

Chapter 2

Intrinsic Value, Pain, and Method

We can often gain new insights into old issues by changing our approach. This dissertation is part of a larger project unified by a particular philosophical methodology. I believe that we can make progress on many of value theory's central questions by asking them just about pain. Instead of asking about the nature of intrinsic value, I shall ask about the nature of pain's intrinsic value without assuming that this will translate into a more general account of intrinsic value. For example, I shall not assume that what goes for pain's intrinsic value goes for pleasure's. By narrowing our investigation in this way, we can make a great deal of progress in our understanding of the nature of intrinsic value.

To do this, I'll begin with some background by discussing the philosophical methodology most modern writers use and some of its consequences in §2.1. In §2.2 I'll set out two central debates surrounding intrinsic value. Then after sketching my particular approach in §2.3, I'll flesh it out and lay groundwork for the forthcoming chapters by setting out a virtually universal conception of what pain is in §2.4. I'll then do some conceptual cartography by taxonomizing the existing accounts of pain's intrinsic badness in §2.5 and §2.6.

§2.1

The coherence method

It is difficult to discuss philosophical methodology in the abstract. At a fine-enough grain, the number of approaches to philosophy approaches the number of philosophers, and approaches are not readily separable from the topics they approach. But I believe there are (at least) three distinct ways of approaching value theory. I'll adumbrate these three methods in §2.1.1. In §2.1.2 and §2.1.3 I'll give a bit more detail to the most popular approach.

2.1.1 Three methods

On the condescension method we begin from the very top. Writers like Kant, Brandt, and Hare, begin with purely formal concepts, or rarified claims about the nature of rationality, or thin axiological and metaethical claims about the structure of value and normative discourse. From those heights these writers then derive and justify answers to more substantive questions about value — for example, why is pain intrinsically bad? — and the problems of practical ethics.¹

Of course, as many have pointed out, if one begins with conceptions of rationality and such which are too rarified and insubstantial, there will not be enough to derive answers to substantive questions. At the same time, if one builds too many substantive assumptions about human nature and motivation into the starting point, the results will be tainted and will fall short of the universality at which such approaches aim.

¹ For example, Kant (1785), (1788); Hare (1952), (1972), (1971), (1981); and Brandt (1979).

On the swamp-draining method we proceed by casting a jaundiced eye on normative topics as they've been traditionally raised. On this view, normative theory and discourse is a fever-swamp which we must drain before we can understand and apply normative concepts. In the shadow of Hume, the this method's proponents discard traditional questions like 'what is the nature of intrinsic value?' or 'why is pain intrinsically bad?' in favor of the careful study of human normative talk and behavior. These writers thus draw heavily upon the resources of behavioral economics, psychology, evolutionary biology, ethnography, and other sciences. The new aim of moral philosophy becomes cataloging and finding common structures in the norms and customs of the world's peoples. We might find, for example, that there is a widely held aversion to certain kinds of sensations that most cultures regard as in of themselves undesirable. And with enough evolutionary background as to why an aversion to such sensations is biologically optimal, we've said what there is to be said about the intrinsic badness of pain.

The coherence method is by far the most common modern approach to value theory. On this method, we proceed in value theory by seeking coherence in our answers to a host of metaethical, normative, axiological, and practical questions on the one hand; and our judgments about example cases on the other. Griffin summarizes the approach:

The best procedure for ethics...[is one of] going back and forth between intuitions about fairly specific situations on the one side and the fairly general principles that we use to try to make sense of our moral practice on the other, adjusting both, until eventually we bring them all into coherence.²

Like both of the other methodologies, the coherence method is implemented in many different forms by different writers. My own approach will be a version of the coherence method.

I shall have no more to say about the condescension and swamp-draining methods. Both can be fruitful. Though I suspect that their fruits will be most significant as data for the coherence method. That is a question far beyond the purview of this dissertation. This chapter will set out my particular version of this method. In this section I'll discuss this method as it is generally applied before setting out my own version of it in §2.2.

2.1.2 The coherence method in outline

Let me begin with an extremely rough and idealized picture of how work in value theory proceeds on the coherence method. It will help to imagine that we are just starting out on the broadest questions of value theory. We begin with a large set of normative and metaethical theses, and a set of prima facie judgments about substantive cases. Each thesis and judgment has its own weight —we find them to various degrees more and less plausible, and more and less difficult to give up. I shall not address the details of this notion of weight or its

² Griffin (1996), 9. My thoughts about method have been very much influenced by Griffin's work.

relationship to the justification of our moral beliefs. For my purposes it will be enough to take 'weight' as something like 'degree of conviction.'³

Taken individually, some theses and judgments seem to have a particularly heavy weight. This is true of the thesis that equals are to be treated equally; and the judgment that something very wrong occurs when one wantonly kills. Others have lesser but still significant weight. We might be attracted to, yet somewhat ambivalent about, the doctrine of double effect; or about the judgment that we ought to save the ten trapped miners instead of diverting the resources to preventing future catastrophes. And so on, all the way down to theses which we believe are simply false and judgments in cases with no normative significance.

