Metanormative Regress: An Escape Plan

Christian Tarsney*

Abstract

How should you decide what to do when you're uncertain about basic normative principles? A natural suggestion is to follow some "second-order" norm: e.g., *obey the most probable norm* 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, any norm-guided response to normative uncertainty appears doomed to a vicious regress. This paper aims to rescue second-order norms from the threat of regress. I first elaborate and defend the claim some philosophers have made that the regress problem forces us to accept *normative externalism*, the view that at least one norm is incumbent on all agents regardless of their normative beliefs. But, I then argue, we need not accept externalism about first-order norms, thus closing off any question of how agents should respond to normative uncertainty. Rather, we can head off the threat of regress by ascribing external force to a single second-order norm: the enkratic principle.

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),

*Population Wellbeing Initiative, University of Texas at Austin.   Email: christian.tarsney@austin.utexas.edu

1

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 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 norms must decide what to do by means of some second-order norms, must not an agent who is uncertain about *second*-order norms decide what to do by means of some *third*-order norms—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 norm 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

---

[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 (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.

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.

The paper therefore has two aims: 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.

§2 introduces some conceptual scaffolding for the rest of the paper. §3 sets out the regress problem as an argument for externalism, considers two internalist responses, and concludes that they are unsatisfactory. §4 sets out my response to the regress problem, which posits a single belief-independent norm of practical rationality: the enkratic principle. §5 summarizes my conclusions.

2   Internalism, externalism, and metanormativism

2.1   Choice situations and norms

A *choice situation* is an ordered triple $S = \langle A, \mathbf{O}, \mathrm{Cr} \rangle$, where A is an agent, $\mathbf{O}$ is a finite set of options $\{O_1, O_2, ..., O_n\}$ available to A in S, and Cr is A'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

---

[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).

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, 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 $O_k^i \in \mathbf{O}_i$ and $O_l^j \in \mathbf{O}_j$ that differ only with respect to that feature, N asserts that O_k^i is permissible in S_i but O_l^j 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 pair of choice situations S_i and S_j that differ only with respect to the agent's credences over propositions in Σ, where N designates some option permissible in S_i while designating the corresponding option impermissible in S_j. In other words, a norm is sensitive to an agent's beliefs 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 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,

---

[4]These notions should be understood very thinly. To say that an option is "permissible" just means that it's possible for an agent who is in some relevant sense normatively ideal (e.g. fully rational and, in the case of objective norms, fully informed) to choose that option.

[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 probability axioms) and that her evidence gives rise to either evidential probabilities or 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)".

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 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 below are meant to require no commitments regarding either the existence of other kinds of subjective norms or the nature 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.

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 don't intend to presuppose this identification.[7]

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

---

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

[7]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).

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.)[8]

2.4 Internal and external subjective norms

Say that an option O in 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 unless she *believes* either the norm itself or at least its particular permissibility judgements in the choice situation she confronts. 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).

---

[8]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.

If an agent is sufficiently confident that MFO is correct and MFT is incorrect, either in general or in her particular situation, she cannot be rationally required to obey MFT (the thought goes). This is so even if MFT is the *true* second-order norm in some sense—e.g., in the sense that sufficient rational deliberation would lead us 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 has sufficient credence (e.g. certainty, or credence greater than 0.5) in either N^n itself or the rational permissibility of O. When only condition (i) is satisfied, N^n remains silent.

Because norms can be internal, it's possible for two apparently competing subjective norms to be true, even though they 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 an agent ought to do, and so can't both be true. But if the first-order norm applies only to agents who believe it 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* is the thesis that there is at least one true external norm N, which applies to agents regardless of their beliefs about N itself. *Internalism* is the thesis that all true norms are internal.[9]

---

[9]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).

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. 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 reasoning capacities (of whom I will say more shortly). I will then argue that if externalism is true of these unbounded agents, it is true of bounded agents as well.

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.[10]

---

[10]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

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 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][11]

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

1 that authorizes O (namely, 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.

[11]This presentation of the regress problem was originally inspired by remarks in Weatherson (2014), though I now take Weatherson to be making a 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).

new evidence, and (v) has perfect introspective/recollective access to her own beliefs and evidence.

Unbounded agents are "computationally omniscient," in the sense that they face no purely computational constraints. Any reasoning that a human 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 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's 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.[12] 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 reasoning 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 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 bounded agent could overcome. A bounded and an unbounded agent in the same 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

---

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

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.

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 this for granted.

