Reply to commentators

From *Philosophical Studies* symposium on *Inquiry under bounds*
Penultimate draft, please cite published version

Abstract

This paper responds to commentaries on *Inquiry under bounds* by Sara Aronowitz, José Luis Bermúdez and InJoon Seo, and Julia Staffel.

1 Introduction

I would like to thank Aronowitz, Bermúdez, Seo and Staffel for their insightful commentaries. Let me say a bit about each commentary in turn.

2 Aronowitz

Aronowitz asks four important questions about my view. Can agents have reason to improve their bounds (Section 2.1)? Are there epistemic reasons for action (Section 2.2)? What exactly is clutter avoidance (Section 2.3)? And whose beliefs and desires matter to normative evaluation (Section 2.4)? I address each question below.

2.1 Improvement for bounded agents

Bounded rationality theorists are not fatalists. We think it is important for agents to strive to move beyond their current bounds (Musslick and Masís 2023). A good theory of bounded rationality should tell agents when and how to improve their bounds.

Aronowitz notes two projects that my book does not have. First, I do not give an account of the degree to which agents are rational, but only of rational permissions and requirements. Rational permissions and requirements do not apply to agents, but only to features of agency. As such, I cannot directly ground interpersonal comparisons between

the rationality of an adult and a toad, a toddler, or even another adult. Another commentator in this symposium has made substantial progress towards a theory of degrees of rationality, which might be of use in grounding these evaluations (Staffel 2020).

Second, I do not allow bounds to be objects of rational assessment. On my view, rational permissions and requirements apply to features of agency. An agent's current bounds are no more features of agency than her eye color, and it would be a mistake to rationally assess them. This does not mean that we cannot ask other normative questions about bounds. For example, I do allow axiological questions to be asked about the value of bounds, and we will see below that these axiological questions play their traditional role in explaining the rationality of acts aimed at extending an agent's bounds.

As Aronowitz notes, my account does apply to the rationality of acts aimed at extending our bounds, such as gathering evidence to expand informational bounds. Aronowitz asks an important question: against which bounds are such acts assessed? As always, the answer is that rationality is assessed against an agent's current bounds. Our current abilities limit the options available to us and structure the costs and benefits of those options. And our current evidence is the perspective from which the rationality of those options must be assessed.

This story is familiar from the most popular interdisciplinary account of rational evidence-gathering (Good 1966; Howard 1966; Stigler 1961). On this story, acts of evidence-gathering have costs, such as energy, time, and foregone alternatives. Acts of evidence-gathering also have benefits: they improve the value of an agent's decision problem, understood as the expected value of the best available option. Agents should gather evidence when expected benefits exceed expected costs.

On this story, expectations are taken relative to evidential probabilities on the agent's current evidence, options are limited by her current abilities, and costs and benefits are structured by her current abilities. But agents can readily recognize from within this perspective that evidence-gathering is often required because better information will likely put them in a better position to bring about good outcomes.

2.2 Epistemic reasons for action

Chapter 4 argues that there is no distinctively epistemic reading of four normative terms: rationality, justification, and what agents should or ought to do. I do not deny that there are epistemic reasons for action.

Chapter 3 adopts a rich axiology on which many things may bear final value, including intellectual achievements such as accuracy and knowledge. Chapter 3 argues that we have reason to promote all outcomes of value. In this sense, we have reason to promote intellectual achievements such as true belief and knowledge. Should we draw a circle around the reasons to promote these intellectual achievements and call those reasons epistemic reasons? That is not my preference, but while I might object to some downstream conclusions based on this exercise, I cannot object to a simple act of naming.

Aronowitz asks whether the Standard Picture is committed to versions of evidentialism and agent-neutrality. Although defenders of the Standard Picture may be separately sympathetic to evidentialism and agent-neutrality, neither is a commitment of the view. Here is how Edward Stein (1996) originally introduced the Standard Picture:

> On what I call the *standard picture of rationality* ... to be rational is to reason in accordance with principles of reasoning that are based on rules of logic, probability theory, and so forth. (Stein 1996, p. 4)

My own characterization follows Stein, except I allow focal points beyond reasoning to be subject to rational assessment:

> Rational requirements [are] exemplified by requirements of consistency derived from logic, probability theory, and decision theory. (Thorstad 2024b, p. 9)

Neither formulation is committed to evidentialism or agent-neutrality.

However, it is a virtue of my account that it can be made compatible with traditional statements of evidentialism and agent-neutrality:

(Evidentialism) For all agents S and propositions p, if S believes p then S's belief that p is rational only if S's total evidence supports p and S's belief that p is based on her evidence.

(Agent-neutrality) Value is agent-neutral.

By contrast with some recent consequentializing attempts (Dreier 2011; Portmore 2011; Seidel 2019), my brand of consequentialism uses an agent-neutral conception of value. I show in Chapter 3 how the view can be not only made compatible with evidentialism, but indeed used to protect it against a range of criticisms.

As Aronowitz notes, it does seem plausible that agents can have reason to consider hypotheses beyond the practical benefits of doing so. As before, this follows from a rich axiology together with the consequentialist refrain that all value is to be promoted. Truth and knowledge are important achievements that we, as philosophers, devote much of our lives to because we recognize them as valuable in their own right.

One difficulty in these discussions is that authors often differ in their understandings of what makes a norm epistemic. Aronowitz asks what, if anything, should be taken as the mark of an epistemic norm. I agree wholeheartedly with Aronowitz's concern. In my view, a once tightly constrained notion of epistemic normativity has given way to a proliferation of different conceptions due largely to the methodological insistence that epistemologists describe their ever-expanding projects as continuing to study a type of epistemic normativity.