Taken together, the prima facie weightings conflict. Some theses which fit with intuitively plausible judgments in one case endorse seemingly implausible judgments in other cases. We then must decide how to rearrange the prima facie weights. For example, many find certain claims about the impartial promotion of utility plausible but balk when they imply that those in richer countries ought to transfer much of their wealth to those in poorer countries. That strikes many as making morality too demanding. We must then lower the weight of the alleged moral requirement or convince ourselves that the negative reaction is

³ For some important discussions of the relationship between coherence and justification in ethics, see Rawls (1971), Ch.1-3; Griffin (1996), Ch.1; Scanlon (1998), Ch.4-5.

misleading. In this case, the change in weights need not be all-or-nothing (though it might be). In practical terms, we may change our judgments about the amount we owe while retaining the belief that we do owe.

How we resolve these conflicts depends both on the apparent weights of the thesis and judgment, and on their relationships to other theses and other judgments. Some conflicts require that we revise just one thesis or judgment. With others we must revise large sets of interrelated theses and judgments. The ultimate goal is a normative theory —a maximally coherent set of weighted theses and substantive judgments.⁴

Like coherence theories of justification or truth, the coherence method in ethics faces several well-known objections. Famously, for whatever set of weightings of theses and judgments we come to, there will be other equally coherent sets of weightings and no principled way of choosing between them. I must leave this and other objections to others. In any event, the ultimate defense of a taxonomy or methodology is its fruits. And, like many others, I believe that some version of this approach will give us much to harvest.

Different writers apply the coherence method in importantly different ways. But the general approach is the same. Most writers also share a common

⁴ We need not think that ‘maximally coherent’ means ‘completely coherent’. The coherence method is compatible with the possibility that the best normative theory will contain intractable disputes between some theses and/or substantive judgments. Larry Temkin makes something like this point in Temkin (1993), Ch.10.

conception of the scope of their projects. They focus on general questions like ‘what is intrinsic value?’ and use particular phenomena like the intrinsic value of pain as test cases. That is, the intrinsic values of pain, pleasure, autonomy, and other putative exemplars, function as proving grounds for answers to these general questions.

2.1.3 Ship-building on the open seas

As the coherence method is normally applied, there can be no a priori guarantees that any of our normative beliefs will be included in the final maximally coherent normative theory. That in turn suggests that there are no unshakable foundations in value theory. This is because the concepts and therefore the questions we ask in value theory are deeply theory-laden.

In building a theory we are, in a sense, at the same time building the very concepts that are at issue. For example, in inquiring about what has intrinsic value, we are at the same time crystallizing what we mean by ‘intrinsic value’. Similarly, our concept of rights acquires a great deal of content from our broader theoretical concerns. If we are Kantians, we may see rights as (nearly) inviolable restrictions on conduct which have a central and, in a sense, underived moral status. Whereas, if we are rule utilitarians, rights may be relatively strict constraints on conduct that have their normative status in virtue of our inability to know what actions will maximize utility in every particular case we face. All

of these views share some common core which fixes the concept in a rough position in logical space. But while there is a core to the concept, the content of our concept of a right is in large part a nexus of interrelated normative theses and judgments about cases. Thus, in a sense, we are never talking just about rights.

If the concepts involved are theory-laden in this way, the questions we address will be similarly laden. We can only take on one neighborhood of issues at a time. Thus when we take up a topic, we must hold seemingly remotely related theses and judgments in abeyance. The fact that different writers will bring different concepts to the table will often mean that they are holding different issues in abeyance.

If this is correct, when we do value theory there are no theory-neutral concepts and the questions we ask are somewhat artificial subsets of interconnected issues. These two observations show that there can be no a priori guarantee that any of our normative beliefs —no matter how deeply held— will not be overturned in the final analysis. Thus there cannot be any theory-neutral foundation on which to stand that is not itself subject to crumbling as we proceed. There are no Archimedean points in value theory.

This isn't a new observation. Nor is it a problem for the coherence method as it is usually applied. There are better and worse places to stand. On better

ground the concepts and questions have wide theoretical stability and widespread intuitive backing and thus seem less likely to be overturned.

I believe that the claim that pain is intrinsically bad is as close as we can possibly come to an Archimedean point. If we narrow our scope to just pain and do not consider any other examples of intrinsic value, we can get the concepts involved thin enough that they will be neutral between the competing theories. To help flesh this out, I'll now turn to some issues about intrinsic value.⁵

§2.2

Conceptions of intrinsic value

Let me now turn to the nature of intrinsic value. In adumbrating two important and interrelated debates about intrinsic value, I'll be both introducing one of this dissertation's main topics and giving an example of the sorts of interconnections that I discussed above. I'll begin with the competing conceptions of intrinsic value.

