Metanormative Regress: An Escape Plan

Christian J. Tarsney*

(Version 7, August 2019)
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

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

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 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

[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 (forthcoming). Expectational views are defended by Lockhart (2000), Ross (2006), Sepielli (2009), MacAskill and Ord (forthcoming), and Riedener (forthcoming), among others. Other metanorms have been proposed by Guerrero (2007), Nissan-Rozen (2012), MacAskill (2016), Tarsney (2018, forthcoming), 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.

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.

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. I examine two internalist responses and conclude that they are unsatisfactory. In §4, I introduce my own response to the regress problem, which posits a single belief-independent norm of practical rationality: the enkratic principle, correctly formulated. I argue that this form of externalism is both *prima facie* plausible and draws support from considerations independent of the regress problem. This "enkratic externalist" position satisfies the demands of the regress argument for ex-

[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).

ternalism, but that doesn't necessarily mean it solves the regress problem. In §5, I describe two ways in which the enkratic externalism might fail to block the regress and try to plug both these gaps. §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 $\{O_1^i, O_2^i, ..., O_n^i\}$ 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.)

[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.

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 implies 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, 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.

An *objective* norm is sensitive to empirical and counterfactual features of choice situations, like the fact that a particular option would have a particular consequence. And its sensitivity to these features is not mediated by the agent's credences: For instance, an objective norm may assess a given option as impermissible because it would harm some third party, even if the agent does not believe or have any evidence that it would. The output of an objective norm is an assessment of options in terms of *choiceworthiness*, i.e., the degree of objective reason the agent has to choose a given option, and a designation of options as either objectively permissible or objective prohibited.

A *subjective* norm, on the other hand, is sensitive only to facts about the agent's mental states, in particular her beliefs and/or evidence.[5] Just as objective norms yield an assessment of options in terms of objective reasons and designate options as objectively permissible/prohibited, so subjective norms yield an assessment of

[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 both. Likewise, references to an agent's "credences" should be understood to mean "either subjective credences, or evidential/epistemic probabilities, or some combination of the two."

options in terms of subjective reasons and designate options as *subjectively* or *rationally* permissible/prohibited.

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 omit the word "subjective" and simply refer to "first-order norms," "second-order norms," etc.)[6,7]

[6]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.

[7]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 useful discussion of the challenges

I have said that subjective norms are norms of *rational requirement*. This deserves a little explanation. A key substantive presupposition throughout this paper is that subjective norms *tell us how to respond to our beliefs* about normative phenomena in the world. That is, a true subjective norm specifies the appropriate response to a given belief state concerning objective and/or subjective reasons. Thus, a subjective norm is a kind of *coherence* requirement: It demands coherence between my beliefs about reasons and my choices/intentions. I find it convenient to identify "rationality" with this sort of coherence: Responding correctly to, or making choices that cohere with, one's beliefs about reasons, sounds a lot like rationality.[8] But this is a terminological matter on which, as far as I can see, nothing of great substance depends. I would be equally happy, for instance, to defend all the arguments and conclusions of this paper if a concept like "subjective ought" or "subjective rightness" were substituted for "rational requirement" throughout—and the reader should feel free to make the substitution.[9]

2.4 Internal and external subjective norms

Say that a choice situation S is *in the domain of* a norm N if N asserts of some option in S that it is permissible or asserts of some option in S that it is impermissible—i.e., if either N's choice set for S or N's prohibited set for S or both are non-empty. Say that N is *domain-restricted* by an agent's beliefs about some set of propositions Σ if, for some possible belief state with respect to Σ, no choice situation in which the

raised by indefinite extensibility, see for instance Hellman (2006) and Shapiro and Wright (2006).

[8]The link between rationality and coherence is not uncontroversial, of course. For a contrary view, see for instance Lord (2017).

[9]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).

agent has that belief state is in the domain of N.

