

Additive Theories of Rationality: A Critique

Matthew Boyle

Abstract: Additive theories of rationality, as I use the term, are theories that hold that an account of our capacity to reflect on perceptually-given reasons for belief and desire-based reasons for action can begin with an account of what it is to perceive and desire, in terms that do not presuppose any connection to the capacity to reflect on reasons, and then can add an account of the capacity for rational reflection, conceived as an independent capacity to ‘monitor’ and ‘regulate’ our believing-on-the-basis-of-perception and our acting-on-the-basis-of-desire. I show that a number of recent discussions of human rationality are committed to an additive approach, and I raise two difficulties for this approach, each analogous to a classic problem for Cartesian dualism. The *interaction problem* concerns how capacities conceived as intrinsically independent of the power of reason can interact with this power in what is intuitively the right way. The *unity problem* concerns how an additive theorist can explain a rational subject’s entitlement to conceive of the animal whose perceptual and desiderative life he or she oversees as ‘I’ rather than ‘it’. I argue that these difficulties motivate a general skepticism about the additive approach, and I sketch an alternative, ‘transformative’ framework in which to think about the cognitive and practical capacities of a rational animal.

If the human being had animal *drives*, he could not have that which we now call *reason* in him; for precisely these drives would naturally tear his forces so obscurely towards a single point that no free circle of reflection would arise for him... If the human being had animal *senses*, then he would have no *reason*; for precisely his senses’ strong susceptibility to stimulation, precisely the representations mightily pressing on him through them, would inevitably choke all cold reflectiveness.

— J. G. Herder, *Treatise on the Origin of Languages*

1. Two Conceptions of Human Rationality

If, as many contemporary philosophers still hold, human beings are rational animals, then we share certain cognitive capacities with nonrational animals, but there is also something that sets our cognition apart from theirs.¹ Like nonrational animals, we have capacities to learn about our environment through perception and to act in pursuit of things we desire. Unlike them—so the traditional story goes—we also have the capacity to reflect, in an efficacious way, on the adequacy

of our reasons for believing what perception inclines us to believe and for pursuing what desire inclines us to pursue.

My aim in the present essay is not to defend the traditional distinction between rational and nonrational animals, or the classification of human beings as animals of the former sort, but to highlight a question that arises if we grant this distinction and this classification: the question whether our capacity for rational reflection merely *adds* a further power to the capacities for perception-induced belief formation and desire-driven action that we share with other animals or rather *transforms* the latter powers in a way that makes our perceptual and desiderative capacities essentially different from those of nonrational animals. I believe contemporary philosophers often overlook this question because they assume the first answer is unavoidable. To bring out the importance of the question, it will be necessary to criticize this seemingly unavoidable answer, and to clarify the significance of the 'transformative' alternative.

1.1. *Additive Theories*

The conception of human rationality that I aim to criticize—expressed in a rough and preliminary way—is that the capacity to 'reflect on reasons' for belief and action is a sort of special module that rational minds possess, over and above the modules for accumulating information through perception and for desire-governed action, which we share with nonrational animals.² Our additional rationality module, it is held, gives us the capacity to *monitor* and *regulate* our believing-on-the-basis-of-perception and our acting-on-the-basis-of-desire in ways that nonrational animals cannot, but it does not make our perceiving and desiring themselves essentially different from the perceiving and desiring of any animal. We rational animals perceive and desire in the same sense in which any animal perceives and desires; the power that differentiates our minds is something separate and additional.

I will call views that take this shape *additive theories* of rationality, to mark a significant implication of such views: namely that an account of our minds might begin with an account of what it is to perceive and desire, in terms that do not presuppose the capacity to reflect on reasons, and then supplement this with an account of the 'monitoring' and 'regulating' of belief-on-the-basis-of-perception and action-on-the-basis-of-desire that only rational creatures can perform. In this sense, such theories regard the package of capacities that make us rational as something that might be 'added on' to a mind that already forms an intelligible system apart from this addition.

These characterizations of the explanatory commitments of additive theories need further clarification, and I shall shortly try to make them more precise. First, though, it will be helpful to introduce some quotations to illustrate the sort of view I have in mind. Consider, then, the following pair of remarks, the first from Gareth Evans and the second from David Velleman:

[W]e arrive at conscious perceptual experience when sensory input is not only connected to behavioral dispositions...—perhaps in some phylogenetically more ancient part of the brain—but also serves as the input to a

thinking, concept-applying, and reasoning system; so that the subject's thoughts, plans, and deliberations are also systematically dependent on the informational properties of the input... Of course the thoughts are not epiphenomena; what a conscious subject does depends critically upon his thoughts, and so there must be links between the thinking and concept-applying system, on the one hand, and behavior, on the other... Further, the intelligibility of the system I have described depends on there being a *harmony* between the thoughts and the behavior to which a given sensory state gives rise. (Evans 1982: 158–159)

Suppose that you were charged with the task of designing an autonomous agent, given the design for a mere subject of motivation... [You would not] start from scratch. Rather, you would add practical reason to the existing design for motivated creatures, and you would add it in the form of a mechanism modifying the motivational forces already at work... A creature endowed with such a mechanism would reflect on forces within him that were already capable of producing behavior by themselves, as they do in nonautonomous creatures or in his own nonautonomous behavior. His practical reasoning would be a process of assessing these springs of action and intervening in their operations (Velleman 2000: 11–12)³

Both Evans and Velleman suggest that a central aspect of our ability to engage, as rational creatures, with the world around us (for Evans, our power to learn about the world through perception; for Velleman, our capacity to implement our desires in action) can be thought of as constituted from two components:

- (1) a more primordial system that we share with nonrational creatures (in one case, a perceptual system that adjusts our behavioral dispositions in response to changing sensory inputs; in the other, a motivational system that translates desires for things into behavior directed toward the pursuit of them); and
- (2) a 'reasoning system' that 'monitors' the activities of the more primordial system, 'assesses' the rational warrant for those activities, and 'regulates' these activities in response to its assessments.

As both authors note, these two systems must not merely coexist; they must normally exhibit a certain harmony: states of the primordial system must in general serve as inputs to the reasoning system, so that the thinking of the latter is informed by the condition of the former, and assessments of the reasoning system must in general produce predictable modifications in the primordial system, so that our reasoned judgments make a difference to what we believe and what we do. This requirement of harmony implies that, at least when the two systems are functioning normally, we can think of them as constituting a single total system in a single total state, a state for which we can reserve a special term, as Evans reserves the term 'conscious perceptual experience' for the kind of condition produced when sensory input is received in a way that seamlessly

governs both our behavioral dispositions and our reflective thought. And having coined such a term, we can say that only rational creatures can be in *such* states. But, we should not let this mislead us into thinking that such views recognize a basic difference between rational and merely animal perception or rational and merely animal desire. They recognize a difference, but not a *basic* difference: they hold that when a rational animal perceives something, this consists in its having a perception of a merely animal kind harmoniously integrated with the operations of a distinct 'thinking, concept-applying, and reasoning system', and that when a rational animal desires something, this consists of its having a merely animal desire harmoniously integrated with the operations of a distinct faculty of 'practical reason', conceived as 'a mechanism modifying the motivational forces already at work'.

1.2. Transformative Theories

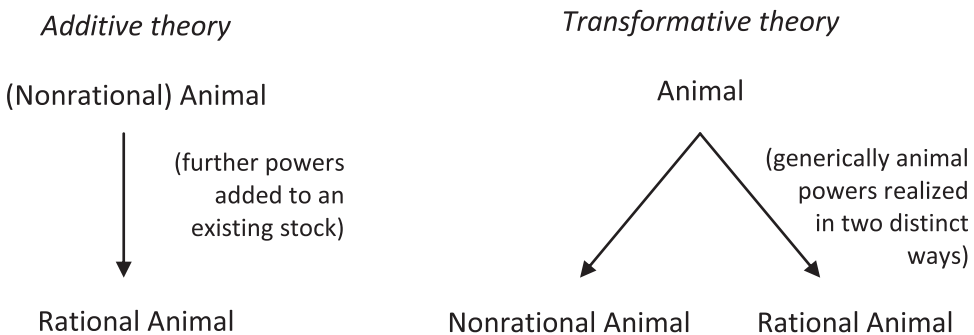
Evans and Velleman express their commitment to the additive approach with uncommon clarity, but I believe this conception of rationality is widespread. Indeed, some philosophers take the additive approach to be inevitable once we admit that we rational beings share our capacities for perception and desiderative motivation with other animals. But the consensus is not universal. Dissent is voiced, for instance, by John McDowell:

If we share perception with mere animals, then of course we have something in common with them. Now there is a temptation to think it must be possible to isolate what we have in common with them by stripping off what is special about us, so as to arrive at a residue that we can recognize as what figures in the perceptual lives of mere animals... But it is not compulsory to attempt to accommodate the combination of something in common and a striking difference in this factorizing way: to suppose our perceptual lives include a core that we can recognize in the perceptual life of a mere animal, and an extra ingredient in addition... Instead we can say that we have what mere animals have, perceptual sensitivity to features of our environment, but we have it in a special form. (McDowell 1994: 64)

On the alternative McDowell proposes, our rationality *does* make a basic difference to the nature of our perceiving: it gives us a 'special form' of perceptual sensitivity to our environment, one whose operations are themselves informed by our capacity to weigh reasons. If this is right, then our power to acquire knowledge from perception cannot be accounted for in an additive framework. For if what 'perception' signifies in the case of rational creatures cannot be explained without reference to the capacity for rational reflection, then rational perceiving cannot be explained as perceiving supplemented by the further power to monitor and regulate this activity in the light of reasoning. Rather, an account of our sort of perceiving must itself appeal to capacities connected with rational thought and judgment. We can thus call the sort of view that McDowell recommends a *transformative theory* of rationality. Such theories take the very nature of perceptual

and desiderative capacities to be transformed by the presence of rationality, in a way that makes rational perceiving and rational desiring essentially different from their merely animal counterparts.