⁵ In Ernie Sosa's metaphor, coherence methods attempt to assemble theories in the same way that one might struggle to build a raft while floating in the middle of the ocean. Extending the metaphor may help clarify my point about guarantees. Imagine yourself floating in the debris field of a shipwreck. The reasonable thing to do is to clamber onto the biggest and sturdiest bit of driftwood you can find. You would then paddle around looking for other suitable pieces to lash to the first piece, gradually building a boat as you go. Sometimes you will find new, better, pieces and discard older more ill-fitting components. Indeed, at some point you may find that you ought to abandon the large piece that you began with in favor of something else which better fits the whole. Even though it is a big and sturdy piece, there's no telling what else you'll find.

2.2.1 Six varieties of value

Different writers may mean very different things by 'intrinsic value'. But there is a common core. Our putative concept of intrinsic value is of a value that is most tightly tied to its bearer and which plays a particular and important role in our normative thought. Moore captures a large part of this role:

That which is meant by '[intrinsically] good' is...the *only* simple object of thought which is peculiar to Ethics. Its definition is therefore, the most essential point in the definition of Ethics and moreover a mistake with regard to it entails a far larger number of erroneous ethical judgments than any other. Unless...its true answer [is] clearly recognized, the rest of Ethics is as good as useless from the point of view of systematic knowledge.⁶

His acolyte Zimmerman adds

At the heart of ethics lie the concepts of good and bad; they are at work when we assess whether a person is virtuous or vicious, an act right or wrong, a decision defensible or indefensible, a goal desirable or undesirable....It is in virtue of intrinsic goodness and badness that other types of goodness and badness may be understood⁷

With 'intrinsic value' we are referring to one of several possible combinations of six kinds of value which fits this role. Many of these normally overlap. For perspicuity I shall separate them here. My goal here is to frame some of the debates and what's at stake. Thus I'll gloss over several details. I shall discuss them in §5.1.

The first variety is non-derivative value. An x has non-derivative value when we can want x strictly for its own sake. x is desired as an end; it does not

⁶ Moore (1903), 5.

⁷ Zimmerman (2001), 4.

inherit its value from anything else. The goodness of pleasure and happiness are standard examples. Its opposite is derivative value. Instrumental value is a paradigm example of derivative value. The instrumentally valuable is only valuable insofar as it conduces to the achievement of some non-derivative value. Money is valuable because it allows one to buy new records, the listening to which gives one pleasure. Contributive value —the value an x has in virtue of being part of a valuable whole— is another form of derivative value.⁸

Non-relational value is the value an x has solely in virtue of its non-relational properties. Its opposite is relational value which something has (at least partially) in virtue of its relational properties. While instrumental values are again paradigm relational values, there can be others. As Korsgaard points out

Certain kinds of things, such as luxurious instruments...are valued for their own sakes under the condition of their usefulness. Mink coats and handsome china and gorgeously enameled frying pans are all things that human being might choose partly for their own sakes under the condition of their instrumentality: that is given the role such things play in our lives.⁹

Thus the value of a mink coat depends on the role it plays in our lives as an object of aspiration. In a similar vein, Kagan suggests that the pen Lincoln used to sign the Emancipation Proclamation is valuable in virtue of its causal and

⁸ There is another conception of non-derivative value on which an x has non-derivative value if there is no helpful explanation of why x is good. I shall ignore this alternative in this chapter. I discuss it in §5.1.3.

⁹ Korsgaard (1983), 185.

historical properties (these are also allegedly cases of things with non-derivative value).¹⁰

Finally, essential value is the sort of value a thing has necessarily. That is, an *x* has its essential value no matter in what circumstances it occurs; *x*'s essential value thus depends solely on *x*'s essential properties. This is what Moore had in mind with his famous isolation test for intrinsic value.

In order to arrive at a correct [answer to the question 'what things have intrinsic value?'] it is necessary to consider what things are such that, if they existed *by themselves* in absolute isolation, we should yet judge their existence to be good¹¹

Many things have non-essential value. That is, value which depends on some contingent property. Ice cream is good only insofar as people like it. Though they tend to be closely related, non-relational value and essential value need not always correspond. A thing's non-relational and essential properties are often different. The box's being square is a non-relational property, but not an essential property — it remains the same box even after being squished.¹²

2.2.2 Two answers

When we ask what intrinsic value is, we want to know which mélange of these six values properly fills the conceptual role carved out by the putative

¹⁰ Kagan (1998).

¹¹ Moore (1903), 187.

¹² We can distinguish still other sorts of value. For example, C.I. Lewis introduced the notion of inherent value to cover the objects of an intrinsically good experience. See, Lewis (1946), Nonetheless, I think these six values are at the heart of the debates about intrinsic value.

conception. The existing accounts of intrinsic value divide into two broad camps.