Although an nth-order norm is not *sensitive to* an agent's nth-order normative beliefs, it can be *domain-restricted* by those beliefs. In this case, we will call it an *internal* norm. The paradigm of an internal norm is a norm N that applies only to choice situations in which the agents satisfies some threshold of belief with respect to N itself. Such a norm can be expressed as a conditional: "For any agent who believes N to such-and-such degree [e.g., with certainty], φ," where φ makes reference only to the agent's non-normative beliefs and her normative beliefs of order less than n. Because subjective norms can be internal, it is 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 ought to maximize expected hedonic welfare, while the true second-order norm says that agents ought to act on the first-order norm they consider most probable. 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 that norm with certainty, then there is no conflict: Whenever both norms apply to a given choice situation, they yield the same permissions and prohibitions.

An *external* nth-order norm, by contrast, has no such domain restriction: Whether it applies to a given agent does not depend on that agent's nth-order normative beliefs. *Externalism*, then, is the thesis that there is at least one true external norm. *Internalism* is the thesis that all true norms are internal.[10]

[10]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

2.5 Metanormativism and first-order externalism

An nth-order subjective norm N is *comprehensive* if (i) for every choice situation S in its domain, it classifies every option in S as either rationally permissible or rationally impermissible and (ii) its domain includes (at least) every choice situation in which the agent believes all the propositions in N with probability 1. Condition (ii) allows that internal norms may be comprehensive. Because whether a given choice situation is in the domain of an *external* norm never depends on the agent's nth-order normative beliefs, a comprehensive external norm applies to every choice situation whatsoever. (Note, however, that even a comprehensive external norm need not be *complete* in the formal sense: It may imply, for instance, that certain options are incomparable.)

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 applies to every choice situation and fully determines the subjective normative properties of all options in any choice situation to which it applies. Therefore, any true subjective norm can only yield normative 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 N^1_* 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

applies to any "fundamental policy, rule, or method" (p. 185).)

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.

3.1 Stating the argument

Here is an intuitive gloss of the argument: If the normative 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, the internalist view that norms only have normative force to the extent that the agent believes them implies that, for agents who are generally uncertain about basic normative principles, rational action 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* an option O in situation S if either (i) S is in the domain of N and N implies that O is permissible in S or (ii) in any minimal variant of S that

merely alters A's nth-order normative beliefs to place her within the domain of N, N would imply that O is permissible. Likewise, N *deauthorizes* O in S if either (i) S is in the domain of N and N implies that O is impermissible in S or (ii) in any minimal variant of S that merely alters A's nth-order normative beliefs to place her within the domain of S, N would imply 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 her in the domain of N.

P2. If internalism is true, then for any true nth-order norm N_i^n that authorizes an option O, if A also assigns positive credence to some rival nth-order norm N_j^n that deauthorizes O, then A's beliefs do not place her 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 pth-order normative beliefs place her in the domain of N^p.[11]

L1. If (i) internalism is true and (ii) 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][12]

[11]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, I assume that norms can be freely conjoined and disjoined, with the disjunction $N_1 \vee N_2 \vee ... \vee N_n$ yielding a (generally non-comprehensive) norm 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 imply 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 her in 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.

[12]How does this follow? By P1, A is permitted to choose O only if there is some norm on the

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 an unbounded agent 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

basis of which she is rationally permitted to choose it. By P2, a norm does not meet this condition if there is a rival norm of the same order that deauthorizes O and to which A assigns positive credence. Any subjective norm is either of order $\leq m$ or of order $> m$, since any norm either is or is not sensitive to the agent's beliefs about norms of some order $\geq m$. So, if no norm of order $\leq m$ meets the conditions necessary for A to permissibly choose O on the basis of it, then either *no* norm meets these condition (and A is not rationally permitted to choose O, as per P1), or there is a norm that meets these conditions of order $> m$.

[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 (more like the "Simple Argument" for externalism, introduced 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).

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 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 says they should 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 de-

[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, 1983) Lipman (1991), Smith (1991), and Lin (2014), among others). I discuss these issues in a companion piece (Tarsney, ms).