The difference between additive and transformative theories of rationality is not that additive theorists admit, whereas transformative theorists deny, that the minds of rational and nonrational creatures have something in common. As McDowell observes, the dispute is about how to understand this commonality. Additive theorists advocate a certain way of understanding what we have in common with nonrational animals: they hold that there must be a distinguishable *factor* in rational powers of perception and action that is of the very same kind as the factor that wholly constitutes merely animal powers of perception and action. Transformative theorists, by contrast, locate the similarity between rational and nonrational mentality in a different sort of explanatory structure. They hold that rational mentality and nonrational mentality are different *species* of the *genus* of animal mentality. What the two 'have in common', on this view, is not a separable *factor* that is present in both, but a generic *structure* that is realized in different ways in the two cases. Rational and nonrational animals do not share in the sensory and conative powers of *nonrational animals*; they share in the sensory and conative powers of *animals*, where this is a generic category of power that admits of two more specific sorts of realization. So the difference between the explanatory commitments of the two approaches can be diagrammed as follows:



The thesis that our rationality transforms our perceptual capacities captures at least part of the significance of McDowell's well-known claim that the content of our perception is 'conceptual': given McDowell's understanding of what 'conceptual content' is, this characterization implies that the kinds of perceptual episodes that we rational creatures undergo must be characterized in terms that imply a power to reason about the import of such episodes.⁴ I believe, however, that this feature of McDowell's position has not received the attention it deserves. Most of the critical discussion of McDowell's view has focused on various further claims he made about the nature of human perceptual content: that it must be propositional in structure, that everything a given perceptual experience puts a subject in a position to know noninferentially must be somehow written into the 'conceptual content' of that experience, that the specificity of perceptual content can be

captured by appeal to the notion of a 'demonstrative concept', and so on.⁵ McDowell has subsequently changed his mind about some of these points,⁶ but in any case, there were always two parts to his position: on the one hand, a more abstract claim about how our rationality is related to our perceptual capacities, and on the other hand, a set of more specific views about how to think of what perception presents.

Although the former idea is more fundamental, most of McDowell's critics have not directly confronted it. Part of the reason for this, presumably, is that it has not seemed clear what could be at stake in this claim taken by itself: without further theses about the nature of perceptual content to flesh it out, what can it mean to say that our perceptual capacities are 'permeated' by rationality?⁷ I want to suggest, however, that McDowell's point is significant even in this abstract form, because, if correct, it would rule out a widely held and tempting view of the *structure* that an account of rational mentality must exhibit.

1.3. Historical Roots of the Transformative Approach

It is an old idea to think of our rationality as transforming the powers we share with other animals. Aristotle famously thinks of 'rational' as a predicate that differentiates the genus 'animal', and when characterizing the relation of genera to the species that fall under them, he remarks that

by genus I mean that one identical thing which is predicated of both and is differentiated in no merely accidental way... For not only must the common nature attach to the different things, e.g. not only must both be animals, but *this very animality must also be different for each*... For I give the name of 'difference in the genus' to *an otherness which makes the genus itself other*.⁸

Commenting on this passage, Thomas Aquinas remarks that

what the Philosopher says here rules out ... the opinion of those who say that whatever pertains to the nature of the genus does not differ specifically in different species, for example, the opinion that the sensory soul of a man does not differ specifically from that in a horse.⁹

For Aristotle and Aquinas, rational and nonrational animals possess specifically different kinds of 'animality', and because they take animality to imply powers of sense perception and desire-governed action, this appears to commit them to holding that rational animals possess specifically different forms of these powers.

In spite of its long history, I believe this way of thinking of the difference between rational and nonrational mentality is not well recognized by contemporary philosophers. A battery of standard objections are brought against philosophers who profess this sort of view, but these objections are, it seems to me, commonly put forward without a real appreciation of how the transformative view works, and what resources it has for defending itself.¹⁰

A first step toward developing such an appreciation would be to understand what there is to object to in the additive approach. I will present two such objections,

which I will call the *interaction problem* and the *unity problem*. Versions of these difficulties have been raised for additive approaches in particular domains, but I think their generality has not been appreciated. I will argue that these problems do not merely affect this or that particular account, but a whole class of views, and that versions of them can be raised wherever the powers of a rational animal are treated in the additive way.

The problems will turn out, interestingly enough, to be similar to classic problems for Cartesian dualism. I think this is not an accident: although the views about mind and explanation that motivate additive theories are quite different from those that motivate mind-body dualism, the two positions bear a structural similarity to one another. Both views have a sort of horse-and-rider structure, in which one source of activity is set over another. The additive theorist's position is not metaphysically extravagant in the way the Cartesian position is: it does not regard our rational powers as inhering in an immaterial substance. Nevertheless, I will argue, the two positions face analogous difficulties, difficulties grounded in the structural features they share.

2. The Interaction Problem

A first problem for additive theories concerns the nature of the *interaction* they posit between our perceiving and desiring, on the one hand, and our judging and choosing on the other. To bring out this difficulty, it will help to examine the problem McDowell originally raised for the idea that the content of our perception is nonconceptual.

2.1. McDowell on Nonconceptual Content

When McDowell first argued against the idea that a nonconceptual 'given' might play a role in an account of human perceptual knowledge, his basic objection was simply that we cannot make sense of the role this nonconceptual factor is required to play:

[W]e cannot really understand the relations in virtue of which a judgment is warranted except as relations within the space of concepts: relations such as implication or probabilification, which hold between potential exercises of conceptual capacities. The attempt to extend the scope of justificatory relations outside the conceptual sphere cannot do what it is supposed to do. (McDowell 1994: 7)

Any satisfactory account of perception must, McDowell held, explain how our perception can exercise an intelligible 'constraint' on what we judge, so that our enterprise of forming a reasonable view about the world does not appear to be an activity in which we proceed without input, a 'frictionless spinning in a void' (McDowell 1994: 11). But, he claimed, if we try to introduce this constraint by supposing that perception supplies a *nonconceptual* content with which our application

of concepts in judgment must agree, we impose a limitation on this content that makes it unintelligible how it could supply the needed constraint. For what is needed is not just any sort of constraint on the subject's judging; what is needed is something intelligible as a constraint *from the subject's own point of view*—something she could see as a *reason* for judging the world to be thus-and-so, if she were to reflect on the question 'Why should I believe that P?' But, to suppose that the content of perceptual experience is nonconceptual is to conceive of it in a way that rules out its playing such a role.

McDowell's case for these claims turned on a thought about the nature of the capacity to judge, on the one hand, and a thought about the nature of conceptual content, on the other. The capacity to judge, he held, is a capacity for 'spontaneity'—for conscious self-determination in the light of reasons recognized as such. But, a perceptual state has conceptual content just if it

has its content by virtue of the drawing into operation ... of capacities that are genuinely elements in a faculty of spontaneity. The very same capacities must also be able to be exercised in judgments, and that requires them to be rationally linked into a whole system of concepts and conceptions within which their possessor engages in a continuing activity of adjusting her thinking to experience. (McDowell 1994: 46–47)

A perceptual state whose content was nonconceptual, by contrast, would be one whose having its specific content did not engage capacities for such reflection. Hence, McDowell argued, such a state would be one whose impact on judgment would not be open to critical reflection:

[T]he putatively rational relations between experiences, which this position does not conceive as operations of spontaneity, and judgments, which it does conceive as operations of spontaneity, cannot themselves be within the scope of spontaneity—liable to revision, if that were to be what the self-scrutiny of active thinking recommends. And that means that we cannot genuinely recognize the relations as potentially reason-constituting. (McDowell 1994: 52)

The content of a subject's judgment might certainly 'match' such a nonconceptual perceptual content, but her having a perceptual experience with this content could not itself be her reason for so judging. McDowell summed this up by saying that nonconceptual contents could at best supply 'exculpations', not 'justifications', for a subject's judgments: they could at best make it explicable, and thus excusable, that the subject judged as she did; they could not constitute the reason she saw for so judging.¹¹

McDowell made his case for this conclusion in an evocative but somewhat metaphorical idiom, and it is not easy to understand exactly how the argument is supposed to work. A natural objection, raised by a number of authors, is that McDowell's argument trades on a non-sequitur.¹² Even if we grant both that a rational subject must be able to reflect on the probative force of her own reasons for judgment and that such *reflection* must draw on conceptual capacities, it does

not obviously follow that the mere *having* of the perception whose rational significance is assessed must draw on such capacities. Why could my capacity to perceive not present me with representations whose content was not *intrinsically* conceptual, but whose rational significance I was able to consider in virtue of my possession of a further, distinct capacity for conceptual thought and judgment?