In the prime example of the first sort of view, Moore writes that

When I say, with regard to any particular kind of value, that the question whether and in what degree anything possesses it *depends solely on the intrinsic nature of the thing in question*, I mean to say two different things at the same time. I mean to say (1) that it is *impossible* for what is strictly *one and the same* thing to possess that kind of value at one time, or in one set of circumstances, and *not* to possess it at another; and equally *impossible* for it to possess it in one degree at one time, or in one set of circumstances, and to possess it in a different degree at another, or in a different set. This, I think, is obviously part of what is naturally conveyed by saying that the question whether and in what degree a thing possesses the kind of value in question always depends *solely* on the intrinsic nature of the thing. For if *x* and *y* have different intrinsic natures, it follows that *x* cannot be quite strictly one and the same thing as *y*; and hence if *x* and *y* can have a different intrinsic value, only where their intrinsic natures are different, it follows that one and the same thing must always have the same intrinsic value....(2) The second part of what is meant is that if a given thing possesses any kind of intrinsic value in a certain degree, then...anything *exactly like* it...must...possess it in exactly the same degree.¹³

Thus the first competitor for the concept of intrinsic value is the

Moorean view: The intrinsic value of *x* is *x*'s non-relational, non-derivative, and essential value.

The alternative, is the

Final value view: The intrinsic value of *x* is *x*'s non-derivative value.

These are, I think, the two basic positions on what intrinsic value is.

2.2.3 The bearers of intrinsic value

On the coherence method, we cannot simply compare the theses involved in each view and the substantive judgments they yield, for these are all bound up

¹³ Moore (1951), 260-61. Italics original.

with other claims and judgments. They are, for example, deeply entangled with questions about the bearers of intrinsic value.

There are many different candidates for what exactly has intrinsic value. Some writers, hold that concrete objects can bear intrinsic value.¹⁴ On these views, people, books, works of art, and other concrete objects, are what is intrinsically good. Others locate intrinsic value in more abstract entities. On some views it is the state of affairs consisting in someone feeling pleasure or an artwork existing that is intrinsically good.¹⁵ Still others hold that facts¹⁶ or properties¹⁷ or tropes¹⁸ are intrinsic value's bearers. For simplicity, in discussing this bearers debate I will talk only about objects and facts as the candidate bearers. These two views can reasonably stand in for other potential bearers. They are also, I think, the two most plausible competitors.¹⁹

The metaphysical issues of the bearers debate are important in the dispute between the Moorean and final value views. Suppose the proponent of final value introduces a case wherein an *x* seems to have intrinsic value in virtue of its relational properties. As Kagan notes, if a view on which facts or states of affairs are the bearers of intrinsic value is correct,

¹⁴ For example, Kagan (1998); Korsgaard (1983); Rabinowicz and Ronnow-Rasmussen (2004).

¹⁵ For example, Chisholm (1986); Zimmerman (2001); and Lemos (1994).

¹⁶ For example, Ross (1930), 137.

¹⁷ For example, Butchvarov (1989).

¹⁸ Zimmerman seems to suggest this view in. Zimmerman (2001), Rabinowicz and Ronnow-Rasmussen attack it in. Rabinowicz and Ronnow-Rasmussen (2003),

¹⁹ Facts, on this understanding, are instantiated states of affairs.

there will of course be a fairly easy translation from the common, informal object-based idiom to the strictly correct fact-based idiom. Instead of saying that Lincoln's *pen* has intrinsic value, for example, by virtue of its having been used to sign the Emancipation Proclamation, we will say that what has value is the *fact* that there exists a pen which was used to sign the Emancipation Proclamation.²⁰

Thus the proponent of the Moorean view can accept the suggestion that the pen has intrinsic value.

We can recognize that [the fact that such a pen exists] has intrinsic value by virtue of its being about a pen being used in a particular way —but since we are assuming that this is an intrinsic property of the fact in question, it will still be true that only intrinsic properties of the fact are relevant to its intrinsic value.²¹

That is, by making facts the bearers of intrinsic value, the Moorean can suck up any relevant relational properties of the object and make them non-relational properties of the fact.

2.2.4 Why is pain intrinsically bad?

Our most ground-level axiological questions are laden with these and other disputes. The Moorean/final value and bearers disputes infect the question 'why are pains intrinsically bad?' Depending on which views we adopt, some independently plausible substantive accounts of pain's intrinsic badness may be non-starters.

²⁰ Kagan (1998), 293. Italics original.

²¹ Kagan (1998), 294.

Consider two simple substantive accounts. On the mental state theory, pains are bad because they are unpleasant. On the dislike theory, pains are bad because we dislike them. On the final value view, both are candidate accounts of pain's intrinsic value. Pain's unpleasantness could be non-derivatively bad. The same is true of being disliked. But on the Moorean view, the dislike theory cannot be an account of pain's intrinsic badness. Being disliked is a relational property. Thus if pain is bad in virtue of its relationship to the sufferer, that value cannot be intrinsic value.

The bearers debate is also enmeshed with our understanding of pain's intrinsic badness. The Moorean cannot say that pain is intrinsically bad because it's disliked. But she can say that the fact that there is a disliked pain is intrinsically bad. Thus by coupling the Moorean view with the claim that facts bear intrinsic value, one does not rule out dislike theories as accounts of intrinsic value.

Of course, on the coherence method, we are not deciding these issues first and then turning to pain's evil. Indeed, since pain is a key exemplar of intrinsic value, our substantive judgments about it carry heavy weight in the more general debates. For example, if we found the dislike theory overwhelmingly plausible for pain, that would be a substantial consideration in favor of the final value view or the Moorean view coupled with the claim that facts bear value.