I am not sure what to do about this situation. Chapter 4 notes that many of our most traditional motivations for positing and privileging a distinctively epistemic type of normativity governing belief do not apply to inquiry. We might react by describing a more expansive type of epistemic normativity governing inquiry. Alternatively, we might react by questioning the need to recover distinctively epistemic norms of inquiry. Why not simply begin by studying norms of inquiry and follow the argument where it leads (Kelly 2011)?

2.3 What is clutter? I don't know

Gilbert Harman (1986) originally offered three motivations for principles of clutter avoidance. Agents should refrain from cluttering their minds to avoid wasted effort, conserve memory storage space, and avoid taxing limited capacities to retrieve and process stored information.

Chapter 6 uses Harman's motivations to ground norms of clutter avoidance on investigation, memory encoding and forgetting:

> **(Junk Non-Investigation)** Agents should not expend effort investigating junk subject matters.
>
> **(Junk Non-Encoding)** Agents should take steps to ensure that junk beliefs are not encoded in long-term memory.
>
> **(Junk Forgetting)** Agents should take steps to ensure that junk beliefs are forgotten.

Aronowitz mounts a series of challenges to existing views about clutter avoidance based around an attack on the underlying metaphor of a cluttered memory space. Aronowitz notes that it is often difficult to individuate beliefs, that beliefs come in different types, and that on many ways of typing and individuating beliefs it is not clear that there is rigorous sense to be made of the idea that beliefs clutter a fixed memory space which could soon be full. Aronowitz proposes that we abandon the metaphor of spatial clutter and instead focus on the ways in which superfluous representations can interfere with downstream cognitive processes.

This is an important and plausible set of corrective remarks. If true, they have at least three implications for future discussions of clutter avoidance. First, in motivating principles of clutter avoidance we may do well to de-emphasize Harman's appeal to limited memory space and focus instead on wasted effort and limited capacities to retrieve and process stored information.

Second, many principles of clutter avoidance – including my own – are stated at a high level of generality. These principles might be clarified by incorporating details about the types of representations available, as well as the processes of encoding, retrieval and downstream information processing.

Third, once we do this we may see that existing principles of clutter avoidance – including my own – pass over many of the most important questions about clutter avoidance. For example, the notion of junk belief will be expanded to incorporate details about representation, encoding and processing, in light of which what counts as a junk belief will be a very complex question sensitive to a great many specific details about agents and contexts. Filling in these details should help to enrich existing principles of clutter avoidance and would be a productive project for future work.

2.4 Whose utility? And when?

Theories of rationality come in two types.[1] *Descriptive* theories of rationality aim to predict and explain behavior. They answer the questions of what agents will do and why they will do it. Descriptive theories such as revealed preference theory (Samuelson 1938; Houthakker 1950), radical interpretation (Davidson 1973) and minimal rationality (Cherniak 1981, 1986) are essential tools in scientific attempts to predict and explain behavior.

Normative theories of rationality aim to normatively evaluate behavior. They answer the questions of what agents are permitted and required to do. On many accounts, normative theories are strictly more demanding than descriptive theories. Agents must satisfy descriptive rationality conditions in order for it to make sense to speak of them as having beliefs and desires, and of taking actions on the basis of their beliefs and desires. When it makes sense to treat agents as having beliefs and desires and taking actions on their basis, normative theories answer the further question of what agents are permitted

[1] Bermúdez (2009) offers a helpful argument for, and application of this distinction. Bermúdez makes a further distinction between two types of normative theories, on which I will remain neutral here.

and required to do.

Scientific theories of rationality are often engaged in the descriptive project of predicting and explaining behavior. This project is necessarily subjectivist.[2] It is an agent's own beliefs and desires that determine her actions, not the beliefs that the evidence supports or the objective values of outcomes. This is why subjective credences and subjective utilities are commonly used in scientific accounts of bounded rationality, as Aronowitz notes.

Normative theories of rationality need not, and probably should not exactly match our descriptive theories. On my approach, rationality is determined by evidential probabilities rather than by an agent's own credences, and by the actual values of outcomes rather than by the subjective values that agents place on them.

I am open to gentle loosenings of this framework, such as a normative uncertainty approach which incorporates uncertainty about the values of outcomes (MacAskill et al. 2020). I am not a subjectivist, and in Chapter 3 I argue that subjectivist accounts may struggle to recover some plausible verdicts about bounded rationality. However, readers who wish to incorporate elements of the subjectivist story can easily do so while retaining the rest of my account. It would be an interesting project to explore the implications of these modifications for questions about inquiry and bounded rationality, as well as the prospects for pluralism about bounded rationality (Berg 2003, 2014).

Aronowitz considers a case study of friendship. She notes that while friendship confers intrinsic and instrumental value on friends, it may nonetheless be the case that some friendships do more harm than good, for example because you already have too many friends. When this is the case, my account implies that you should act with an eye to phasing out harmful friendships.

Aronowitz argues that there is a subjective sense in which what matters is only your concern for this particular friend and not any further questions about other friendships. I do not think that this view is naturally glossed as subjectivist. A natural subjectivist

[2] At least, it is *ontologically* subjectivist in that an agent's own attitudes are taken to be the fundamental explanations for her behavior. It may nonetheless be epistemically helpful to use proxies such as evidential probabilities or objective utilities when an agent's own attitudes are difficult to measure directly.

account would see Aronowitz as adopting a subjective value function while remaining neutral on the relevant probability function. On this view, if you care greatly about your friend, then that raises the expected value of acting to continue the friendship. Only, this account does not get us what we want, since the instrumental harms of having too many friends also impact things that the agent cares about. Hence I do not think that subjectivizing an agent's value function will account for a reading of rationality which screens off questions about other friendships from questions about how to act towards those we care about.