We will be in a better position to assess this objection once we have a clearer understanding of the underlying structure of McDowell's argument. I will return to this topic in the succeeding text (Section 2.4), after drawing two comparisons that should help to clarify the nature McDowell's objection to nonconceptualism. For the moment, let me simply make two observations about the sketch of the argument just given.

First, the argument does not appear to depend on specific commitments about the nature of conceptual contents. 'Conceptual content' figures simply as a term for content the attribution of which implies the engagement of certain capacities—the capacities that enable us to reflect on our reasons for belief and form a considered judgment about their cogency.¹³ McDowell's ambition is to raise a difficulty for any view that does not recognize the operation of such capacities in our perception itself but still maintains that perception give us reasons for judgment. In other words, his ambition is to raise a difficulty for any account that treats the cognitive powers of a rational perceiver in an additive way: as consisting of a not essentially rational power to perceive, whose acts of perception are inputs to a further and independent power to make reflective judgments.

Secondly, the strategy of the argument is to raise a problem about how, if our perceiving does not itself draw on conceptual capacities, our perceptions can *explain* our judgments in the right way. Some of McDowell's phrasing does not foreground this question of explanation, as when he asks how 'relations in virtue of which a judgment is warranted' such as 'implication and probabilification' can hold between perception and judgment. But a careful reader of the surrounding text will know that McDowell's concern is not merely that perception should supply reasons for judgment in the sense in which R might be a reason for me to judge J although I am quite unaware of R and hence in no position to take it into account in judging. His demand is that an account of perception should make it intelligible how a subject's perceiving something can be *her reason for* a certain judgment—a reason her apprehension of which explains her so judging.¹⁴ His strategy is to argue that a view on which perceiving does not 'actualize conceptual capacities' cannot account for this special kind of explanatory dependence: it can at best represent the subject's judgment as depending on her perceptual experience in an exculpatory way.

2.2. Descartes on Mind–Body Interaction

To clarify the structure of the problem McDowell raises for nonconceptualism, it will be useful to compare it to a venerable objection to Cartesian dualism, one

that has been discussed ever since Princess Elisabeth of Bohemia wrote to Descartes to ask

how the soul of a human being (it being only a thinking substance) can determine the bodily spirits, in order to bring about voluntary actions. For it seems that all determination of movement happens through the impulsion of the thing moved, by the manner in which it is pushed by that which moves it, or else by the particular qualities and shape of the surface of the latter. Physical contact is required for the first two conditions, extension for the third. [Yet] you entirely exclude the one [extension] from the notion you have of the soul, and the other [physical contact] appears to me incompatible with an immaterial thing.¹⁵

The difficulty here can be stated without appeal to the specific view of physical causation presupposed in Elisabeth's remark. The general problem is simply that the Cartesian view posits an explanatory relation between *relata* whose natures seem to exclude their standing in any such relation. Bodies are extended things, and Elisabeth takes Descartes to admit that the realm of extended things forms a closed explanatory system, such that any event that occurs in it can be sufficiently explained by physical laws together with facts about other, earlier events that have occurred in that realm. The mind, by contrast, is a non-extended thing, and Descartes famously holds that its properties are in principle wholly independent of facts about the disposition of bodies in space, in such a way that things might proceed just as they do with minds even if there were no bodies and might proceed just as they do with bodies even if there no were minds. Yet our minds are supposed to move our bodies, which appears to require that changes in minds should be capable of causing changes in material bodies that would not otherwise occur. Elisabeth's question is how this sort of causal relation is possible. If the two realms are intrinsically independent of one another, how can a change in one necessitate a change in the other? And if we suppose that changes in the mental realm can influence the behavior of bodies, how can this influence be anything but disruptive of the lawful order of physical causation?

There has been extensive discussion of Elisabeth's challenge and how Descartes might respond to it.¹⁶ For present purposes, however, this sketch will suffice, because my concern is only to bring out a broad, structural analogy between Elisabeth's objection to Descartes and McDowell's objection to the nonconceptualist. In each case, the difficulty concerns how to make sense of a sort of influence that plainly occurs (the shaping of bodily movement by will in the one case, the constraining of judgment by perception in the other) without compromising a basic commitment about the mode of explanation proper to a certain realm of phenomena. Just as Elisabeth assumes that an account of non-disruptive influence on a material body must trace this influence to contact with some other material body, so McDowell assumes that an account of non-disruptive influence on a person's judgment must trace this influence to the presentation of a reason on whose adequacy the judging subject can reflect. The difficulty is to see how any influencing factor that was not

of this sort could operate on the system without disrupting the explanatory order that characterizes its normal operation: a source of bodily movements that do not accord with the laws of nature, or an 'exculpation' for judgments that are not made on the basis of recognized grounds. But *this* is not the sort of influence we aimed to describe; our aim was to explain how the factor in question could *non-disruptively* influence the system in question.

When I speak of an *interaction problem*, this is the sort of difficulty I have in mind. Such a problem will arise for any view that posits a situation with the following structure:

- | | |
|----------------------------|---|
| (Normal Explanation) | For any fact F of type T ₁ , a normal explanation of F must appeal to a fact that relates to F in way W. |
| (Non-disruptive Influence) | Facts of type T ₂ can normally explain facts of type T ₁ . |
| (System Externality) | Facts of type T ₂ do not relate to facts of type T ₁ in way W. |

A view that is committed to versions of these three theses is committed to an incoherent position. But this, I will argue, is exactly the sort of situation that additive theories of rationality characteristically produce: one in which a certain system is supposed to be non-disruptively influenced by a power whose operations are conceived in such a way that they could only influence the system by disrupting it.

Now, there is obviously a very significant difference between Descartes's position and the position advocated by additive theorists. Whereas Descartes held that mind and body are distinct *substances*, each of which could exist without the other, additive theorists posit not distinct substances but distinct *capacities*, and they are only concerned to assert an independence in one direction: they claim that it is possible to explain what is involved in our possessing the generically animal capacities for perception and motivationally efficacious desire without appealing to the specifically rational capacity for reflective judgment.¹⁷ Nevertheless, I will argue, additive theories characteristically make commitments that produce an interaction problem. To show how these commitments arise, and that they are not easily avoidable, I will first consider a problem for additive theories of the relation between desire and reflective choice and then return to McDowell's problem about the relation between perception and reflective judgment.

2.3. *The Influence of Desire on Reflective Choice*

To clarify what an additive theory of the relation between desire and choice might look like, it will help to consider the idea, endorsed by a number of recent authors, that the special and distinctive power of a rational *agent* is the capacity to 'step back' from his or her desires to act. This idea has been given an especially vivid expression by Christine Korsgaard:

A lower animal's attention is fixed on the world. Its perceptions are its beliefs and its desires are its will. It is engaged in conscious activities, but it is not conscious of them... But we human animals turn our attention on to our perceptions and desires themselves, on to our own mental

activities, and we are conscious of them. That is why we can think *about* them... I desire and I find myself with a powerful impulse to act. But I back up and bring that impulse into view and then I have a certain distance. Now the impulse doesn't dominate me and now I have a problem. Shall I act? Is this desire really a *reason* to act? (Korsgaard 1996: 92–3)

Korsgaard's metaphor of 'backing up' from one's own desires admits of various interpretations, but on one natural reading, it suggests a view with an additive structure: one on which the power to reflect on one's own desires is added, in a rational creature, to a merely animal power to be impelled by desire, in such a way that this addition does not alter the nature of the desiring itself but merely allows us to make certain sorts of assessments of and interventions in the animal desiderative system, encouraging some 'impulses' and thwarting others.

I should immediately add that I do not take Korsgaard herself to conceive of the relation between reason and desire in this way (although I think she does not make this sufficiently clear).¹⁸ For my purposes, however, it is not necessary to attribute such a view to any particular author: the step from Korsgaard's metaphor to an additive theory is, at any rate, an intelligible and tempting one. The idea that we are subject in the first instance to 'brute impulses', but that reason gives us the power to scrutinize and govern these impulses, embodies a venerable and appealing picture of human motivation. But, appealing though this may be as a picture, I want to suggest that it is unacceptable as a literal theory of the relation between reason and desire. For if taken literally, it generates an interaction problem that renders it mysterious how a rational agent's desires can provide her with reasons to act.

To see this, notice first that the same kinds of considerations that make it attractive to conceive of rational judgment as an exercise of 'spontaneity' make it attractive to conceive of choice in a similar way. Just as a subject has the capacity for reflective judgment only if she has the capacity to accept propositions for reasons she can scrutinize and accept or reject, so a subject has the capacity for reflective choice only if she has the capacity to adopt aims for reasons whose sufficiency she can likewise scrutinize. When a subject has reasons for making a certain choice, these will normally be reasons that are open to scrutiny in this sense, and she will choose what she does because she regards it as desirable in the light of these reasons. This gives us a version of

(Normal Explanation _D)	For any choice C of a rational subject S, a normal explanation of C must appeal to reasons grasped by S and regarded as making C desirable.
------------------------------------	---

Now, it is also natural to assume that our desiring something can at least sometimes present us with a reason to choose to pursue that thing. Korsgaard's own phrasing reflects this assumption: in speaking of my reflectively asking myself whether a certain desire 'is really a *reason* to act', she implicitly presupposes that my desiring X *can* present me with a reason in the way that canonically explains choice, even if not all desires actually present me with such reasons. If we assume this, we are committed to

(Non-disruptive Influence_D) The fact that a rational subject S desires E can normally explain S's choosing to pursue E.