But pain is just one of many exemplars of intrinsic value. Thus the support the dislike view gives to these views might be outweighed by theoretical or metaphysical concerns, or by what we find most plausible with other exemplars of intrinsic value like pleasure, well-being, and knowledge.

Therefore we cannot approach the intrinsic badness of pain without simultaneously addressing (at least) the Moorean/final value and bearers debates. What we say about pain's intrinsic badness depends on much more than what we think about pain.

§2.3

Privileging pain

Let me turn now the particular way I shall apply the coherence method. In this dissertation, instead of asking 'what is intrinsic value?' I shall ask only

Q1: Why is pain intrinsically bad?

Given that our beliefs about pain's intrinsic value have a particular firmness and centrality in our thought throughout value theory, I think we can make progress on many issues by focusing solely on how these debates play out with respect to pain. The remainder of this chapter will discuss this approach and lay the groundwork for the rest of this dissertation. I'll begin by clarifying the scope of this project and the strategy I shall use.

2.3.1 The univocality assumption

This proposal to attempt to make progress in our understanding of the nature of intrinsic value by focusing solely on pain's intrinsic value, may seem to give short shrift to our intuitions about other exemplars of intrinsic value. Why not, one might wonder, try to draw conclusions about intrinsic value by talking only about pleasure or knowledge or Korsgaard's cookware?

My answer may be surprising. I do not propose to draw conclusions about everything that has intrinsic value. I shall only be seeking coherence in our beliefs about intrinsic value as they fit with pain. I shall not assume that what goes for pain goes for anything else.

In pursuing the coherence method, we usually proceed by weighing competing views of the nature of intrinsic value by examining a wide range of cases. This depends on the assumption that a concept like intrinsic value is univocal. The univocality assumption entails that the weight we give our intuitions about the intrinsic value of one exemplar of intrinsic value can be compared to the weights we give intuitions about the intrinsic value of other exemplars. Suppose that α is a claim about the nature of intrinsic value. α entails plausible results when applied to the intrinsic value of pleasure, but implausible results when applied to the intrinsic value of close personal relationships. The univocality assumption entails that the positive results in one case and the

negative results in the other represent a genuine conflict and test for α . I propose to proceed by suspending the univocality assumption. I'll now flesh out the strategy this suspension allows me to pursue.

2.3.2 Strategy

Once we've restricted ourselves to the intrinsic value of pain, we can set aside the bearers debate until we've answered a more basic question:

Q2: What are pains insofar as they are normatively significant?

As I'll explain in §2.4.1 this is a narrower question than 'What are pains?' but for now the details won't matter. It would be a mistake to ask whether pain or the fact that someone is in pain is what is intrinsically bad without first being sure we know what pains are. And, as I'll show in chapter three, the view that virtually everyone accepts is wrong.

As I argued above, there is an intimate tie between the Moorean/final value debate, the bearers debate, and Q1. What position we take on any of the three can affect our views about the other two. If we can legitimately set aside the bearers debate until we understand what pains are, then we can also set aside the question of whether intrinsic value can depend on relational properties. If we don't know what pains are, we can hardly get started on asking whether their relational properties figure into their intrinsic value.

Once we have a picture of what pains are, we can address the Moorean/final value debate via a proxy. I shall ask:

Q3: In virtue of what are pains intrinsically bad?

If we are not considering any other exemplars of intrinsic value, we'll be limited to investigating how a pain's relational and non-relational properties figure into our judgments about its intrinsic value. That is, we will be examining substantive views of the value most tightly tied to pain which plays a particular and central role in our normative thought. Thus when we limit the debate to pains, we can make progress on the Moorean/final value debate by discussing the mental state and dislike theories (and their competitors).

In the following sections I'll lay the necessary background for Q2 and Q3. I'll begin by discussing a common conception of what pains are that nearly everyone holds. I'll then describe the most prominent theories of pain's intrinsic badness —that is, answers to Q3.

Once we come to a view of the best substantive account of pain's intrinsic value and what that leads us to say about other issues with respect to pain, we can then reinstate the univocality assumption and see what this shows us about intrinsic value in general. I shall not do so in this dissertation. My project will be limited to answering Q2 and Q3, and drawing conclusions about some issues in value theory as they stand to pains.

§2.4

The kernel view

In this section I'll lay the groundwork for my answer to Q2 in the next chapter. This requires distinguishing two conceptions of pain's nature. In this dissertation, I shall be only interested in the question:

Q2: What are pains insofar as they are normatively significant?

I shall not discuss the broader question:

Q4: What are pains?

In §2.4.1 I'll describe the difference between these two questions, and why it is significant for my project. I'll then turn to the virtually universal answer to Q2 in §2.4.2 and §2.4.3.