I think it is better to regard Aronowitz's intuition as a claim about a normative category other than rationality. For example, it could be a view about directed duties. On this view, friendships generate relationship obligations that depend only on the friendship itself and not on considerations about competing friendships (Scanlon 1998). Alternatively, Aronowitz might be read as making a point about the nature or virtue of friendship. To be a friend may involve being moved directly by your concern for your friend's well-being in a way that is at least partly independent of cost-benefit analysis (Badhwar 1987, 1991).

However, at the level of rationality, consequentialists remind us of the importance of tradeoffs, which are ubiquitous in the lives of bounded agents. Most of us have, at some point, had more friendships and caring relationships than we could manage. In this situation, it is natural to think that we should act to gradually reduce our commitments to a more manageable level. In doing so, we may well violate relationship obligations and other directed duties or even begin to think and act in ways that are incompatible with friendship. We may nonetheless be doing what rationality requires.

3 Staffel

Staffel asks important questions about how level separation should be formulated and how cases of level tension should be resolved. Let us take each in turn.

3.1 Formulating level separation

I was not sufficiently clear in formulating level separation or the direct normative theory that underlies it. Nor, to my understanding, were leading global consequentialists.[3] We should have done better, and Staffel's discussion provides a needed opportunity to clarify the view. I will do my best to provide a first-pass sketch of how the view should be developed, with the understanding that this account is preliminary and likely requires both clarification and adjustment. I promise to provide a better account soon (Thorstad ms).

Normative terms, such as 'rational', 'ought', or 'blameworthy', are applied at *evaluative focal points*, such as action, inquiry and belief (Kagan 2000). We can ask, for example, what it is right or rational to believe, or how it is right and rational to inquire.

On a first pass, *indirect* normative theories tie the normative status of some focal points to the normative status of others. For example, rule consequentialists hold that an act is right just in case it results from the right general rule (Brandt 1959; Harsanyi 1977). *Direct* normative theories do not do this.

It is possible to adopt direct normative theories for some focal points or evaluative terms while adopting indirect theories for other points or terms.[4] However, many direct theorists have been *everywhere direct*, evaluating all focal points directly with respect to all normative terms.[5] In this paper, I will use 'direct' to mean 'everywhere direct' unless otherwise noted.

On this first pass, direct normative theories make the normative status of all focal points independent:

> **Direct normative theory V1 (DNT-1):** For all normative terms N and focal points F_1, F_2, the N-status of F_1 is independent of the N-status of F_2.

[3] See for example Driver (2001, 2012); Kagan (2000); Ord (2009); Parfit (1984) and Pettit and Smith (2000).

[4] Kagan (2000) explores this possibility in some detail.

[5] On some ways of formulating the notion of direct normative theory, this definition may be too strong to capture some leading views. It is possible to hold direct theories for all focal points, but only some normative terms, or for some focal points and all normative terms.

DNT-1 does not make the stronger claim that the N-status of F_1 is independent of other normative statuses of F_2. Indeed, consequentialists might wish to claim that the rightness of inquiry depends on the value of its outcomes (a point to which we will return later).

Level separation is the restriction of direct normative theory to governing focal points. A focal point F_1, such as practical inquiry, *governs* another focal point F_2, such as intention, if members of the governed point (intention) are typically produced and modified by members of the governing point (practical inquiry). Level separation asks us to apply direct rather than indirect normative theories across governing focal points:

> **Level Separation V1 (LS-1):** For all normative terms N and focal points F_1, F_2, if F_1 governs or is governed by F_2, then the N-status of F_1 is independent of the N-status of F_2.

Let's focus on correctly formulating direct normative theory, then we can recover the corresponding statement of level separation.

DNT-1 is a good first pass, but it is at least under-specified. Consider a constant act a which is guaranteed to produce a single outcome o. Consequentialists would like to say that the value of a depends on the value of o. The value of a is a matter of the consequences it promotes, and the value of o captures the value of these consequences. But we might equally well say that the value of o depends on the value of a. After all, for most consequentialists it is a necessary claim that an outcome receives the same value as the constant act which produces it. But then purportedly direct consequentialists would seem to be no better than indirect consequentialists who say, one level above this, that the value of an act depends on the value of the rule that produces it.

The lesson of this discussion is that the notion of independence in DNT-1 needs to be strengthened to something stronger than necessary coextensiveness. Readers are welcome to reformulate DNT-1 using their preferred heavyweight metaphysical notion. I will reformulate DNT-1 in terms of metaphysical grounding:

> **Direct normative theory V2 (DNT-2):** For all normative terms N and focal

points F_1, F_2, the N-status of F_1 is not grounded in the N-status of F_2.

For generality, I will take the notion of grounding in DNT-2 and throughout this paper to be (weak) partial grounding unless otherwise specified. Full grounding falls out as a special case.

DNT-2 resolves our previous concern. For consequentialists the value of a constant act is grounded in the value of its outcomes. But the value of those outcomes is not grounded in the value of the constant act that produces them. Grounding, unlike necessary coextensiveness, is not a symmetric relation.

However, mainline consequentialists do seem to allow grounding claims in the other direction. As Staffel notes, I use expected value theory to ground the value of inquiries in terms of the value of the attitudes they produce. And we saw above that consequentialists do seem to ground the value of constant acts in the value of their outcomes. This violates DNT-2.

The echoes of what we should like to say are sketched by Shelly Kagan (2000), though we will see below that this is a bit too strong. The grounding claims that really matter relate evaluative focal points to the features of outcomes they produce and not to normative claims about those outcomes or other focal points. For example, it may seem that consequentialists make the value of an act a depend on the value of its outcomes together with the evidential probabilities Pr_E of those outcomes coming about:

$$V(a) = \Sigma_{o_i} Pr_E(o_i|a)V(o_i). \tag{1}$$

(1) makes the value of a depend on facts of the form f_1 = [a has evidential probability n of bringing about an outcome with value r]. This violates DNT-2 because of its reliance on the value of outcomes.