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

liberative capacities are relaxed (e.g., as the cost of a unit of computation in time or other resources goes to zero), our 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.[16]

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.[17]

P4 should also be relatively uncontroversial. It simply asserts that the consequent

[16]For instance, suppose that (i) the true second-order norm is MFO ("choose an option for which the total probability of comprehensive first-order norms that permit it is maximal"), (ii) this norm has external force with respect to unbounded agents, (iii) you are in a choice situation where MFO requires choosing option O, and this is not computationally difficult to figure out, but (iv) you are certain or nearly certain that the true second-order norm is MFT ("choose an option permitted by a comprehensive first-order norm whose probability is maximal"), which prescribes option P. In this case, the gap between bounded and unbounded agents seems irrelevant (since the computational demands of the situation are modest), but if externalism were true of unbounded agents and internalism true of bounded agents, then you would be rationally required to make a different choice than an unbounded agent in the same situation.

[17]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 imply the permissibility of O. If N is an external norm whose scope does not depend on A's beliefs, then this condition is trivially satisfied.

of L2 should be considered a *reductio*, i.e., we should not accept the conclusion that an unbounded and epistemically rational agent is never 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 asserts, formally, is that whether an nth-order norm applies to an agent A depends on A's nth-order normative beliefs. What motivates this assertion, presumably, is the idea that rational choice must be guided by (or at least cohere with) norms that the agents accepts 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 could be justified by the standard 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 in order 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* norm, but merely that at every level of normativity, she should be at least a little uncertain about the choiceworthiness 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.[18]

P3 can also be substantially weakened, at the cost of strengthening P4. For instance, we could weaken P3 to assert merely that, for *most* agents in *most* choice situations, it is *permissible* to be uncertain at each level of normativity 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* uncertain in this way 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 such objections, based on internalist responses to the regress problem in the recent literature.

3.2 Fixed-point approaches: convergence and contraction

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

[18]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).

P3. But why think that it's true? The most promising argument in this direction comes from Trammell (forthcoming), who shows that convergence is guaranteed under certain strong assumptions: in particular, when for every n, the agent has positive credence in only finitely many nth-order norms, all of which are cardinal, complete, and "compromising" (meaning that they assign each option a cardinal value strictly in between the minimum and maximum values assigned to it by the $(n-1)$-order norms in which she has positive credence). (Under these assumptions, convergence is guaranteed only at transfinite levels of the metanormative hierarchy. To guarantee convergence at finite levels, further strong assumptions are needed.) 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 on its face non-cardinal. And more obviously, many *first*-order normative theories are non-cardinal and/or incomplete. So these results, while interesting and important, do not seem like a general solution to the regress problem.

Sepielli (2014b) suggests another hypothesis that bears some resemblance to Convergence.

Contraction For any agent A in any situation S (perhaps excluding a few pathological cases), if A's credences are epistemically rational, then the set of options that some nth-order norm authorizes will contract monotonically as n increases.[19]

If Contraction is true, in other words, the set of options that *might* be permissible can only ever get smaller, never larger, as we ascend the hierarchy toward higher-and-higher-order norms. This means that, if A has only finitely many options in S, then the set of possibly-permissible options must eventually stabilize: There must be some n such that for all $p > n$, the set of options authorized by some pth-order norm in which A has positive credence is identical to the set of options authorized

[19]Sepielli suggests this hypothesis, but doesn't fully endorse it or claim that it solves the regress problem—see Sepielli (2014b, p. 539).

by some nth-order norm in which A has positive credence.

Contraction is more modest and hence more plausible than Convergence (though not strictly weaker). But it's not clear that, even if it were true, it would do anything to resolve the regress problem, unless some more ambitious hypothesis like Convergence were also true. The fact that, across all nth- and higher-order norms, the same set of options are *possibly* permissible (i.e., authorized by *some* norm in which I have positive credence) does not let me conclude that any of those options in particular is *in fact* permissible. To illustrate the point, Contraction is trivially true for any agent whose credences satisfy regularity, since for every option O and every n, she will have positive credence in some nth-order norm implying that O is permissible. For such an agent, the set of possibly-permissible options will never expand as she ascends the hierarchy, but only because it never contracts. This presumably does not mean that she is rationally permitted to choose any option in any choice situation.