4.1 Two conceptions of pain

Q4 is the most general metaphysical question about pain. It demands a complete account of the nature of pain. Q4's answer will likely include an answer to Q2. Any answer to Q4 will likely be the product and synthesis of several disciplines: philosophy of mind, cognitive science, psychology, medicine, and neurophysiology, on the one hand; literature, art, and personal experience, on the other.

Q2 is much narrower. It is confined to the nature of pains insofar as they figure into our normative judgments and theories. When we inquire about the nature of pain from the normative standpoint of Q2, the firing of A δ nociceptors and activation levels in the reticular formation are hardly likely to be relevant. Thus when Korsgaard claims that

Pain is the perception of a reason²²

or Ryle writes that

we only call pains those [sensations] which we dislike. And if there are sensations which we ordinarily dislike but on some occasions like having, then we do not call them pains on those occasions on which we like having them.²³

they are making claims about the essential nature of pain qua normative phenomenon. If these claims were supposed to imply full accounts of pain — answers to Q4— we should dismiss them. If Korsgaard or Ryle were correct, pain scientists studying the sensory qualities of pain would be fundamentally mistaken about their subject. Studying pain tolerances to different patterns of noxious stimuli would be a questionable project if pains were only reasons or whatever we dislike.

Limiting discussion to Q2 may seem to beg the question on many important issues in the philosophy of mind. Some writers have vigorously

²² Korsgaard (1996), 149.

²³ Ryle (1949), 273.

argued that there are no such things as pains.²⁴ I shall assume that pains exist and have normative significance. I shall also distinguish pains from other mental states and each other via their normatively significant properties. That simply bypasses some of the most controversial questions about mental content and the metaphysics of mind.

But philosophers of mind should see my project as more helpful than threatening. I shall be fleshing out a commonsense conception of pain as it figures in our normative thought. Indeed, in our day-to-day lives our foremost concern with pain is as something bad. Thus the answer to Q2 is the conception of pain that should be at issue when philosophers discuss the entities referred to in our folk-theories of the mind. Once we have a well-described account of what we believe pains qua normative phenomena to be, we can then ask the philosopher of mind whether there are any such things.

Thus I shall henceforth use 'pain' only to refer to a normatively significant phenomenon. Let me now turn to what nearly everyone believes this phenomenon is.

2.4.2 The kernel view

On the kernel view of the nature of pain, a pain is just a painful sensation. Insofar as it is normatively significant, its nature and intrinsic badness lie solely

²⁴ Most notably, Hardcastle (1997) and Hardcastle (1999).

in the way it hurts. When I stub my toe, the pain is wholly before my mind in the way it stings and throbs. Insofar as it affects our lives, there is nothing else to pain.

The kernel view holds that pains are the atoms of experience which hurt. As an experiential atom, a pain is necessarily distinct from the other elements of one's experiential milieu. The arthritic pain in my hand as I type this sentence is distinct from my experience of the cat draped drooling across my forearms, though I am simultaneously conscious of both. Thus my reaction to a painful sensation is not part of the pain; it is a reaction to the kernel.

On this view, the character of the painful sensation exhausts the properties in virtue of which a pain is intrinsically bad. Stubbed toes throb; cuts sting and burn; migraines pound and crush. Hence if pains are bad because they are unpleasant, these properties constitute a pain's unpleasantness. If pains are bad because we dislike them, the kernel composed of these properties is what we dislike.

Terms like 'sensation', 'feeling', and 'hurting' are vague. There is a good deal of room for disagreement about what exactly they refer to. Nonetheless, we have a sufficient grasp on the way a pain hurts to distinguish it from other elements of our experiential milieu. A person in pain may also be, for example,

anxious, depressed, and have negative beliefs about her predicament. But these are not part of the pain on the kernel view. Her pain kernel could have occurred in another situation where she did not have these beliefs.

For example, Hare imagines jumping repeatedly into cold water to generate an analogy to how one could feel a pain without disliking it.

Suppose...that I do this diving act many times in the hope of getting not to mind this degree of cold; and that in the end I succeed. *It is not necessary to suppose that there is any change in the degree of cold that I feel (even subjectively);* there might be, but that would spoil the example. It may be merely that through habituation I stop minding my skin feeling like that. We do not even need to suppose any course of habituation. Whether I found the cold unpleasant or invigorating might depend on my general state of mind —on whether I was feeling depressed or elated. [80 ital added]

The sensation itself (the cold kernel) could persist between the two cases while other factors change around it and its value thereby changes. The depression or elation which matter in Hare's example do not seem to be part of the pain.

Hopefully, this makes the kernel view as I've described it plausible. But it is a further question whether the kernel view is as widely held as I've claimed. The answer will come in the next chapter. I shall argue that nearly every existing account of pain's intrinsic badness relies on the truth of the kernel view. I'll now briefly sketch the existing accounts of pain's intrinsic badness.

§2.5

Containment views

Let me now turn from what pains are to why they are intrinsically bad.

Substantive accounts of pain's intrinsic badness are answers to the question:

Q3: In virtue of what is pain intrinsically bad?

I believe the existing answers to Q3 divide into two camps.