Additive theorists must hold, however, that a rational subject's faculty of desire is not intrinsically different from the desiderative faculty of a nonrational animal, one lacking the capacity to reflect on reasons for pursuing ends. Hence, they must hold that a rational subject's desiring E does not itself involve E being presented to her *as* something there is reason to pursue. By this I mean not merely that the subject's desiring E must not involve her having representations of the form

($\exists R$) R is a reason to pursue E

This much is surely true, because by hypothesis, the subject's desiring E cannot involve anything that could not occur in the life of a nonrational animal, which presumably lacks concepts like *__ is a reason to pursue __*. The deeper point, however, concerns the manner in which the object of desire itself is presented. Even when a rational subject does not explicitly believe that there is a reason to pursue a certain end, it is attractive to think of her desires as (at least normally) presenting ends as *desirable*.¹⁹ To characterize our desires as presenting ends 'as desirable' is not to specify a further element of the content represented by the desiring subject but simply to make explicit the specific commendatory mode of presentation that characterizes our desires themselves: they normally present their objects as things that would, other things equal, merit desire (and indeed, attainment). This too is implicit in Korsgaard's characterization: in supposing that it is possible for desires to 'withstanding reflective scrutiny' (Korsgaard, 1996: 93 et pass.), she implies that, even prior to reflection, a desire *recommends* its object for endorsement.

The difficulty, however, is to see how a rational subject's desires *could* present their objects in this way if they were—as additive theorists must suppose—not intrinsically different from the desires of a nonrational animal. To present an object as desirable is to present it as meeting a certain kind of standard, a standard that is explicitly invoked when a rational agent asks herself Korsgaard's 'normative question': 'Is this desire really a reason to act?' But, a nonrational animal has, by hypothesis, no cognizance of this standard: the objects it desires may in fact be things it has reason to pursue, but their meeting this standard is not itself a fact within its ken. So while it may be correct to say that a nonrational animal's desires present their objects as *attractive* (for instance, as promising pleasure or promising to relieve some distress), it cannot be correct to say more specifically that they present their objects as *desirable* (i.e., as *meriting* desire): this way for something to be attractive lies beyond the scope of a nonrational mind. But then additive theorists must hold that, even in the case of rational animals, desiring E does not itself involve E's being presented as desirable. Additive theorists are therefore prohibited by the structure of their project from conceiving of the desiderative capacity of a rational subject as itself a capacity for *prima facie* rational assessment of ends, a capacity actualized in presentations of certain ends as worth pursuing. The representation of ends as meriting pursuit must be introduced by the actualization of a further

faculty of practical rationality, conceived—as Velleman puts it—as ‘a mechanism modifying the motivational forces already at work’.

The idea that the relevant forces are ‘already at work’ expresses the basic commitment of the additive theory: that the desires monitored and regulated by practical rationality can be conceived as elements that might exist and play their motivational role in the absence of any capacity for rational assessment of ends. The problem is to see how such a desire could present the subject with a *prima facie* reason to pursue an end. If it could not, then the additive approach implies

(System Externality_D) A rational subject S’s desiring E does not present S with a reason that S can grasp and regard as making E desirable.

And then we have a version of the interaction problem.

To see how the additive approach leads to System Externality_D, suppose for the sake of argument that our desiring E does not normally involve E’s being presented as desirable. The question to consider is how such a desire can be a presentation of an apparent reason for pursuing E, rather than a mere fact about the situation with which the subject has to cope in deciding what to do. There is a palpable contrast between a normal human desire for an end and, for example, a feeling of nausea that disposes one to vomit. A feeling of nausea does not present an ostensible reason for vomiting: it presents this result, not as *prima facie* reasonable to pursue, but—if anything—as tending to become unavoidable. One indication of the difference between nausea and ordinary desire is this: if I can relieve my nausea not by vomiting but by taking a pill that alleviates the nausea, then—absent some independent reason to believe that it is important for me to vomit—I take myself to have as much if not more reason to take the pill. Not so for ordinary desire: perhaps I might eliminate my desire for E by taking a pill, but unless I subscribe to some stoical philosophy that rejects the claims of desire on general grounds, I do not take this outcome to be one I have just as much presumptive reason to pursue as the outcome in which I obtain E. A normal human desire presents its object as *prima facie* to-be-pursued in a way that nausea does not present vomiting. But it is not clear how an additive theory of the relation between reason and desire can account for this presumptive reasonableness of ordinary desire.

There is a well-known hypothetical example, originally due to Warren Quinn, that helps to highlight the difficulty here. Quinn asks us to imagine a person with a brute impulse to turn on radios whenever he sees them: not in order to hear the broadcast, or for any other purpose, but unaccountably (perhaps as the result of hypnotic suggestion or of some scientist’s having established odd connections among his brain synapses). Imagining himself to be such a person, Quinn remarks:

I cannot see how this bizarre functional state in itself gives me even a *prima facie* reason to turn on radios... It may help explain, causally, why I turn on a particular radio, but it does not make the act sensible, except in so far as

resisting the attendant disposition is painful and giving in pleasant. But in that case it is not the present state that is the reason but the future prospect of relief. (Quinn 1994: 237)

Quinn constructs this example with a view to arguing against the Humean claim that my having a brute desire for E by itself constitutes my having a reason to pursue E. Considered in the present context, however, I think the example suggests that a certain sort of ‘impulse’ would not even be a normal human *desire*.²⁰ A normal human desire presents its object as *prima facie to-be-pursued*, where this means something like *prima facie meriting the endorsement of rational reflection*. Reflection may of course overrule immediate desire, but if our account of desire does not make intelligible how our desires can, so to speak, present a verdict that our reason must recognize as at least presumptively significant, then we lose the intuitive contrast between ordinary desires and Quinn-ish impulses to turn on radios. But a view on which my desiring E does not normally involve E’s being presented as desirable could, it seems, only represent desire in this way: as a disposition to pursue E that did itself not engage my sense of what there is reason to do.

An additive theorist might reply that, although her approach must hold that a rational subject S’s desiring E does not *itself* involve its appearing to S that there is reason to pursue E, the approach can still allow for an *extrinsic* relationship between desiring and being presented with a reason. Might there not be some linking disposition, not belonging to S’s capacity for desire itself, but nevertheless normally present in rational subjects, which ensures that normally, when S desires E, E appears worth pursuing to S?²¹

Well, if this disposition is to operate in a way consistent with the thought that a rational subject must be able to ‘back up’ from her dispositions to be moved and scrutinize their rational basis—if, to put matters in McDowellian terms, it is to explain her *prima facie* reason assessments not merely in an ‘exculpatory’ but in a ‘justifying’ way—then the relevant disposition must be one she can rationally consider and whose continued operation depends on her acceptance of it as sound. We can capture this requirement by insisting that the relevant disposition is grounded in a rationally reviewable *belief* of the subject, a belief to the effect that

(D) My desires normally direct me toward kinds of objects that are desirable.

If a person believed something like (D), then even if her desiring E did not itself involve E’s being presented as worth pursuing, still she might intelligibly regard her desire for E as making it *prima facie* reasonable for her to pursue E. The connection between her desiring and her sense of what there was reason to do would be established, not intrinsically through the nature of her desiring itself, but extrinsically through a belief about the normal connection between her desires and her reasons for choice.

This is indeed a possible way in which the fact that I desire E might come to bear on my reasoning about whether to pursue E. We should, however, note two points about this sort of connection between desire and reason. First, this proposal seems intuitively wrong as a characterization of the *primary* relation between our normal

desires and our sense of what there is reason to do. It is true that I might suppose I have reason to pursue E in virtue of feeling a desire for E and holding a further belief like (D). Perhaps in cases where I am in doubt about whether a certain desire is worth fulfilling, I might reinforce my sense that it is by recalling such a conviction. But the claim of my desires on my rational attention does not normally seem to be mediated in this way. In the normal case, to desire E just is to find the prospect of obtaining E presenting itself as *choiceworthy* in some respect. A desire for E of which this was not true would be, in this respect, akin to nausea: it would be an impulse toward E about which I might believe, as a further, independent conviction, that it should not be resisted, but which did not itself present E as *prima facie* desirable. But, as we have seen, human desires are not normally like that.