However, we can also specify outcomes o_i by their features as $< o_i >= (o_i^1, \ldots, o_i^n)$. A mature axiology tells us which features of outcomes matter and how they contribute to the overall value of outcomes, yielding a function f from features of outcomes to outcome

values. Consequentialists then also hold:

$$V(a) = \Sigma_{o_i} Pr_E(o_i|a) f(< o_i >). \tag{2}$$

(2) makes the value of a depend on facts of the form f_2 = [a has evidential probability n of bringing about an outcome with features $< o >$]. This does not violate DNT-2.[6]

A hardline response could be that (1) expresses a relation of necessary coextensiveness but not a grounding claim. This response denies that facts of the form [a has value n] can be partly grounded in individual facts of the form f_1 or fully grounded in a collection of facts with form f_1.

There may be some pull to this response.[7] We might think that making the value of acts depend on the value of outcomes, as opposed to their features, involves one thought too many (Williams 1981). On this line, the value of acts is already fully grounded in the features of their outcomes. To say that those outcomes are valuable is to add nothing to what is already a complete account of the values of our acts. Alternatively, we might say that making the value of acts depend on the value of outcomes fetishizes normative properties such as value by reifying them as objects of pursuit in their own right (Smith 1994), or that it ignores what Ross called the *resultant* nature of moral properties: they can only be possessed in virtue of possessing other properties (Berker 2018; Ross 1930).

At the same time, there is no need to be so restrictive. Some grounding relations are *mediate*. For example, the fact [$A \wedge (B \wedge C)$] is fully mediately grounded in the facts [A], [B] and [C], because [B] and [C] fully immediately ground [$B \wedge C$], and [A] and [$B \wedge C$] fully immediately ground [$A \wedge (B \wedge C)$]. Others grounding relations are *immediate*. For example, the fact [$A \wedge (A \wedge A)$] is fully immediately grounded in [A] (Fine 2012).

Crucially, to say that some collection of facts immediately grounds another is not to say that we cannot add superfluous mediating grounds. After all, [A] fully immediately

[6]It might be thought that facts about f must also figure in f_2, which would violate many readings of DNT-2. See Berker (2018) for an argument against giving this grounding role to normative theories.

[7]However, the first two of these arguments may work best against the idea that agents should show *de dicto* concern for morality, rather than against claims of moral grounding, and even here those arguments are controversial (Copp 1997; Lerner 2018).

grounds [$A \wedge A$], and [$A \wedge A$] and [A] fully immediately ground [$A \wedge (A \wedge A)$]. To say that some collection of facts fully immediately grounds a further fact is simply to say that in addition to any mediate grounding, the collection of facts also fully immediately grounds the further fact (Fine 2012).

Consequentialists think that acts and other features of agency are valuable because of the outcomes they produce. In the present instance, this is naturally read as claiming that facts of form f_2 fully immediately ground facts of the form [a has value n]. To say what outcomes an act may bring about, what their features are, and how likely they are to arise is enough to say why the act has the value that it does. Similar arguments can be made for other normative terms and focal points.

If this is right, then we should reformulate direct normative theory in terms of immediate grounding:

> **Direct normative theory V3 (DNT-3):** For all normative terms N and focal points F_1, F_2, the N-status of F_1 is fully immediately grounded in a collection of facts not involving the N-status of F_2.

DNT-3 allows direct normative theories to accept that facts about the value of acts may be fully immediately grounded in facts of the form f_1, so long as they are also grounded in facts of the form f_2.

DNT-3 also explains why rule consequentialism and related views are not direct normative theories. Rule consequentialists hold that acts are best because they result from the best rules. They also hold that rules are best because they have the best consequences. Glossing over many complications, this grounds the value of rules R in facts of the form f_3 = [R has evidential probability n of bringing about an outcome with features $< o >$].

Motive consequentialists could certainly *claim* to be offering a direct theory on which facts of the form f_3 immediately ground facts about the value of acts, of the form [a has value n]. Only, this is not a plausible claim. Facts about the consequences of rules are not immediately relevant to the values of actions that result from rules. Any grounding

that f_3 offers for facts about the value of actions that result from rules is mediated by the grounding they offer for facts about the best rules.

If this is right, then we have finally arrived at a formulation of direct normative theory that explains why direct theorists can make the claims considered above, and which also offers a plausible explanation of why leading indirect theories should not count as direct. It yields a corresponding version of level separation:

> **Level Separation V3 (LS-3):** For all normative terms N and focal points F_1, F_2, if F_1 governs or is governed by F_2, then the N-status of F_1 is fully immediately grounded in a collection of facts not involving the N-status of F_2.

That is not to say that we have told the final story about the matter.

Several features of DNT-3 require clarification. For example, what does it mean to say that a collection of facts involves the N-status of F_2? DNT-3 also bears further reformulation. For one thing, DNT-3 should be weakened to allow for ungrounded normative claims. For another, it does not look necessary for DNT-3 to make any reference to a second focal point F_2. That reference is needed in LS-3, but why does it belong in DNT-3?

These and other modifications will have to wait for future work (Thorstad ms). I hope that in the meantime, it is plausible that future work could lead to a correct and defensible specification of the distinction between direct and indirect normative theories as well as the weaker notion of level separation. I am sorry that these clarifications were so long in coming, and I hope that the next attempt to make them will be my last.

3.2 Resolving conflicts between process and attitude norms

Level separation allows for cases of *level tension*, in which the same normative term is applied with opposite valence across governing focal points. For example, it may be that a rational process of inquiry leads agents to produce irrational attitudes. Sometimes, there is no way to avoid level tension. The only way to produce a rational belief would be to use an irrational process of inquiry. How should this conflict be resolved?