On containment views, the presence of a pain kernel by itself is both necessary and sufficient for the pain being intrinsically bad; the source of the badness is contained entirely within the pain. One simple account holds that a pain is bad insofar as it is unpleasant. The more unpleasant a pain, the more intrinsically bad. On another, pain's badness lies in the power with which it motivates the sufferer to escape it.

Stance views hold that the presence of a pain kernel by itself is a necessary but not sufficient condition of its being intrinsically bad. The sufficient condition lies in some relationship between the sufferer and her pain. On some stance views, pain is intrinsically bad only insofar as its sufferer wants not to have it. Suppose you and I both stub our toes and experience qualitatively identical sensations (pain kernels). If I don't really mind my pain but you intensely dislike yours. Your pain is intrinsically worse than mine, even though they involve qualitatively identical pain kernels.²⁵

²⁵ There is a third kind of account. On privation views the intrinsic badness of pain consists in its being the absence of the intrinsically good. These views have, for good reason, fallen into ill-repute. I shall not discuss them in this chapter. I will discuss them in chapter four when I set out my own privation view in answer to Q3.

The crucial difference between the two is that, on containment views, the sufferer contributes little to the intrinsic value of her pain; she is, in a sense, merely the location where something bad occurs. Stance views hold that the intrinsic badness of a pain depends on some contribution made by the sufferer to the intrinsic value of her pain. It's not enough that the pain kernel simply occur in her, she must be involved with it. She must, for example, take some attitude toward the pain kernel.

I shall discuss the containment views and their three genera in this section. I'll discuss the stance views in §2.6.

2.5.1 Mental state theories

Mental state theories are the putative exemplars of containment views. On these views, pain is bad in virtue of its being unpleasant (or some other intrinsic phenomenological property). As Mill writes,

pleasure, and freedom from pain, are the only things desirable as ends; and that all desirable things...are desirable either for the pleasure inherent in themselves as a means to the promotion of pleasure and the prevention of pain.²⁶

Rachels emphasizes that on mental state theories:

It is an intrinsic, nonrelational fact about certain experiences that they are unpleasant [and therefore intrinsically bad]. Unpleasantness, on this view, supervenes on qualia: there cannot be a change in unpleasantness intensity without a change in qualia. Also, unpleasantness does not reduce to

²⁶ Mill (1863), 7.

motivation or disliking bodily damage relating in the right way to experience.²⁷

At least in this general form, these theories should be familiar.

On a mental state theory, pains are always unpleasant. But it is also true that we usually dislike pains when they occur. Thus it is worth emphasizing that on a mental state theory pains are disliked because they are bad, not vice-versa.

To clarify this relationship, consider a slightly cryptic passage from Mill:

desiring a thing and finding it pleasant, aversion to it and thinking of it as painful, are phenomena entirely inseparable or, rather, two parts of the same phenomenon —in strictness of language, two different modes of naming the same psychological fact; that to think of an object as desirable (unless for the sake of its consequences) and to think of it as pleasant are one and the same thing; that that to desire anything except in proportion as the idea of it is pleasant is a physical and metaphysical impossibility.²⁸

In this passage he is ostensibly claiming that being undesired and unpleasant are the same thing. That may seem to deny that the badness of unpleasantness is prior to the pain being disliked. But the appearances are misleading. Mill is being loose with 'desiring x' and 'x being desirable' —shifting back and forth between them as though they are equivalent. Mill believes that unpleasantness is an intrinsic property of a pain kernel. If unpleasantness and being undesired are equivalent, he must be claiming that being undesired is an intrinsic property of the pain kernel. But desires are attitudes that we take toward pain kernels.

²⁷ Rachels (2000), 195. The added phrase clarifies that on his view unpleasantness implies intrinsic badness.

²⁸ Mill (1863), 38.

Therefore, he cannot mean that 'x is undesired' and 'x is unpleasant' are equivalent. Rather, 'undesired' here means 'undesirable'. As with all mental state theories, on Mill's view pains are undesired because they are undesirable. They are undesirable because they are unpleasant (things can be undesirable without being unpleasant).

2.5.2 Motivation theories

Motivation theories are an often overlooked form of containment view. On these theories a pain is intrinsically bad in virtue of its intrinsic power to motivate its sufferer to escape the pain. Motivation theories thus tie degree of intrinsic badness to the strength of the motivation. Mild pains nudge; intense pains shove.

On Nagel's view:

Physical pleasure and pain do not usually depend on activities or desires which themselves raise questions of justification and value. They are just sensory experiences in relation to which we feel involuntary desire or aversion. Almost everyone takes the avoidance of his own pain...as subjective reasons for action in a fairly simple way; they are not backed up by any further reason.²⁹

Thus we dislike and desire to escape pain kernels in virtue of their intrinsically and necessarily spurring these feelings. Nagel explicitly rejects the claim that this motivational power could lie outside of the pain. On such a view,

²⁹ Nagel (1986), 156.

aversion to pain is a useful phobia —having nothing to do with the intrinsic undesirability of pain itself—which helps us avoid or escape the injuries that are signaled by pain....There would then be nothing wrong with pain in itself, and someone who was never motivated deliberately to do anything just because he knew it would reduce or avoid pain would have nothing the matter with him.³⁰

Since he believes that this view is insane, Nagel holds a motivation theory of pain's intrinsic badness.