Secondly, it is open to question whether this proposal *could* characterize the primary relationship between our desiderative and our rational faculties, for this proposal bears a structural resemblance to a familiar sort of 'foundationalist' view about the relationship between perceptual appearances and reasons for belief. On such a view, the primary rational bearing of perception on belief must be established through my having a non-question-begging reason to believe that

- (P) Having a perceptual appearance as of X's being F normally presents me with a reason to believe that X is F.

where a reason for (P) is non-question-begging only if it does not itself presuppose that my perceptions normally give me reasons for corresponding beliefs. This is not the place for a detailed examination of the prospects of such a foundationalist project, but the well-known objections to such projects give us some grounds for skepticism on this point.²² Likewise, I suggest we should be skeptical about whether we could 'bootstrap' ourselves into a conviction in the rational bearing of desire on choice if, in the basic case, our desiring E did not itself involve E's being presented to us as desirable.

It thus appears that additive theorists are committed to System Externality_D. And then, unless they are willing to reject either Normal Explanation_D or Non-disruptive Influence_D, they face an interaction problem. For a philosopher who finds Normal Explanation_D and Non-disruptive Influence_D attractive, on the other hand, the best way out is to reject the additive approach.

2.4. McDowell's Objection Revisited

Let us now return to McDowell's objection to nonconceptualism about perceptual content and consider how it resembles the other interaction problems we have examined.

McDowell's reason for claiming that the content of our perception must be conceptual, we saw, was that otherwise episodes of perception could place only the wrong kind of constraint on our acts of judgment. We can schematize his argument as follows:

- (1) For any judgment J of a rational subject S, a normal explanation of S's judging J must appeal to reasons available to S's reflective scrutiny.
 - (2) The fact that a rational subject has a perceptual experience of some object O's being F can normally explain S's judging O to be F.
 - (3) If the content of perceptual experience were nonconceptual, it could not present a rational subject with reasons available to her reflective scrutiny.
- So (4) the content of perceptual experience cannot be nonconceptual (not insofar as it is to figures in normal explanations of judgment, at any rate).

Premise (1) articulates McDowell's conception of what distinguishes the capacity for reflective judgment from the capacity for merely 'instinctive' belief: a subject who can judge must be one who can scrutinize her own reasons for belief and reflect on their cogency. It should be clear that this premise amounts to a version of Canonical Explanation for the present case. Premise (2), which states the corresponding version of Non-disruptive Influence, captures what is at stake in McDowell's claim that perception 'rationally constrains' judgment. To deny either premise would be, in effect, to reject the terms of the debate between McDowell and his nonconceptualist opponents, so for present purposes, I will assume that neither (1) nor (2) is in dispute.

The premise characteristically rejected by McDowell's nonconceptualist opponents is (3): that if the content of perceptual experience was nonconceptual, it could not supply reasons available to the reflective scrutiny of a rational subject.²³ This premise amounts to a version of System Externality, and the resulting interaction problem is the engine of McDowell's objection to nonconceptualism. A common objection to (3) is that McDowell fails to distinguish between the capacities that enable us to have perceptual experiences that present us with reasons and the capacities that enable us to reflect on those reasons.²⁴ Critics who make this objection grant that conceptual capacities may be required in order to *reflect on* the reasons supplied by our perceptual experiences, but they maintain that it does not follow that such capacities are required for us simply to *have* perceptual experiences that supply the relevant reasons. If this is right, there appears to be space to admit both that perception supplies us with reasons on which we can reflect, and that reflection calls on conceptual capacities, without granting (3).

This response leaves us, however, with a problem that should by now feel familiar. The reasons with which our perceptual experience supplies us are supposed to be ones whose availability does not itself involve an actualization of the capacities we exercise in assessing reasons for judgment. When we reflect on such a reason, this is supposed to consist in our *conceptualizing* a content that is intrinsically nonconceptual. But now consider this supposed act of conceptualizing. To say that a certain reflective thought 'conceptualizes' a certain perceptual content involves positing some sort of dependence of the thought on the perception, but what sort of dependence can this be? Can I, who conceptualize my perceptual experience in a certain way, reflect on this act and see a reason for so conceptualizing, or can I not?

It does not seem that I *can* see a reason for conceptualizing as I do, for what could this reason be? It cannot be the very reason the perceptual state itself was supposed to supply, for my ability to reflect on this reason was supposed to be the upshot of my act of conceptualizing and so cannot be available to me as something I can see as my ground for this very act. And surely there is no other candidate reason in the picture: to try to insert one would initiate a regress.

But if my act of conceptualizing is one for which I *cannot* see a reason, how can I regard the reflective thought that is the upshot of this supposed act as a potential justifier of judgments? This reflective thought will make a claim about what I am perceiving, or how things perceptually appear to me, and this claim will itself be something I take to be true. On the proposal under consideration, I am correct to take this to be true just in case I have had a perceptual experience with a certain (nonconceptual) content. But my having had this experience cannot be *my* ground for thinking what I do, for, by hypothesis, this ground only becomes available to my reflection via this very thought. Then what can my reason be for thinking myself to have had a perceptual experience with a certain content? If my thought needs a ground, but I cannot see any ground for it, then it is not clear how I can regard it as giving me a reason for any further judgment. But if it does not need a ground, then the supposed nonconceptual content of my perceptual state drops out of the picture as rationally irrelevant. In either case, the supposed nonconceptual content cannot supply me with a reason for judgment.

One defender of nonconceptualism who has discussed this sort of defense of McDowell's position is Richard Heck (2000). According to Heck, this response unfairly saddles the nonconceptualist with an unreasonable view of perceptual epistemology, one on which perceptually based judgments about how the world is must rest on judgments about how things perceptually appear to me. Heck replies that

we need no epistemic intermediary between our perceptions and our beliefs, and it would not help if we had one: if there is a problem about how I can form justified beliefs about the world on the basis of my perceptions—one allegedly solved by letting me form them on the basis of judgments about how things appear to me—why is there not a similar problem about how I can form justified beliefs about how things *appear*. If judgments about how things appear can justifiably be made without any intermediary, why can't judgments about how things are justifiably be made without one too. (Heck 2000: 517–518)

I think this is an attractive position in its own right; the question is whether it is available to nonconceptualists, given their other commitments. In granting the first two premises of McDowell's argument, nonconceptualists grant that a rational subject must be able to advert to what her perception presents when she reflects on her reasons for judging that things in her environment are thus-and-so. When she reflects in this way, she will be making another judgment, one about how things are presented in her perceptual experience.²⁵ But on the nonconceptualist's view, the subject has a reason for making such a judgment just if she is in a certain nonconceptual perceptual state. Now, a rational subject will

be able to ask *herself* what reason she has for making the judgment in question, as she can ask this about any judgment she makes. But if her perceptual state is nonconceptual, then, I have argued, the relevant reason is not available to her. No doubt it would be better to hold that her judgment does not stand in need of such a ground, but I cannot see how the position taken by nonconceptualists permits them to say this.

I conclude that nonconceptualism about perceptual content implies

(System Externality_P) A rational subject's perception does not present her with reasons available to her reflective scrutiny.

Hence, unless nonconceptualists are willing to reject either

(Canonical Explanation_P) For any judgment J of a rational subject S, a normal explanation of S's judging J must appeal to reasons available to S's reflective scrutiny.

or

(Non-disruptive Influence_P) The fact that a rational subject has a perceptual experience of some object O's being F can normally explain S's judging O to be F.

they face a version of the interaction problem. For philosophers who find Canonical Explanation_P and Non-disruptive Influence_P attractive, however, the natural way out is to reject the additive conception of perceptual rationality that gives rise to System Externality_P.

In several respects, this conclusion is modest. It does not address the question how we *should* conceive of the reasons that perception gives us for judgment. The claim that a rational subject's perception must have 'conceptual content' has figured in my discussion only as an abbreviation for the claim that her perceiving must, inasmuch as it presents her with reasons, 'actualize conceptual capacities', where the latter phrase means simply that her perceiving itself draws on the capacities that enable her to reflect on reasons for judgment. How exactly to characterize the role of these capacities in informing the perceptual experience of a rational subject is beyond the scope of my discussion here; I have only been defending an abstract thesis about the order of explanation that an account of the relation between perception and reason must follow.

My conclusion is also modest in another respect. It does not rule out—and it is no part of my agenda to deny the value of—describing forms of perceptual content whose presence bears no connection to the actualization of the subject's conceptual capacities. There is, indeed, a rich body of work in the cognitive science of vision that does just this. The value of positing such forms of perceptual content (e.g., Marr's '2½-D Sketches' and '3D Models') is to be assessed by reference to the explanatory success of the theories that posit them: no a priori argument rules them out. But ruling them out is no part of my agenda here.²⁶ My question concerns the grounds for judgment that perception makes *available* to us in a specific sense: the ones that we can reflectively consider and take as our reasons for judging as we do.

It seems clear that perception can supply us with such reasons, as comes out in our readiness to point to our having perceived something when asked why we accept a certain proposition. *This* notion of what perception makes available belongs not to a theory of perception to be judged by its explanatory success but to the internal standpoint of the judger herself. The question I have sought to address is whether explaining how perception can play this role requires relating the capacity to perceive to capacities that enable us to reflect on reasons. This is the question at issue in the dispute between additive and transformative theories.