3.3 Sepielli's view

Elsewhere, Sepielli suggests a different approach to the regress problem.[20] He starts by drawing a distinction between *conscious* uncertainty and *dispositional* uncertainty. A rational agent, he claims, may be dispositionally uncertain without being consciously uncertain. And if she acts on a norm N of which she is dispositionally but not consciously uncertain, 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 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 about which options are permissible, so long as that nth-order uncertainty remains merely dispositional

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

rather than conscious.

To assess this strategy, we need to know exactly what is meant by "conscious" 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, 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 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, 2014a, 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 the fixed-point approach nor the distinction between conscious and dispositional uncertainty seems 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.

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.

[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 three 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, a knowledge norm, 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 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 Titlebaum 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. I don't find these responses particularly promising, but I won't try to evaluate them here.

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*.

4.1 Motivating enkratic externalism

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.[23]

There is a substantial debate about how exactly to formulate EP, much of which need not concern us here. But one issue that very much does concern us is that familiar formulations of EP take no account of uncertainty. If "belief" is compatible with uncertainty, then EP1 is simply false, since 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. If "belief" implies certainty, then EP1 is never or almost never applicable to actual agents.

Wedgwood (2013) points this out and tries to generalize EP in a way that accommodates uncertainty about what one objectively ought to do. He concludes that the right generalization is a principle requiring agents to *maximize expected choiceworthiness*. Adapted slightly to fit the idiolect of this paper, Wedgwood's version of EP can be stated as follows:

EP2 (MEC) It is rationally required of any agent in any choice situation that she choose an option that maximizes expected choiceworthiness.

[23]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 convincingly in Reisner (2013).

Wedgwood suggests that this expectational version of EP is "the fundamental principle of rationality" (p. 505). The true, narrow reading of EP1 asserts a rational requirement that, if an agent is certain that she objectively ought to choose O—in other words, if she is certain that O is her most choiceworthy option—then she chooses O. But this is, Wedgwood suggests, just a limiting case of the more general rational requirement given by EP2, that agents make choices (or, for Wedgwood, form intentions) that cohere with their beliefs about objective reasons, weighing each potential reason for or against O in proportion to both (i) the probability that it obtains and (ii) its strength if it does obtain.

So that we have a reasonably spelled-out version of enkratic externalism to evaluate as a response to the regress problem, I will assume that Wedgwood is correct and that EP2 is the correct generalization of EP1. Needless to say, there is a towering literature on the strengths and weaknesses of expectational decision rules (though largely focused on axiomatic expected utility theory, which differs in important ways from MEC). 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 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 change the discussion that follows.[24]

[24]Wedgwood offers further arguments for MEC in Wedgwood (2017). MEC is also defended in MacAskill (2014), Lazar (2017), and MacAskill and Ord (forthcoming), 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.

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 another matter, which we will take up 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). 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 she can act on MEC without considering conflicting second-order norms, there is no need to resort to higher-order metanorms.[25]

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.

[25]It is worth noting here an interesting parallel between the metanormative regress problem and Lewis Carroll's famous paradox of Achilles and the Tortoise (Carroll, 1895). (Thanks to Ben Blumson for pointing this out to me.) The regress that emerges when the Tortoise demands that *modus ponens* itself be included as a premise in what started out as a simple *modus ponens* inference has often been interpreted as showing that we must recognize a fundamental distinction between *rules* and *premises* of theoretical inference. Likewise, the enkratic externalist solution to the regress problem can be understood as claiming that EP is a rule rather than a potential premise of practical reasoning, just as *modus ponens* is a rule rather than a potential premise of theoretical reasoning.

Notably, however, Carroll's regress doesn't involve any uncertainty about the rules of inference. (The Tortoise doesn't doubt the validity of *modus ponens* inferences, but simply requests that each instance of the *modus ponens* schema that figures in the argument be "written down" as a premise.) And it does not obviously support the conclusion that the normative force of valid inference rules is independent of the agent's *beliefs* about those rules. Plausibly, then, there are two distinct regress problems that can each arise in both the theoretical and practical domains: (i) the regress of theoretical/practical reasoning under *certainty* about basic norms, which shows that at least some norms must be accorded a special role distinct from that of accepted premises (for useful discussion of the direct practical analogue of Carroll's paradox, see §4.2 of Kolodny and Brunero (2018) and citations therein) and (ii) the regress of theoretical/practical reasoning under *uncertainty* about basic norms, which shows that the normative force of at least some norms is belief-independent.