The standard response to this question is to insist that level tension is not a conflict to be resolved, but an unavoidable feature of our normative landscape. Learning to live with level tension was the project of Part 1 of Parfit's (1984) *Reasons and persons*. Global consequentialists have largely followed Parfit in this matter.

Learning to live with level tension can be difficult. If rational inquiry will not produce a rational attitude, then we are tempted to ask, what should we *do*? Should we inquire rationally, or believe rationally? Following Parfit, global consequentialists have urged us to come to see this question as resting on a mistake.

There is no easy way to articulate the mistake, since there are many things that can be meant by the simple question of what should be done in cases of level tension. Perhaps the most straightforward attempt to articulate what is going on here is due to Shelly Kagan:

> If the best rules direct us to different acts from the best motives, which in turn direct us to different acts from the best norms, and so on, what are we to do? Despite the appearance of difficulty, however, this question wears its answer on its own sleeve. If the question is what are we to *do* in the face of such conflicts, then the question is one about acts, and we already know the standard by which acts are to be evaluated: directly in terms of the good. Sometimes this means that the right thing to do will be to instill the best motive, and sometimes it will mean that the right thing to do is to promulgate some rule; sometimes we will have to choose between the two ... But even if rightfully neglected, the best rule (say) is still the best rule for all that. So long as we remain clear about what *exactly* we are evaluating, the difficulty disappears. (Kagan 2000, p. 152)

I recognize that this answer is frustrating. I required many rounds of quasi-Wittgenstenian therapy to convince myself that there was no adequate way to reformulate the challenge that Kagan and others urge us to make do without. Certainly more discussion is required – this is why the first part of *Reasons and persons* contained so many intricate distinc-

tions. Continuing that discussion should help us to make peace with level tension by understanding how it answers or dissolves the various attempts to resolve level tension.

As Staffel notes, in many cases of level tension the governing focal point will answer to many or all of the same normative considerations as the governed focal point. This does not mean that assessments at the governing focal point are in any sense primary or should replace or take precedence over assessments at the governed point. It makes perfect sense to ask how I should decide whether to buy coffee, but also perfect sense to ask if I should buy coffee.[8] Both questions have a rich role in our descriptive and philosophical theorizing and our everyday lives.

4 Bermúdez and Seo

4.1 Epistemic norms and constitutive norms

Until recently, epistemologists largely thought that there were no epistemic norms of inquiry. In recent years, many epistemologists have begun to report the opposite intuition. Certainly, the view that there are epistemic norms of inquiry is a respectable one. But this does not mean we should treat the view as an intuitive datum. To do so risks wielding intuition reports in the manner that experimental philosophers warn us against: as means of enforcing contingent intuitions situated in a particular time and place that are far from universally shared (Machery 2017; Weinberg et al. 2001).

When I first argued against the claim that there are epistemic norms of inquiry, there were few published defenses of the contrary view. In recent years, a number of authors have given detailed defenses of the claim that there are epistemic norms of inquiry (Flores and Woodard 2023; Kelp 2021a,b). Bermúdez and Seo take one popular approach, which treats epistemic norms of inquiry as constitutive norms of inquiry. The arrival of this and other approaches is an important development.

There are a great many accounts of constitutive normativity in ethics (Korsgaard 2009;

[8] And yes, you should buy coffee. Coffee is great.

Velleman 2000) and epistemology (Maguire and Woods 2020; Kelp and Simion 2021), including some prominent constitutivist views within the epistemology of inquiry (Kelp 2021a,b). There are also a great many objections to constitutivist accounts (Enoch 2006; Friedman 2024) and responses to them.

To a large extent, I suspect that readers will come out of this brief discussion holding the same opinions about constitutivism they began with. Readers friendly to constitutivism should retain their favorite story about constitutive norms of inquiry and read me as giving an account of the non-constitutive norms governing inquiry. But it might be worth taking some time to briefly formulate constitutivism (Section 4.2) and recall some common doubts about constitutivism as well as the link between constitutivism and epistemic normativity (Section 4.3).

4.2 Formulating constitutivism

There are a great many varieties of constitutivism. Most agree that activities have constitutive aims.[9] This often means at least the following (Katsafanas 2018):

> **(Aims)** For all activities A and aims G, if G is the constitutive aim of A then:
>
> > **(All Tokens Aim)** All tokens of A aim at G.
> >
> > **(Aim Constitutes Activity)** Aiming at G is part of what constitutes an act as a token of A.

All Tokens Aim can be further clarified by saying what it means for an act token to have an aim. For example, on a *personal* reading, act tokens of A aim at G just in case all agents tokening A aim at G. By contrast, on a *metaphysical* reading, to say that tokens of A aim at G is to say that it is in the nature of A to be governed by G.[10]

Many constitutivists link aims to standards of correctness or success.[11] On a traditional view:

[9] Korsgaard (2009) prefers to bypass aims and move straight to constitutive principles.
[10] Further positions are possible, but I am trying to be brief.
[11] I do not distinguish between claims about correctness and success in this paper.

(Aim-Standard Link) For all activities A and aims G, if G is the constitutive aim of A then tokens of A are successful just in case they achieve G.

Some constitutivists stop there, but many aim to use claims about successful activity to ground other norms. The most ambitious constitutivists (call them *global constitutivists*) aim to derive all norms from constitutive norms (Korsgaard 2009). Most constitutivists have more limited ambitions, such as establishing an internal sense in which agents ought or are rationally required to A in a way that promotes G.