3. The Unity Problem

The preceding section sought to raise a problem about how additive theorists can account for the *interaction* between our rational capacities and our capacities for perception and desire. I illustrated the problem in a few cases, but I believe it can be raised for any additive theory. For any additive theory will be forced to posit a relation of explanatory dependence that is subject to versions of the difficulties I have been raising. The names for such relations are manifold: ‘monitoring’, ‘basing upon’, ‘conceptualizing’, ‘intervening’, ‘blocking’, ‘reinforcing’, ‘redirecting’, and so on. But the structure of the problem will be the same.

Before closing, I want to mention a second difficulty for additive theories, which I will call the *unity problem*. My presentation of this difficulty will be considerably more schematic than my discussion of the interaction problem. My aim in presenting it is not primarily to strengthen the case against additive theories, but to shed further light on the source of the difficulties we have already encountered, and to extend the comparison between additive theories of rationality and dualistic accounts of the relation between mind and body.

3.1. A Unity Problem for Descartes’s Dualism

Again, it will help to recall a classic objection to Cartesian dualism: that, because Cartesians hold that mind and body are ‘really distinct’ (i.e., are distinct substances, each of which could exist in the absence of the other), the Cartesian view cannot account for the *unity* of mind and body that we all know a living human being to be. Arnauld raised such a difficulty for Descartes in the ‘Fourth Objections’ to the *Meditations*:

It seems to me, moreover, that the argument [viz., Descartes argument for the conclusion that I am a thinking thing] proves too much, and takes us back to the Platonic view (which M. Descartes nonetheless rejects) that nothing corporeal belongs to our essence, so that man is merely a rational soul and the body merely a vehicle for the soul – a view which gives rise to the definition of man as ‘a soul which makes use of a body’.²⁷

The (reputedly) Platonic position that I am a soul using a body was much discussed in scholastic philosophy, and the slogan 'I am in my body as a sailor in a ship' was taken to epitomize the position.²⁸ Descartes repudiates this position in the Sixth Meditation, where he insists that he is *not* in his body as a sailor in a ship.²⁹ Arnauld's question, however, is whether Descartes is entitled to say this, given his other claims about mind and body. To say that mind and body are distinct substances is to say that each can in principle exist in its own right, in the absence of the other. But, Arnauld objects, if they are intrinsically two, it is not clear how the composite they are supposed to form can be genuinely one. Moreover, if I am a thinking thing, and mind and body are really distinct, it seems that I can only be a mind that makes use of a body.

Descartes replies to Arnauld that

[i]t is... possible to call a substance incomplete in the sense that, although it has nothing incomplete about it *qua* substance, it is incomplete in so far as it is referred to some other substance in conjunction with which it forms something which is a unity in its own right.³⁰

That is, although neither mind nor body is incomplete in itself, still we can say that each is incomplete inasmuch as each by nature belongs to a third substance, a human being, which is a unity 'in its own right' consisting of a mind and a body. The difficulty, however, is to see what can entitle Descartes to hold that a human being is a unity in its own right. Something that is a unity in its own right (or *per se*) is standardly contrasted in scholastic philosophy with something that is a unity *per accidens*, that is, a unity of things that do not belong together in virtue of their essential natures. A heap of stones is a unity *per accidens*, for it is composed of distinct existences, and nothing in the nature of these several existences implies that there must be such a heap. A man who is white is also, in another way, a unity *per accidens*, for being white is accidental to him *qua* man, because it is not in the nature of man as man to be this color as opposed to another. By contrast, a man who is an animal is a *per se* unity, because it belongs to the essence of human beings to be animals. And the organs of a living human body form a *per se* unity, because (according to Aristotle and the Scholastics who follow him) they cannot genuinely exist apart from the whole they form: a hand severed from a living body is a hand 'in name only', as is shown by the fact that it can no longer perform the characteristic functions of a hand, and by the fact that, soon enough, it decomposes.

If human beings are composites of mind and body, however, and if these are substances in their own right, then it is hard to see how human beings themselves can be unities in their own right: their existence as wholes is not prior but posterior to the existence of the parts from which they are composed. The difficulty facing this Cartesian view becomes particularly acute when we ask 'What am I?': for if mind and body do not form a *per se* unity, and if I am the thing that thinks, then it seems that I cannot also *be* another, distinct thing, my body. I must rather stand to my body as a sailor to his ship.

I believe additive theorists face a partial analog of this problem. The problem is not strictly analogous to the one facing Descartes, because additive theorists do not

take our animal powers and our rational powers to constitute two independent entities, each capable of existing in its own right. Rather, they suppose that our animal powers can exist independently of our rational powers, but not conversely: rational powers are taken to be powers precisely to 'step back' from, 'monitor', and 'intervene in' the operations of our animal powers. So the apt comparison in this case is not a heap of stones, whose constituents are all independent existences, but a man who is white: just as whiteness cannot exist apart from a substance in which it inheres, but it does not belong essentially to man to be white, so our rationality cannot exist apart from an animal whose life it oversees, but it does not belong essentially to any animal *qua* animal to be rational.³¹ Nevertheless, I want to suggest, this is still a kind of unity *per accidens*, and difficulties analogous to those facing Descartes's position can be raised for views that conceive of our rationality as united with our animality only in this way.

3.2. Additive Theories and the Problem of Unity

As in the Cartesian case, the way to bring out the problem is to ask: What am I? Our problem will be to reconcile two natural answers to this question. The first we may call the *Aristotelian response*: I am a certain kind of animal (namely a rational one). The second we may call the *Cartesian response*: I am the thing that thinks. To be sure, these responses are not obviously inconsistent: the thing that thinks might well *be* a certain kind of animal. But we shall see that the additive theories render this commonsense reconciliation surprisingly problematic.

To see the difficulty, consider a subject who feels an immediate desire for X but who reflects and chooses not to pursue X.³² She might express her decision—a little stiltedly—by thinking to herself:

(C) I choose not to pursue X.

The occurrence of 'I' in (C) evidently refers to the subject who thinks (C). In asserting (C), the *thinker* has expressed her decision about what to do. But has she thereby expressed the decision of the *animal* whose life she governs? Is the subject of whom this choice is predicated the very same subject to whom the animal activities of perception and desire are to be ascribed?

On the additive view, our animal capacities for perception and desire are not themselves capacities whose actualization involves the actualization of our rational capacities. They are capacities of a kind that might be present in a nonrational animal, one whose 'perceptions are its beliefs and [whose] desires are its will' (to borrow Korsgaard's phrase from the passage quoted earlier).³³ Such an animal would not, by hypothesis, have the capacity to reflect on reasons, but it would surely have a point of view, including beliefs about what is so and desires bearing on what to pursue. The question to consider is how the additive theorist can explain the fact that, in our own case, although capacities of the very same kind are present, they constitute no point of view distinct from the rational one. It is,

after all, a familiar fact that our immediate desires can resist the judgment of rational reflection and can seemingly, in some cases, even overpower our considered judgment. What entitles us to hold that this reflects a fracture within a single subjective standpoint, rather than a struggle between two essentially distinct standpoints for control of a certain body?

A transformative theorist has a straightforward answer to this question, because the transformative view implies that *our* animal capacities are essentially such as to submit their deliverances to the judgment of rational reflection. It is certainly possible for our will to prove weak in the face of a powerful desire, and we may if we like describe such cases picturesquely as ones in which our 'animal nature' overpowers our rationality. But, this sort of possibility presents no deep threat to the unity of the relevant subject, for although it shows that our desires can in particular cases move us to act contrary to our reflective judgment, it does not show that our *capacity* for desire is of such a nature as to constitute a self-sufficient rival to the standpoint constituted by our capacity for rational choice. For the additive theorist, however, the possibility of a discrepancy between what we desire and what we reflectively choose presents at least a *prima facie* problem. If the relation between our rational and our desiderative capacities is conceived in the additive way—as a relation in which reason monitors, assesses, and intervenes in the operation of 'forces already at work'—then what differentiates this situation from the operation of one subject on another, as a sailor acts on a ship to change its course?

To reply that these two systems belong to a single body is not satisfactory, any more than it was satisfactory for Descartes to reply to Arnauld's objection by asserting that mind and body belong to a single composite. The question is whether the additive approach gives us the resources to understand what makes this body the locus of a single subjectivity, rather than the scene of a relatively harmonious accord between two subjectivities. Nor can we appeal to the idea that our animal and rational capacities have a single bearer. The point in question is whether there is a *single* bearer of capacities present here. To avoid tortured syntax, I have been speaking as if there is something to which these two capacities both belong, 'the rational animal'. But the question is whether this putative entity is like the entity Descartes calls 'a man': not fundamentally one thing, but two things standing in a relation.

The question at issue here is not one the eye can judge, it concerns how rightly to *conceive* of a rational animal, whether it is properly regarded as a single subjectivity of which both desires and choices are predicated, or two subjectivities standing in a relation. The difficulty for the additive theorist is to explain how, if the person's desiderative capacities are intrinsically independent of his or her capacity for reflective choice, the operations of these two capacities can express one and the same point of view: not an 'I' confronted with a resistant 'it', but a single, evolving conception of what is to be done. To the extent that this question remains unanswered, the additive approach faces difficulties in explaining what it is for me to *be* a certain animal analogous to the difficulties Cartesianism faces in explaining what it is for me to *be* an embodied living thing.

