *Essays in the Foundations of Decision Theory*. Blackwell, 1991.

N. J. J. Smith. Is evaluative compositionality a requirement of rationality? *Mind*, 123(490):457–502, 2014.

C. Tarsney. Maximizing expected choiceworthiness: A guide for bounded agents. Unpublished ms, August 2019.

C. Tarsney. *Rationality and Moral Risk: A Moderate Defense of Hedging*. PhD thesis, University of Maryland, 2017. URL: https://philarchive.org/archive/TARRAM-7.

C. Tarsney. Exceeding expectations: Stochastic dominance as a general decision theory. *arXiv e-prints*, art. arXiv:1807.10895, Jul 2018.

C. Tarsney. Moral uncertainty for deontologists. *Ethical Theory and Moral Practice*, 21(3):505–520, 2018.

C. Tarsney. Normative uncertainty and social choice. *Mind*, 128(512):1285–1308, 2019.

M. G. Titelbaum. Rationality's fixed point (or: In defense of right reason). In T. S. Gendler and J. Hawthorne, editors, *Oxford Studies in Epistemology*, volume 5, pages 253–294. Oxford University Press, Oxford, 2015.

P. Trammell. Fixed-point solutions to the regress problem in normative uncertainty. *Synthese*, 198(2):1177–1199, 2021.

S. Vineberg. Dutch Book Arguments. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2022 edition, 2022.

B. Weatherson. Running risks morally. *Philosophical Studies*, 167(1):141–163, 2014.

B. Weatherson. *Normative Externalism*. Oxford University Press, Oxford, 2019.

R. Wedgwood. Akrasia and uncertainty. *Organon F*, 20(4):484–506, 2013.

R. Wedgwood. Must rational intentions maximize utility? *Philosophical Explorations*, 20(sup2):73–92, 2017a.

R. Wedgwood. *The Value of Rationality*. Oxford, UK: Oxford University Press, 2017b.

S. G. Winter. Optimization and evolution in the theory of the firm. In R. H. Day and T. Groves, editors, *Adaptive Economic Models*, pages 73–118. Elsevier, 1975.