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 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, 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. In assessing and assigning probabilities to norms, we have much less to go on than we do in, say, the physical sciences, which

are the paradigm example of a domain in which certainty is unattainable.[13]

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's 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 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 subjec-

---

[13]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).

tive choiceworthiness, unless these are the same).[14] 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. 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.[15] 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 norm, even if she has credence in rival nth-order norms that disagree 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 what is meant by "conscious" and "dispositional" uncertainty. Here is Sepielli's explanation:

> 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

---

[14]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.

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

> 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, 2014a, 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 an uncertain norm] 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, 2014a, p. 92)

This suggests that practical agents can never fully satisfy the demands of rationality. That's plausible with respect to bounded agents, which seems to be what Sepielli has in mind. But with respect to unbounded agents it's much less plausible, and the sort of conclusion we should adopt only if we are forced into it—which, I will shortly argue, we are not.[16]

Neither convergence results nor the distinction between conscious and dispositional uncertainty seem to rescue internalism from the threat of regress. So I conclude that the Regress Argument gives us good reason to accept externalism.[17] In the next section, we will see where this leaves us vis-à-vis normative uncertainty.

---

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

[17]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 on the basis of convergence results 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 *n*th-order normative beliefs: e.g., a "Lockean threshold" for full belief or a requirement that the probability assigned to *n*th-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

4 Enkratic externalism

On pain of regress, it seems, we must accept that at least one norm has belief-independent force, such that an unbounded agent is permitted (if not required) to follow its 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 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), appropriately formulated. Call this view *enkratic externalism.* To make the 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).

This view avoids the regress problem. 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).[18] And it is a *comprehensive* norm: For any option O in any choice situation, it asserts that O is permissible if it has

---

requirement of epistemic 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 (2014b) 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.

[18]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.

maximal expected choiceworthiness, and impermissible otherwise.[19] Thus, it settles all questions of rational permissibility without any need for higher-order norms. 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.

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 and first-order externalism. 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 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.

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 is simply any consideration that counts for or against some option, from a normative point of view

---

[19]One might claim that MEC is comprehensive only when combined with a rule for comparing the strength of reasons posited by rival objective norms (i.e., a solution to the "problem of intertheoretic value comparisons"). Without such a rule, it is typically indeterminate (except in cases of dominance) whether a given option maximizes expected choiceworthiness. The solution to this worry, I think, is to adopt what MacAskill (2014) calls the "universal scale" approach to intertheoretic comparisons. On this approach, objective norms are understood as different assignments to options of a shared set of cardinal choiceworthiness properties. When we are uncertain how to make intertheoretic comparisons between two objective norms (say, total and average utilitarianism), this should be understood as uncertainty between different "amplifications" (affine transformations) of one or both objective norms. If objective norms are characterized in this way, as rival assignments on a shared cardinal scale, then we do not need to resolve the problem of intertheoretic comparisons to apply MEC. For ideas in a similar spirit to MacAskill's that would serve the same purpose, see the discussions of "absolutism" in Riedener (2015, Ch. 3) and of "*de dicto* utilities" in Carr (2020).

(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 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). Thus, an agent's beliefs about rationality can vary independently of her beliefs about choiceworthiness. Even if we in fact have objective reason (even *decisive* objective reason) to be rational, it's possible to believe otherwise, and your belief that O is rationally required does not *constitute* a belief that you have objective reason to choose it. Therefore the enkratic conception of rationality, which makes rational requirements exclusively sensitive to your beliefs about objective reasons, does not make them (directly) sensitive to your beliefs about rationality. For instance, suppose you falsely believe MFT, and believe on that basis that some option O is rationally required. This need not mean that your *beliefs about your objective reasons* favor O, or (therefore) that O *is in fact* rationally required. 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.[20] 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.

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-

---

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

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*.[21]

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*.

MEC might be defended on the basis of the enkratic conception of rationality by claiming that to respond *appropriately* to one's objective reason belief is to respond *proportionately* both to the probability that one has a certain reason and to the strength of that reason, if one has it. Or it could be defended on instrumental grounds, along the lines of familiar money-pump or long-run arguments for expected utility maximization.[22]

---

[21]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).

[22]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

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. There is a towering literature on the strengths and weaknesses of expectational decision rules, and I don't mean to take a stand in that debate. Rather, 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.

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 $\Box \forall x (Rational(x) \leftrightarrow \varphi(x))$ (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.[23]

---

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.

[23]Arguments in this spirit are made by Broome (2013, p. 93), Bykvist (2013, p. 133), and

A bit more carefully: The motivating idea of internalism is that agents cannot be rationally bound by norms that they disbelieve. But the true general theory of rationality, by virtue of applying to all agents in all 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 rationality of their options. That would follow, 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:

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]

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... [T]he objector who searches for a thoroughly subjective standard is going to end up like Ponce de Leon."