3.3. Unity and Interaction

Are the unity problem and the interaction problem two independent difficulties? Why do additive theories give rise to both? I will end this section with a brief and speculative remark about this.

The interaction problem was a difficulty about how, if our perceptual and desiderative capacities are conceived as additive theorists recommend, their operations can have the right kind of influence on our judgments and choices, an influence that is not merely exculpatory but justifying. Our consideration of the unity problem suggests a moral about what the right kind of influence would be. It suggests that such influence would be, not a form of *interaction* at all, but a form of *intra-action*, as we might put it: not one capacity monitoring and intervening in the operation of another, distinct capacity, but a single capacity to determine what to do actualized first in a more immediate, then in a more reflective way. These characterizations of how we might think about the relation between the actualizations of our rational and our animal powers are only gestures toward a topic in need of further investigation, but to the extent that they capture something, they suggest that the reason why additive theories fail to give a satisfactory account of the interdependence of our rationality and our animality is precisely because they cannot represent them as belonging, in the right way, to a *unified* capacity to exercise reason in negotiating the demands of animal life. If this is right, then the interaction problem and the unity problem are not simply two distinct difficulties. In an important sense, the latter underlies the former.

4. Conclusion: The Transformative Alternative

I have been arguing for an abstract but nevertheless real constraint on the relationship between our capacity for rational reflection and our capacities for perception and desire: namely that the latter capacities must be themselves informed by our rationality, in a way that renders them distinct in species (although certainly the same in genus) as the perceptual and desiderative capacities of nonrational animals.

My aim has been, so far as possible, to argue for this constraint without making specific commitments about how we *should* think of a rational subject's perceptual and desiderative capacities—or indeed, of the power of rational reflection itself. One thing I hope to have shown is that there is a debate to be conducted at this abstract level, one that does not turn on how specifically to conceive of our perceiving and desiring and their role in grounding judgment and choice. The upshot of our discussion is that, *however* these matters are to be understood, our perceiving and desiring must be actualizations of powers of an essentially rational form. If this is right, then our capacity for rationality does not merely complement our animal capacities for perception and desire, it transforms them in a way that distinguishes these capacities essentially from the corresponding capacities of a nonrational animal.

In its classic, Aristotelian version, the transformative conception of the relation between our rationality and our animality was motivated by two fundamental

ideas: first, that the soul is the *substantial form* of a living thing, that in virtue of which the living thing exists at all; and secondly, that 'rational' names the *specific difference* of our kind of soul. That rationality is our specific difference implies, as we have seen (Section 1.3), that animality, which is our genus, must take a distinctive form in our case. And, general Aristotelian principles about the relation of form to matter dictate that what receives this form must be of the right sort to receive it. In a slogan, our generic matter (*our* sort of animality) must be such as to receive our specific form (rationality).³⁴ A way of putting the upshot of this paper is to say that, if I have succeeded, we should be able to see something attractive in this fundamental Aristotelian thought.³⁵

Matthew Boyle

Department of Philosophy, Emerson Hall,
Harvard University, USA
boyle2@fas.harvard.edu

NOTES

¹ Recent defenders of a broadly classical distinction between rational and nonrational animals include Davidson (1982), Korsgaard (1996, 2009), McDowell (1994), and O'Shaughnessy (2003). For discussion of how to draw this distinction, and of its empirical application, see the essays in Hurley and Nudds 2006. For skepticism about the depth of the distinction, see for instance Stich 1990, Kornblith 2012, and Doris 2015.

² I invoke the idea of a module here without meaning to signal any direct connection to debates about 'modularity' in the philosophy of mind. What I mean by saying that our capacity for rational reflection is conceived as a distinct module is simply that it is conceived as a capacity distinct from our animal capacities to perceive and desire, so that its presence in our minds does not alter the nature of these other capacities. I will shortly give a sharper characterization of the relevant distinctness.

³ Velleman introduces this proposal in a hypothetical mode, but he subsequently makes clear that it is not just a suggestion about how practical reason *might* be added to a faculty of motivation, but a proposal about how to understand the power of practical reason we actually possess.

⁴ Compare McDowell's remark that 'It is essential to conceptual capacities, in the demanding sense, that they can be exploited in active thinking, thinking that is open to reflection about its own rational credentials. When I say the content of experience is conceptual, that is what I mean by "conceptual"' (McDowell 1994: 47).

⁵ See for instance Stalnaker, 2003, Peacocke, 1998, 2001, Heck 2000, Kelly 2001, Byrne 2005.

⁶ See McDowell, 2009.

⁷ Cf. McDowell 1994: 69, 109.

⁸ Aristotle 1984, X. 8 (1057b39–1058a7): 1671; emphases mine.

⁹ Aquinas 1995, Bk. X, Ch. 10, §2119: 760. Elsewhere, speaking for himself rather than for Aristotle, Aquinas considers the following objection:

[A] human being and a horse are alike in being animal. Now an animal is called an animal because of its sensitive soul; therefore a human being and a horse are alike in their sensitive souls. But the sensitive soul of a horse is not rational. Consequently neither is that of a human being.

Aquinas replies:

Just as animal, precisely as animal, is neither rational nor nonrational, but 'rational animal' is a human being, whereas 'nonrational animal' is a brute; so also the sensitive soul precisely as sensitive is neither rational nor nonrational; rather the sensitive soul in a human being is rational, whereas in brutes it is nonrational. (Aquinas 1984, q. 11, ad 19)

The problems I raise for additive theories of rationality are modeled on difficulties Aquinas raises for views that hold that a rational animal has a sensitive 'soul' that is not intrinsically rational and a further 'soul' in virtue of which it is rational—a position Aquinas associates with Plato. For Aquinas's criticisms of this position, see for instance Aquinas 1948, I^a, q. 76, aa. 3–4, and Aquinas 1984, q. 11.

¹⁰ Standard objections include the following: that the view is in tension with the fact that we 'rational animals' have evolved from animals that are not rational; that it forces us to deny that nonrational animals genuinely perceive, desire, and know things and exhibit intelligent activity in pursuit of goals; that it must deny that human beings often believe, judge, desire, and intend irrationally; that it faces difficulties in explaining the perceiving and desiring of human infants, or in accounting for their cognitive development into mature rational subjects; and that it is in some vaguer way 'unscientific' or not sufficiently 'naturalistic' in its whole approach. I believe all of these objections rest either on misrepresentations of the transformative theorist's position or on disputable assumptions about the shape that all sound understanding of the natural world must take. Because my aim here is simply to raise difficulties for additive theories, I will not discuss these objections in the present paper. For discussion of some of them, see my 'Essentially Rational Animals' (2012).

¹¹ See McDowell 1994, Lecture I, §3, and for further discussion, see Lecture III and Afterword, Part II.

¹² Compare for instance Heck 2000: 512–514 and Peacocke 2001: 255–256.

¹³ This capacity-oriented way of thinking about the nature of representational contents is certainly not universal (for rejections of this way of framing issues about conceptual content; see for instance Stalnaker 2003 and Byrne 2005). It is, however, common ground among the authors with whom McDowell is primarily engaged: compare Evans's 'Generality Constraint' on conceptual representation (1982: 100–105) and Peacocke's 'Principle of Dependence' (1992: 5, and cf. pp. 42–51).

¹⁴ On this point, see esp. McDowell 1994, Afterword, Part II.

¹⁵ Elisabeth to Descartes, La Haye, 16 May 1643 (in Shapiro 2007: 62).

¹⁶ For recent discussion of Elisabeth's objection and Descartes's options in replying to it, see for instance Garber 2001 and Schmaltz 2008, Ch. 4.

¹⁷ An additive theorist might well admit a dependence in the converse direction: she might admit that explaining what the power reason is requires describing how it is related to the powers of perception and desire. I say more about this contrast between dualistic theories and additive theories in Section 3.2.

¹⁸ An author who does appear to accept a version of the view I discuss is Schapiro (2009), who advocates a position she calls 'inclination as animal action': that is, that our 'inclinations' are to be conceived as having essentially the same nature as the sorts of impulses that directly govern the activity of a nonrational animal.

¹⁹ Although I think this claim about the characteristic mode in which our desires present their objects is intuitively attractive, it is certainly not undeniable. I will not seek

to defend it further here, but just to explore the consequences that would follow from its acceptance. For defense of the claim itself, see Boyle and Lavin 2010.

²⁰ A similar conclusion about Quinn's case is drawn at Scanlon 1998: 38. For related discussion of the 'intelligibility' of ordinary desire, see Stampe 1987.

²¹ I am indebted to an anonymous reader for pressing me to address this point.

²² For a survey of standard attempts and objections to them, see Brewer 1999, Ch. 4.

²³ A different kind of objection to McDowell's argument, which I cannot discuss here, questions the very idea of perceptual *content*. Both McDowell and his original nonconceptualist opponents took it for granted that perception supplies us with reasons by presenting certain representational contents as characterizing the subject's environment. In the past decade, however, there have been important challenges to this notion of perceptual content (see, e.g., Travis 2004 and Brewer 2006). I think these challenges have significant merit, but to treat them adequately would require another paper; so for the sake of expository simplicity, I have adhered to McDowell's original formulation of the issue. Even if the notion of perceptual content is rejected, I believe it is still possible to raise the question at issue between additive and transformative theories of perceptual rationality. Briefly, even if the role of perception is simply to *present* us with worldly things, not to supply us with *representational contents* concerning those things, there is still room for a contrast between additive and transformative conceptions of such presentation. In work now in progress, I offer a fuller account of how a transformative theory of perceptual rationality can take on board the insights of critics of the notion of perceptual content.