Carroll's is just one of several regress problems in epistemology that in some ways parallel the metanormative regress problem. These include, for instance, the classic skeptical regress generated by the demand that all beliefs have a non-circular inferential justification; the problem of higher-order peer disagreement (Weatherson, 2013; Elga, 2010; Kelly, 2010; Christensen, 2013); and regress problems associated with higher-order evidence more generally (Lasonen-Aarnio, 2014). There are important parallels among these problems, though I think it would be a mistake to treat them all as mere facets of a single underlying problem. But constraints of both space and expertise forbid me from attempting any detailed exploration of these parallels.

4.2 Enkratic externalism vs. internalism

Even if it can resolve the regress problem, enkratic externalism still faces challenges from two directions: from internalists like Sepielli, and from first-order externalists like Weatherson. From the internalist direction, the challenge is to explain how it could be rational—let alone rationally required—of an agent to act on a norm that she does not accept. This challenge need not rely on the tendentious claim that an agent cannot rationally act on any norm of which she is *at all* uncertain. The internalist can instead ask: Suppose an agent believes that MEC is certainly or almost certainly false. Suppose she is certain, or nearly certain, of a rival subjective norm that gives conflicting advice for the choice she confronts. And suppose that her evidence supports these beliefs—e.g., she has been exposed to compelling arguments against MEC and for the rival norm. How can it be rational for *this* agent to follow MEC rather than the norm that she justifiably believes to be much more probably correct? This seems to gratuitously offend the idea that leads us to subjective norms in the first place, that rational choice should be guided by an agent's beliefs about the normatively significant features of her choice situation. Indeed, for this reason, enkratic externalism seems paradoxically to offend the essential spirit of the enkratic principle itself.

There is certainly a *prima facie* tension here, but I think there is ultimately no contradiction. Enkratic externalism follows naturally from a particular conception of rationality: Rationality is about coherence, and practical rationality is about coherence between an agent's normative beliefs and her choices. But more specifically, rationality is about coherence between an agent's choices and her *beliefs about objective reasons*—not her beliefs about *rationality*.

Why make this distinction? The normative force of rationality, plausibly, is that it aligns my choices with my beliefs about the features of my actions that *matter*. And while it matters that I do what I have most objective reason to do, it does not matter that I act rationally, except insofar as acting rationally is generally conducive

to doing what I have most objective reason to do. (For a forceful exposition of this point, see Kolodny (2005).) Thus, there is nothing *necessarily* irrational in doing something I believe to be irrational, except insofar as I also believe myself to be acting against my objective reasons. An agent who is motivated by the things that matter need not guide her choices by the former sort of belief. This resolves the apparent incongruity of a normative theory that is belief-sensitive in some regards and belief-insensitive in others: Rationality is, as a conceptual matter, about making your choices cohere with your beliefs *about objective reasons*, so its requirements are sensitive to these beliefs alone.

This defense of enkratic externalism in particular also points to a second argument for externalism in general, distinct from the Regress Argument. Whatever the correct conceptual analysis of rationality turns out to be, it will yield *some* conclusion of the form $\Box\forall x(\mathit{Rational}(x) \leftrightarrow \varphi(x))$ (where x might range over agents, attitudes, options/choices, or something else). The simplest argument for externalism is that, *whatever* the content of φ turns out to be, there is some rational requirement—namely, to satisfy φ—that is incumbent on any agent regardless of her normative beliefs, because satisfying φ is just what it *means* to be rational. Put a bit more generally, the argument is this:

The Simple Argument for Externalism

P1. There is some φ such that (i) $\Box\forall x(\mathit{Rational}(x) \rightarrow \varphi(x))$ and (ii) it's possible for an agent to have credence less than 1 that she is rationally required to satisfy φ.[26]

C. There is at least one true norm N (viz., the rational requirement to satisfy φ) that applies to all agents regardless of their beliefs about N.