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.[24] 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.[25]

    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.)

    But most importantly for present purposes, while the Argument from Fallibility favors externalism 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 rationality.

---

[24] For a similar observation regarding beliefs about moral requirements, see the discussion of ProbWrong in Weatherson (2014, p. 146).

[25] 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.

4.4   Against first-order externalism

The debate between metanormativism and first-order externalism has generated a substantial literature, with many arguments on both sides that I won't try to survey here.[26] But the most compelling argument against first-order externalism, in my opinion, 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 this can happen is for an agent to be certain of a false normative theory. For instance, if impartial utilitarianism is the true external first-order norm, then an agent who is certain that she has objective reason to favor 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 should 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

---

[26]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).

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-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 (though see Robinson (2022) for a contrary view). 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 O_1 is certainly at least as choiceworthy as O_2, and possibility more choiceworthy, then its expected choiceworthiness must be greater.[27]

Thus, there is compelling reason to favor enkratic externalism over first-order externalism as a way of avoiding metanormative regress. Moreover, the considerations that favor enkratic externalism over first-order externalism do not also favor internalism over enkratic externalism; enkratic

---

[27]Is enkratic externalism vulnerable to an analogue of Podgorski's argument, showing that it recommends options that are certainly *irrational*? No—at least, not any argument with the same force as Podgorski's. First, a rational norm can imply that an agent is sometimes rationally required to choose an option that she's correctly certain is less choiceworthy than an available alternative (as first-order externalist norms do, in Podgorski's example). But no rational norm can imply that an agent is rationally required to choose an option that she's *correctly* certain is *irrational*. (Or rather, any norm that implied this would be inconsistent, and MEC is clearly consistent.) So enkratic externalism cannot require an agent to act against her *correct* certainties either about choiceworthiness or about rationality. Second, enkratic externalism can plausibly explain why acting against your *false* certainties about rationality is unproblematic: The false certainty that O is irrational need not count against it at all from your perspective (since it is not in itself a belief about objective reasons, i.e., considerations that count for/against options); nor need it be accompanied by objective reason beliefs that count decisively against it from your perspective (since you may believe that you have no reason, or only *pro tanto* reason, to be rational). Third, constructing an argument analogous to Podgorski's would require that, just as the agent in Podgorski's example is uncertain which option is more choiceworthy according to a particular objective norm (because she is uncertain what outcome each option will have), an agent can be uncertain which of two options is *rationally required* according to a given *subjective* norm. But this may not be possible, if subjective norms are sensitive *exclusively* to an agent's explicit beliefs or other aspects of her mental state to which she has perfect introspective access.

24

externalism is not merely a compromise that mitigates the costs of two "extreme" views, but a principled, first-best approach to decision-making under normative uncertainty.

5   Conclusion

This paper had two goals: 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 should attribute belief-independent force to just one norm: the enkratic principle, comprehensively formulated to cover choices under uncertainty. The lesson of the regress problem is that what you rationally ought to do can perfectly well depend on your beliefs about morality and other sources of objective reasons, but not on your beliefs about rationality.

References

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

Broome, J. (2007). Is rationality normative? *Disputatio 2*(23), 161–178.

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

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

Carr, J. R. (2020). Normative uncertainty without theories. *98*(4), 747–762.

Elga, A. (2010). How to disagree about how to disagree. In R. Feldman and T. A. Warfield (Eds.), *Disagreement*, pp. 175–186. Oxford: Oxford University Press.

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

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

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

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

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

Harman, E. (2015). The irrelevance of moral uncertainty. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics*, Volume 10. Oxford University Press.

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

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

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

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

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

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

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

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

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

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

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

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

Podgorski, A. (2020). Normative uncertainty and the dependence problem. *Mind 129*(513), 43–70.

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

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

Riedener, S. (2020). An axiomatic approach to axiological uncertainty. *Philosophical Studies 177*(2), 483–504.

Robinson, P. (2022). Is normative uncertainty irrelevant if your descriptive uncertainty depends on it? *Pacific Philosophical Quarterly 103*(4), 874–899.

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

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

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

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

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

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

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

Sepielli, A. (2018a). How moral uncertaintism can be both true and interesting. In M. Timmons (Ed.), *Oxford Studies in Normative Ethics*, Volume 7, pp. 98–116. Oxford: Oxford University Press.

Sepielli, A. (2018b). Subjective and objective reasons. In D. Star (Ed.), *The Oxford Handbook of Reasons and Normativity*, pp. 784–799. Oxford: Oxford University Press.

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

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

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

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

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

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

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

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

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

Weatherson, B. (2019). *Normative Externalism*. Oxford: Oxford University Press.

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

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

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