4.3 Questions for constitutivism

4.3.1 Constitutive norms and epistemic norms

A preliminary question to ask is how constitutivism is meant to ground the existence of a specifically epistemic type of normativity. The view cannot be that all constitutive norms of inquiry are epistemic norms. After all, most constitutivists think that action is governed by norms such as the categorical imperative (Korsgaard 2009) or encountering and overcoming resistance to one's ends (Katsafanas 2013), which are not epistemic norms in any standard sense. Inquiry is an activity, so it will also be governed by constitutive non-epistemic norms. Of course, the constitutivist is free to give an account of why some constitutive norms belong to a distinctive variety of epistemic normativity. But it is not clear that constitutivism will play any special role in this story.

4.3.2 Existence of constitutive aims

More generally, it is worth reviewing some standard concerns about constitutivism. The first concern is why we should think that activities have constitutive aims. Constitutivists often motivate this claim by pointing to games such as chess, which seem to aim at their winning conditions, in this case checkmate. But it is far from clear that this should lead us to believe that all activities have constitutive aims, instead of treating the constitutivist

as having made an important point about the nature of games. What, for example, is the constitutive aim of walking?[12]

Nor is it uncontroversial that games have constitutive aims. On a personal reading, to say that chess aims at checkmating the opponent is to say that all chess players aim to checkmate their opponents. This seems to imply that someone who plays chess against a young child and aims to let the child win is not really playing chess.[13]

Of course, constitutivists can say that someone who does not aim to win is really playing the related game of *schmess*, which is identical to chess except that players need not aim to win. But why does this constitute a different game? And by what right is this game described as schmess and the first game described as chess?

Moreover, constitutive norms would seem to have a very brittle hold on me if the only thing necessary to escape them is to change my aims and thereby change the activity (Enoch 2006, 2011; Lewis 1996). In the more general case of constitutive norms governing agency, it might be replied that I cannot decide to just schmee a schmagent in the same way I can decide to schmay schmess or schminquire into questions (Ferrero 2009; Velleman 2009). But even if successful, this reply is little comfort to the defender of constitutive norms governing chess or inquiry.

An increasingly popular response is to claim that my reasons to comply with constitutive norms are proportional to my reasons to engage in the relevant activity (Kiesewetter 2017; Way 2012). If it could be seen that I have reason to inquire rather than schminquire, then I might also have reason to pursue the aims of inquiry rather than schminquiry. But since inquiry and schminquiry differ only in their aims, to ask whether I have reason to inquire rather than schminquire just is to ask whether I have reason to pursue the aims of inquiry or schminquiry. This leaves us back where we started.

[12]Following Friedman (2024), it may be helpful to distinguish *activity* verb phrases such as 'walk' from *accomplishment* verb phrases such as 'walk a mile.' As Friedman notes, accomplishment verb phrases pass standard tests for telicity that activity verb phrases fail. We can accept that there is a constitutive aim of walking a mile – namely, to finish the mile – without accepting that there is a constitutive aim of walking.

[13]On a metaphysical reading, we need to make a related objection: why think it is part of the essence of chess that someone who aims to let a child win is not playing chess?

4.3.3 From aims to standards

A second concern is why aims should generate standards of correctness. A natural way to read the datum that games have constitutive aims is as a descriptive claim about what it is to play a game. Perhaps it is true that to play a game you must intend to win, just as to play a game you must intend to follow the rules. But why should we think that goals generate normative standards of success, such that act tokens are successful just in case they achieve the aim?

I am not sure if this second concern is relevant to Bermúdez and Seo's account. Bermúdez and Seo defend a downstream normative claim about inquiry:

> **(Normativity Condition)** N is a norm governing activity A only if there is a direct, intrinsic, and demonstrable connection between following N and achieving the goal(s) of A.

Although many constitutivists would motivate Normativity Condition through an intermediate claim such as Aim-Standard Link, Bermúdez and Seo may not do this. If that is right, then they will avoid concerns about the link between aims and standards of correctness, but face a heightened form of my next concern.

4.3.4 From standards to further normative claims

A third concern for constitutivism is why standards of correctness should generate further normative claims. Chapter 4 argues that there is no distinctively epistemic reading of four normative terms: rationality, justification, and what agents should or ought to do. Some constitutivists may simply wish to make claims about standards of success or correctness for inquiry. Others may aim to capture a type of attributive goodness, such that a good inquiry is one that achieves its success conditions, where 'good inquiry' is not to be analyzed as the application of a domain-general notion of value to the activity of inquiry. None of these claims are in tension with the claims of Chapter 4.

I suspect that Bermúdez and Seo wish to go further. After all, they hold that Normativity Condition is a norm of requirement, but as stated Normativity Condition does not require anything. On a weak reading of Normativity Condition, it is an account of the constitutive norms of activities, with something like the following form:

> **(Weak Normativity Condition)** *N* is a constitutive norm governing activity *A* only if there is a direct, intrinsic and demonstrable connection between following *N* and achieving the goal(s) of *A*.

Weak Normativity Condition will need to be further specified by stating the normative category involved. Is Weak Normativity Condition intended, for example, to characterize claims about how agents ought or are rationally required to act, at least in some internal or constitutive sense? Weak Normativity Condition will also need to be spelled out by characterizing the relevant type of direct, intrinsic and demonstrable connection.[14] Once this is done, we will need an argument for Weak Normativity Condition. Why does Weak Normativity Condition follow from claims about the constitutive aim or standard of correctness for inquiry?

As stated, Normativity Condition seems to go beyond Weak Normativity Condition in making a claim not just about constitutive norms of inquiry, but rather about all norms of inquiry. One way to motivate this stronger claim would be to follow global constitutivists in deriving all norms from constitutive norms (Korsgaard 2009). However, this is more difficult to achieve. Alternatively, Bermúdez and Seo may just be using the phrase 'norm of *A*' to mean 'constitutive norm governing *A*' so that non-constitutive norms governing *A* aren't classified as norms *of A*. This is not my preferred usage, but we may use words however we like so long as we are clear about their meanings.