This argument, as much as the Regress Argument, convinces me that we must

[26]I substitute a conditional for a biconditional in P1 since it's all the argument requires.

accept externalism. Whatever the true theory of rationality, as soon as we have stated it in its full generality, we have stated principles of rational requirement that apply to all agents whatsoever—meaning, to all agents regardless of their beliefs about rationality.[27] Note also that the Simple Argument 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.

Still, we should acknowledge that any form of externalism carries a genuine cost. We must give up some of the desire for "action-guidance" that seems to motivate the search for first- and second-order subjective norms. If enkratic externalism is right, then the norms of rationality can help us navigate our uncertainties about objective reasons (including both empirical and normative uncertainties), but cannot help us navigate our uncertainties about rationality itself.

4.3 Enkratic externalism vs. first-order externalism

I have defended enkratic externalism against the worries of internalists, but what about the opposite view, first-order externalism? Why should we attribute external normative force to EP, rather than to first-order norms like utilitarianism or Kantianism? The debate between first-order externalism and metanormativism is far too large to enter into here.[28] So I will limit myself to two observations: First, as

[27] Arguments like this 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."

[28] For the first-order externalist position, see Weatherson (2014, 2019), Harman (2015), and Hedden (2016). For metanormativist replies, see Sepielli (2016, 2018a), Johnson-King (2018), and MacAskill and Ord (forthcoming), among others. Podgorski (forthcoming) offers an objection to

a general principle of *rational choice under uncertainty*, MEC seems significantly more plausible than first-order principles like utilitarianism or Kantianism. I have argued that MEC, or some other version of EP, plausibly falls out of a conceptual analysis of rationality. By contrast, first-order norms like *maximize expected total pleasure minus pain* or *don't intentionally deceive other agents* cannot (*pace* certain Kantians) be plausibly derived from the mere concept of rationality. Rather, they reflect particular theories of what objective reasons there are in the world (e.g., reasons for producing pleasure or against deception) which, even if they are true, are not truths about what it means to be a coherent practical agent. Second, the first-order externalist position is incompatible with any version of EP, 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. If, say, subjective hedonistic utilitarianism is an external first-order norm, then an agent who is justifiably certain that she has most objective reason to act in ways favor the interests of her friends and family is rationally *prohibited* from doing what she is certain she has most objective reason to do. So whatever you think of generalized versions of EP like MEC, if you find *any* version of EP compelling, then you have powerful reason to reject first-order externalism.

5 Have we blocked the regress?

Although I initially 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. In the last section I suggested that, since EP is a second-order norm (being sensitive only to an agent's beliefs about objective norms), 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

first-order externalism that is, to my mind, all but conclusive. I give my own defense of metanormativism and reply to the first-order externalists in Tarsney (2017, Chs 2–3).

to consider third- or higher-order norms. But does EP alone allow us to set aside all our second-order normative 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.

5.1 Prohibition without permission

Two worries about the comprehensiveness of EP deserve our attention. First: The requirement to maximize expected choiceworthiness, as stated in the last section, does not necessarily tell an agent that she is *permitted* to choose any option that maximizes expected choiceworthiness. It rather *prohibits* her from choosing an option when there is some alternative that has *greater* expected choiceworthiness. This distinction is significant because more than one option can have maximal expected choiceworthiness. This can happen in at least two ways: First, there can be ties. Second, the expected choiceworthiness of an option can be undefined, in which case its expected choiceworthiness is maximal but the choiceworthiness of the option with the greatest defined expectation is maximal as well (since the expectation of an option with undefined expectation is, trivially, neither greater nor less than the expectation of any other option).

This opens a loophole through which the regress problem can reappear: EP (qua MEC) eliminates all options whose expected choiceworthiness is not maximal, but if this leaves more than one option, then we still have a non-trivial choice situation with multiple options and no belief-independent norm that tells us how to choose between them. The solution is to revise our statement of MEC so that it not only states

a rational requirement, but gives comprehensive necessary and sufficient conditions for rational permissibility.