²⁴ Compare with Peacocke 1998: 383, 386–7 and Peacocke 2001: 255–6; and Heck 2000: 512–9.

²⁵ This need not be a judgment about how things perceptually *appear* to her. It might be a (factive) judgment about how she perceives things to be. The role of the relevant judgment is not to ground a claim about how things are in the subject's environment on an epistemologically less problematic proposition about how things appear to her but to formulate a reflective understanding of how the relevant fact about her environment is known to her—namely in virtue of an actualization of her perceptual capacities. To know that an object with certain properties is (e.g.) visually available to me is not to know something less contentious from which I can infer that my environment contains an object with certain properties; it is to understand something about *how I am presented with* the relevant environmental fact, and thus to understand what kind of reason I have for holding it true.

²⁶ Nor was it part of McDowell's project: cf. McDowell 1994: 55.

²⁷ Fourth Set of Objections, in Descartes 1984: 143 (AT VII, 203). My discussion is indebted to the illuminating account of this objection and its scholastic background in Rozemond 1998, Ch. 5.

²⁸ For the attribution, see for instance Aquinas 1984: 148 (q. 11). Whether this attribution is justified depends partly on whether the 'First Alcibiades' is a genuine work of Plato: there the doctrine that I am a soul that makes use of a body is explicitly maintained by Socrates (see Plato 1997: *First Alcibiades* 129e–130c; but cf. also *Republic*, 580d–581c and *Timaeus*, 69e–70a). The use of the sailor-in-a-ship metaphor to characterize the relation of soul to body does not appear in the existing Platonic corpus, but it is mentioned by Aristotle (in the vicinity of a discussion of Plato's views, although not explicitly as a characterization of them) at *De Anima* II.1.

²⁹ See Descartes 1984: 56 (AT VII, 81).

³⁰ Replies to Fourth Set of Objections, Descartes 1984: 156 (AT VII, 222).

³¹ Note that, on the Aristotelian view described earlier (Section 1.3), the animality that rational animals possess *is* an animality to which rationality is essential. The existence of a wider genus to which this species of animality belongs, and the fact that this *genus* does not imply rationality, does not conflict with this. For on the Aristotelian view, the genus is an abstraction, which can be exemplified in actual cases only by one of its species: either essentially rational animality or essentially nonrational animality. When we say that a human being is an animal, we predicate not the abstract genus, but a certain species of animality. Cf. Aristotle 1984: VIII.8 (1057b35–1058a8) and Aquinas 1949: II, ¶10.

³² A parallel problem about the relation between our capacities for perception and judgment could be raised by considering a subject whose perception presents X as F but who reflects and judges that X is not F.

³³ The phrase is of course an oversimplification: the relations between a nonrational animal's perceptions and its beliefs, and between its desires and its pursuits, might be quite complex. But whatever these relations might be, they would not occur under the governance of a capacity to reflect on reasons. This, I take it, is the real point of Korsgaard's phrase: a nonrational animal's point of view contains no distinction between the factors that function as *prima facie* solicitations to act and the subject's reflective assessment of these solicitations.

³⁴ For the comparison of genus with matter and difference with form, see Aristotle 1984: VIII.2 (1043a14–26) and VIII.6 (1045a20–25). Compare with also Aquinas 1949: 35–36.

³⁵ During the very long gestation period of this paper, I have incurred debts to more people than I can now recall, but I am particularly grateful to Jim Conant, Adrian Haddock, Andrea Kern, Matthias Haase, Richard Moran, Susanna Siegel, Charles Travis, and Jennifer Whiting for advice, criticism, and encouragement.

REFERENCES

- Aquinas, T. (1948), *Summa Theologica*. Trans. Fathers of the English Dominican Province. New York: Benziger Brothers.
- (1949), *On Being and Essence* (De Ente et Essentia). Trans. A. Maurer. Toronto: Pontifical Institute of Mediaeval Studies.
- (1984), *Questions on the Soul* (Quaestiones de Anima). Trans. J. H. Robb. Milwaukee: Marquette University Press.
- (1995), *Commentary on Aristotle's Metaphysics*. Trans. R. J. Blackwell, R. J. Spath, and W. E. Thirlkel. Notre Dame, IN: Dumb Ox Books.
- Aristotle (1984), 'Metaphysics', in J. Barnes (ed.) *The Complete Works of Aristotle*, Vol. 2. Princeton: Princeton University Press.
- Boyle, M. (2012), 'Essentially Rational Animals', in G. Abel and J. Conant (eds.) *Rethinking Epistemology*, Vol. 2. Berlin: Walter de Gruyter.
- Boyle, M. and Lavin, D. (2010), 'Goodness and Desire', in S. Tenenbaum (ed.) *Desire, Practical Reason, and the Good*. Oxford: Oxford University Press.
- Brewer, B. (1999), *Perception and Reason*. Oxford: Oxford University Press.
- (2006), 'Perception and Content', *European Journal of Philosophy*, 14: 165–81.
- Byrne, A. (2005), 'Perception and Nonconceptual Content', in E. Sosa and M. Steup (eds) *Contemporary Debates in Epistemology*. Oxford: Blackwell.
- Davidson, D. (1982), 'Rational Animals', *Dialectica*, 36(4): 317–28.
- Descartes, R. (1984), in J. Cottingham, R. Stoothoff and D. Murdoch (ed.) *The Philosophical Writings of Descartes*, Vol. 2. Cambridge: Cambridge University Press.

- Doris, J. (2015), *Talking to Ourselves: Reflection, Ignorance, and Agency*. Oxford: Oxford University Press.
- Evans, G. (1982), *The Varieties of Reference*. Oxford: Oxford University Press.
- Garber, D. (2001), 'Understanding Interaction: What Descartes Should Have Told Elisabeth', in *Descartes Embodied*. Cambridge: Cambridge University Press.
- Gunther, Y. (ed.), (2003), *Essays on Nonconceptual Content*. Cambridge: MIT Press.
- Heck, R. (2000), 'Nonconceptual Content and the "Space of Reasons"', *Philosophical Review*, 109: 483–523.
- Herder, J. G. (2002), 'Treatise on the Origin of Languages', in M. N. Forster (ed.) *Philosophical Writings*. Cambridge: Cambridge University Press.
- Hurley, S. and Nudds, M. (eds) (2006), *Rational Animals?* Oxford: Oxford University Press.
- Kelly, S. D. (2001), 'Demonstrative Concepts and Experience', *Philosophical Review*, 110: 397–420. Reprinted in Gunther 2003.
- Kornblith, H. (2012), *On Reflection*. Oxford: Oxford University Press.
- Korsgaard, C. (1996), *The Sources of Normativity*. Cambridge: Cambridge University Press.
- (2009), *Self-Constitution: Agency, Identity, Integrity*. Oxford: Oxford University Press.
- McDowell, J. (1994), *Mind and World*. Cambridge: Harvard University Press.
- (2009), 'Avoiding the Myth of the Given', in *Having the World in View*. Cambridge: Harvard University Press.
- O'Shaughnessy, B. (2003), *Consciousness and the World*. Oxford: Oxford University Press.
- Peacocke, C. (1992), *A Study of Concepts*. Oxford: Oxford University Press.
- (1998), 'Nonconceptual Content Defended', *Philosophy and Phenomenological Research*, 58: 381–8.
- (2001), 'Does Perception Have a Nonconceptual Content?', *Journal of Philosophy*, 98(5): 239–64.
- Plato (1997), J. Cooper (ed.) *Complete Works*. Indianapolis: Hackett Publishing Company.
- Quinn, W. (1994), 'Putting Rationality in Its Place', in *Morality and Action*. Cambridge: Cambridge University Press.
- Rozemond, M. (1998), *Descartes's Dualism*. Cambridge, MA: Harvard University Press.
- Scanlon, T. M. (1998), *What We Owe to Each Other*. Cambridge: Harvard University Press.
- Schapiro, T. (2009), 'The Nature of Inclination', *Ethics*, 119: 229–56.
- Schmaltz, T. (2008), *Descartes on Causation*. Oxford: Oxford University Press.
- Shapiro, L. (ed.), (2007), *The Correspondence between Princess Elisabeth of Bohemia and René Descartes*. Chicago: University of Chicago Press.
- Stalnaker, R. (2003), 'What Might Nonconceptual Content Be?' in Gunther 2003.
- Stampe, D. (1987), 'The Authority of Desire', *Philosophical Review*, 96: 335–81.
- Stich, S. (1990), *The Fragmentation of Reason*. Cambridge, MA: MIT Press.
- Travis, C. (2004), 'The Silence of the Senses', *Mind*, 113: 57–94.
- Velleman, J. D. (2000), *The Possibility of Practical Reason*. Oxford: Oxford University Press.