[14]For example, Bermúdez and Seo hold that eating a sandwich does not directly promote the acquisition of knowledge but only enables agents to be in a better position to promote knowledge. The relevant notion of direct promotion versus enabling will need to be further specified so that preliminary steps of inquiry, such as retrieving information from memory and constructing mental models, do not count as mere enabling conditions. We may also need to think about why the constitutive norms of inquiry are relevant to sandwich-eating in the first place. Why not simply say that eating a sandwich is not inquiring, so the constitutive norms of inquiry have no bearing on the activity of sandwich-eating?

4.3.5 Specific pushback against further normative claims

Many constitutivists have encountered resistance to the specific normative claims they wish to derive from the Aim-Standard Link. The reason for this is that when proposed constitutive norms are weighed against non-constitutive norms, such as promoting value, there is pressure to allow non-constitutive norms to outweigh constitutive norms. At the most extreme, a leading constitutivist account due to Barry Maguire and Jack Woods (2020) has it that constitutive norms answer to a distinctive class of *non-authoritative normative reasons* which have no weight in all-things-considered weighing explanations of what agents ought to do. Nevertheless, Maguire and Woods hold, there is a distinctive type of obligation generated by non-authoritative normative reasons and a corresponding criticism to which agents who act against their non-authoritative reasons are liable.

If the worst that can be said against my view is that it captures all and only the authoritative norms of inquiry, then that may not be so bad. At the same time, I have elsewhere argued that this view is in tension with deeply held platitudes about rationality and reasons (Thorstad 2023a):

> **(Authority of Rationality Platitude)** Rationality is authoritative.
>
> **(Authority of Reasons Platitude)** Reasons of rationality are authoritative and participate in weighing explanations of what agents ought to do and believe.
>
> **(Value Platitude)** Rationality has significant value.

I have argued that leading meta-epistemological justifications for positing a distinctive constitutive species of ought or rational requirement do not meet the threshold of plausibility needed to overturn these platitudes (Thorstad 2023a). I think that similar charges can be made against other accounts, though the details will vary considerably from view to view.

4.3.6 Explanation

The leading constitutivist theory of inquiry is due to Christoph Kelp (2021a; 2021b). Kelp urges us to accept his view on explanatory grounds. Kelp thinks that the correct version of constitutivism explains many things. For example, it unifies epistemology through the theory of inquiry, responds to traditional skeptical arguments, and yields insight into the nature of justification and understanding.

Inquiry under bounds also urges readers to judge its account on explanatory grounds. There and elsewhere, I have argued that my aproach provides a compelling account of bounded rationality (Thorstad 2024b,c); allows us to retain the existence and importance of traditional epistemic norms as well as answering objections to traditional epistemic norms (Thorstad 2021, 2024b); gives plausible accounts of how rationality depends on stakes and tradeoffs (Thorstad 2024b), including novel tradeoffs between coherence and competing goals such as accuracy (Thorstad 2024a) and the benefits of complex cognition (Thorstad 2025); explains why stereotypes differ from rational heuristics (Thorstad 2024b); recovers plausible treatments of zetetic norms such as clutter avoidance and the role of logic in reasoning (Thorstad 2024b, 2023b); explains why material poverty does not breed irrationality (Thorstad 2022, forthcoming); solves recalcitrant paradoxes of bounded rationality (Thorstad 2022); and helps us to understand the nature of and motivation for the zetetic turn as well as its relationship to a broader procedural turn within the study of bounded rationality (Thorstad 2024b, forthcoming).

Readers should make up their own minds about the importance and plausibility of these explanatory applications. The explanatory consequences claimed by some constitutivists, such as Kelp, are quite weighty. Those who agree with Kelp's explanations may find constitutivism to be an attractive alternative to my view. By contrast, those who find the explanatory consequences of my view to be more plausible or weighty should be attracted to my view. To a large extent, Kelp and I agree on how the merits of constitutivism should be judged. It is up to readers to do the judging.

References

Badhwar, Neera. 1987. "Friends as ends in themselves." *Philosophy and Phenomenological Research* 48:1–23.

—. 1991. "Why it is wrong to be always guided by the best: Consequentialism and friendship." *Ethics* 101:483–504.

Berg, Nathan. 2003. "Normative behavioral economics." *Journal of Socio-Economics* 32:411–27.

—. 2014. "The *consistency* and *ecological rationality* approaches to normative bounded rationality." *Journal of Economic Methodology* 21:375–97.

Berker, Selim. 2018. "The explanatory ambitions of moral principles." *Noûs* 53:904–36.

Bermúdez, José. 2009. *Decision theory and rationality*. Oxford University Press.

Brandt, Richard. 1959. *Ethical theory*. Prentice-Hall.

Cherniak, Christopher. 1981. "Minimal rationality." *Mind* 90:161–183.

—. 1986. *Minimal rationality*. MIT Press.

Copp, David. 1997. "Belief, reason and motivation: Michael Smith's 'The moral problem'." *Ethics* 108:33–54.

Davidson, Donald. 1973. "Radical interpretation." *Dialectica* 27:313–328.

Dreier, Jamie. 2011. "In defense of consequentializing." *Oxford Studies in Normative Ethics* 1:97–119.

Driver, Julia. 2001. *Uneasy Virtue*. Cambridge University Press.

—. 2012. *Consequentialism*. Routledge.

Enoch, David. 2006. "Agency, shmagency: Why normativity won't come from what is constitutive of action." *Philosophical Review* 115:169–98.

—. 2011. "Schmagency revisited." In Michael Brady (ed.), *New waves in metaethics*, 208–33. Springer.

Ferrero, Luca. 2009. "Constitutivism and the inescapability of agency." *Oxford Studies in Metaethics* 4:303–33.