EP3 (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.

This closes the loophole: Now, MEC not only prohibits options with non-maximal expectations, but permits agents to choose any option with a maximal expectation.

5.2 Normalization uncertainty

There is another, more challenging way in which metanormative regress can reappear within the constraints of enkratic externalism: 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). Even if she is fully committed to maximizing expected choiceworthiness, an agent may nevertheless 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, conditional on which she distributes her credence widely over a range of plausible-seeming normalizations (Riedener, forthcoming). This suggests, once again, that MEC is not genuinely 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? Even if she is fully committed to MEC, therefore, an agent

may 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 or evidence, 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 it is at least plausible that a version of the enkratic principle like MEC can be derived from a conceptual analysis of rationality, and therefore has belief-independent rational requiring force, it is much less plausible that any particular intertheoretic normalization method is somehow baked into the concept of rationality.

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 critical-level theory that assigns twice as much weight to the interests of those below the critical level as to those above it. She might be uncertain whether the critical-level 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 once again, we are just headed for another 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 as such are in the business of assigning to options. Different normative 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 and critical-level utilitarianism—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 critical level 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 (cf. MacAskill (2014, pp. 134–142)). If my uncertainty 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.[29] 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

[29]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), who defends a thesis in the same spirit as MacAskill's, called "absolutism," that would also serve my purposes in this section.

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 and critical level utilitarianism—it simply casts that uncertainty in a way that is tractable to expectational reasoning. Giving a fully worked-out defense of this view is (at least) a paper-length project 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.

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 us with 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 fruitfully 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 discussed in §3.2, or exploring one of 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

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

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.

Carroll, L. (1895). What the Tortoise said to Achilles. *Mind 4*(14), 278–280.

Christensen, D. (2013). Epistemic modesty defended. In D. Christensen and J. Lackey (Eds.), *The Epistemology of Disagreement: New Essays*, pp. 77–97. Oxford: Oxford University Press.

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.

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

Elster, J. (1983). *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge: Cambridge 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.

Hellman, G. (2006). Against 'absolutely everything'! In A. Rayo and G. Uzquiano (Eds.), *Absolute Generality*, pp. 75–97. Oxford: 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.

Kelly, T. (2010). Peer disagreement and higher order evidence. In R. Feldman and T. A. Warfield (Eds.), *Disagreement*, pp. 111–174. Oxford: Oxford University Press.

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

Kolodny, N. and J. Brunero (2018, Winter). Instrumental rationality. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Lasonen-Aarnio, M. (2014). Higher-order evidence and the limits of defeat. *Philosophy and Phenomenological Research 88*(2), 314–345.

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.

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

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

Lord, E. (2017). What you're rationally required to do and what you ought to do (are the same thing!). *Mind 126*(504), 1109–1154.

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 (forthcoming). Why maximize expected choice-worthiness? *Noûs*. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/nous.12264.

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. (forthcoming). Normative uncertainty and the dependence problem. *Mind*. URL: http://www.abelardpodgorski.com/uploads/5/4/0/0/54000997/normativedependence.pdf.

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. (forthcoming). An axiomatic approach to axiological uncertainty. *Philosophical Studies*. URL: https://link.springer.com/article/10.1007/s11098-018-1191-7.

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.

Shapiro, S. and C. Wright (2006). All things indefinitely extensible. In A. Rayo and G. Uzquiano (Eds.), *Absolute Generality*, pp. 255–304. Oxford: Oxford University Press.

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

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. (forthcoming). Normative uncertainty and social choice. *Mind*. URL: https://doi.org/10.1093/mind/fzy051.

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. (forthcoming). Fixed-point solutions to the regress problem in normative uncertainty. *Synthese*. URL: https://link.springer.com/article/10.1007/s11229-019-02098-9.

Weatherson, B. (2013). Disagreements, philosophical and otherwise. In J. Lackey and D. Christensen (Eds.), *The Epistemology of Disagreement: New Essays*, pp. 54–73. Oxford University Press.

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

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