Fine, Kit. 2012. "Guide to ground." In Fabrice Correia and Benjamin Schneider (eds.), *Metaphysical grounding: Understanding the structure of reality*, 37–80. Cambridge University Press.

Flores, Carolina and Woodard, Elise. 2023. "Epistemic norms on evidence-gathering." *Philosophical Studies* 180:2547–71.

Friedman, Jane. 2024. "The aim of inquiry?" *Philosophy and Phenomenological Research* 108:506–23.

Good, I.J. 1966. "On the principle of total evidence." *British Journal for the Philosophy of Science* 17:319–321.

Harman, Gilbert. 1986. *Change in view*. MIT Press.

Harsanyi, John. 1977. "Rule utilitarianism and decision theory." *Erkenntnis* 11:25–53.

Houthakker, Hendrik. 1950. "Revealed preference and the utility function." *Economica* 17:159–74.

Howard, Ronald. 1966. "Information value theory." *IEEE Transactions on Systems Science and Cybernetics* 2:22–6.

Kagan, Shelly. 2000. "Evaluative focal points." In Brad Hooker, Elinor Mason, and Dale E. Miller (eds.), *Morality, rules and consequences*, 134–55. Edinburgh University Press.

Katsafanas, Paul. 2013. *Agency and the foundations of ethics: Nietzschean constitutivism*. Oxford University Press.

—. 2018. "Constitutivism about practical reasons." In Daniel Star (ed.), *Oxford Handbook of Reasons and Normativity*, 367–91. Oxford University Press.

Kelly, Thomas. 2011. "Following the argument where it leads." *Philosophical Studies* 154:105–24.

Kelp, Christoph. 2021a. *Inquiry, knowledge, and understanding*. Oxford University Press.

—. 2021b. "Theory of inquiry." *Philosophy and Phenomenological Research* 103:359–84.

Kelp, Christoph and Simion, Mona. 2021. *Sharing knowledge: A functionalist account of assertion*. Cambridge University Press.

Kiesewetter, Benjamin. 2017. *The normativity of rationality*. Oxford University Press.

Korsgaard, Christine. 2009. *Self-constitution: Agency, identity and integrity*. Oxford University Press.

Lerner, Adam. 2018. "The puzzle of pure moral motivation." *Oxford Studies in Metaethics* 13:123–144.

Lewis, David. 1996. "Desire as belief ii." *Mind* 105:303–13.

MacAskill, William, Bykvist, Krister, and Ord, Toby. 2020. *Moral uncertainty*. Oxford University Press.

Machery, Edouard. 2017. *Philosophy within its proper bounds*. Oxford University Press.

Maguire, Barry and Woods, Jack. 2020. "The game of belief." *Philosophical Review* 129:211–49.

Musslick, Sebastian and Masís, Javier. 2023. "Pushing the bounds of bounded optimality and rationality." *Cognitive Science* 47:e13259.

Ord, Toby. 2009. *Beyond action: Applying consequentialism to decision making and motivation*. Ph.D. thesis, University of Oxford.

Parfit, Derek. 1984. *Reasons and persons*. Oxford University Press.

Pettit, Philip and Smith, Michael. 2000. "Global consequentialism." In Brad Hooker, Elinor Mason, and Dale E. Miller (eds.), *Morality, rules and consequences*, 121–133. Edinburgh University Press.

Portmore, Douglas. 2011. *Commonsense consequentialism: Wherein morality meets rationality*. Oxford University Press.

Ross, William David. 1930. *The right and the good*. Clarendon Press.

Samuelson, Paul A. 1938. "A note on the pure theory of consumer's behaviour." *Economica* 5:61–71.

Scanlon, Thomas. 1998. *What we owe to each other*. Harvard University Press.

Seidel, Christian. 2019. "New wave consequentialism: An introduction." In Christian Seidel (ed.), *Consequentialism: New directions, new problems*, 1–30. Oxford University Press.

Smith, Michael. 1994. *The moral problem*. Wiley-Blackwell.

Staffel, Julia. 2020. *Unsettled thoughts: A theory of degrees of rationality*. Oxford University Press.

Stein, Edward. 1996. *Without good reason: The rationality debate in philosophy and cognitive science*. Clarendon Press.

Stigler, George. 1961. "The economics of information." *Journal of Political Economy* 69:213–25.

Thorstad, David. 2021. "Inquiry and the epistemic." *Philosophical Studies* 178:2913–28.

—. 2022. "Two paradoxes of bounded rationality." *Philosophers' Imprint* 22:1–16.

—. 2023a. "Against the newer evidentialists." *Philosophical Studies* 180:3511–3532.

—. 2023b. "Norms of inquiry." *Philosophical Topics* 51:135–60.

—. 2024a. "The accuracy-coherence tradeoff in cognition." *British Journal for the Philosophy of Science* 75:695–715.

—. 2024b. *Inquiry under bounds*. Oxford University Press.

—. 2024c. "Why bounded rationality (in epistemology)?" *Philosophy and Phenomenological Research* 108:396–413.

—. 2025. "The complexity-coherence tradeoff in cognition." *Mind* 134:422–57.

—. forthcoming. "The zetetic turn and the procedural turn." *Journal of Philosophy* forthcoming.

—. ms. "Level separation in epistemology." Unpublished manuscript.

Velleman, David. 2000. *The possibility of practical reason*. Oxford University Press.

—. 2009. *How we get along*. Cambridge University Press.

Way, Jonathan. 2012. "Transmission and the wrong kind of reason." *Ethics* 122:489–515.

Weinberg, Jonathan, Nichols, Shaun, and Stich, Stephen. 2001. "Normativity and epistemic intuitions." *Philosophical Topics* 29:429–60.

Williams, Bernard. 1981. "Persons, character and morality." In *Moral luck: Philosophical Papers 1973-80*, 1–19. Cambridge University Press.