But others do hold the view Nagel rejects. I'll argue in §2.6.2 that Korsgaard holds a stance view version of the motivation theory. On the stance view competitor, the motivational power lies in a person's disposition to be motivated in reaction to the occurrence of a pain kernel, not in the pain kernel itself.

Interestingly, there seems to be another possible containment version of the motivation theory which skirts very close to the stance view version by holding that a pain's disposition to move one to act may not necessarily be manifested whenever the pain occurs. But this is a containment view because the potentially unmanifested disposition is located within the pain instead of in the sufferer. Sidgwick quotes Mr. Bain as holding that

"pleasure and pain, *in the actual or real experience*, are to be held as identical with motive power." By this Mr. Bain does not, of course, mean that all pleasures when actually felt actually stimulate exertion of some kind; since this is obviously not true of the pleasures of repose, a warm bath, &c. The stimulus must in such cases be understood to be latent and potential; only

³⁰ Nagel (1986), 157.

becoming actual when action is required to prevent the cessation or diminution of that pleasure.³¹

If I read this correctly, the stimulus is allegedly latent in the pleasure itself. That would make this a containment view.

2.5.3 Representation theories

The third genus of containment views are representation theories which hold that pain is bad in virtue of its representing bodily damage. Pitcher writes that

To be aware of a pain is to perceive—in particular to *feel*, by means of the stimulation of one's pain receptors and nerves—a part of one's body that is in a damaged, bruised, irritated, or pathological state, or that is in a state that is dangerously close to being in one or more of these kinds of states.³²

Similarly, Tye holds that

Pains are sensory representations of bodily damage or disorder.³³

Representation theories of pain's intrinsic badness hold the representation Pitcher and Tye describe is the source of pain's intrinsic badness. In Hall's nice image:

[Pains are bad because] they accompany nociceptual reports of bodily damage, and bodily damage is something we don't like to hear about. It is like the ruler who slew the messenger who brought the bad news³⁴

If we take Hall's claim that damage is 'something we don't like to hear about' seriously, his representation theory may be a stance view; indeed it may just be a

³¹ Sidgwick (1884), 122. Italics original.

³² Pitcher (1970), 371.

³³ Tye (1995), 113.

³⁴ Hall (1989), 647.

dislike theory. The fact that a pain represents damage would explain why we tend to dislike pain kernels, but the pain would be bad in virtue of the stance we take toward it.

Notwithstanding, Hall's claim, taken as he intends it, captures what a genuine representation theorist holds. Any genuine representation theory must be a containment view. These theories are untenable if 'representing bodily damage' isn't an intrinsic property of the sensation. On a stance view version of the representation theory, the badness of pain depends on the presence of the pain kernel and the belief that the pain represents damage. This view is false. A person with causalgia may experience a constant and intense burning sensation in her hand. After a few years, it is unlikely that she will believe the sensation to be representing bodily damage. Yet her pain is still intrinsically bad. If that's right, representation theories of pain's intrinsic badness must be containment views.³⁵

§2.6 Stance views

On containment views, the sufferer contributes little to the intrinsic value of her pain; she is, in a sense, merely the location where something bad occurs. Stance views hold that the intrinsic badness of a pain depends on some

³⁵ Rachels makes this point with an unfortunately infelicitous example at Rachels (2000), 188.

contribution made by the sufferer to the intrinsic value of her pain. The presence of a pain kernel is necessary but not sufficient for the pain being intrinsically bad. Taking the relevant stance to the pain kernel is both necessary and sufficient for the pain being intrinsically bad. On all these views, if a person has a pain kernel but lacks the appropriate attitude, then the pain will not be intrinsically bad. That distinguishes them from containment views.

2.6.1 Desire and dislike theories

Desire and dislike theories of pain's intrinsic badness are the paradigm stance views. These views may be the most widely held accounts of pain's intrinsic badness. On a desire theory a pain is intrinsically bad in virtue of its sufferer desiring that it stop (or not occur). Desires are propositional attitudes. Thus on these views the pain kernel is indirectly the object of the desire via its figuring into the proposition. An alternative is a dislike theory on which pain is intrinsically bad in virtue of its being disliked. On these views the pain kernel itself is the object of the attitude. For the argument herein the desire and dislike theories stand and fall together and their differences will not matter. For simplicity I shall only discuss dislike views.

Within this family the many views differ along other dimensions as well. One difference concerns whether we necessarily dislike pain kernels. Views which hold this remain stance views because the pain is bad in virtue of the

stance we take toward it. Sidgwick holds a dislike theory and seems to endorse this view when he writes that

Let then [pain] be defined as feeling which the sentient individual at the time of feeling it implicitly or explicitly apprehends to be [undesirable]; — [undesirable], that is, when considered merely as feeling, and not in respect of its objective conditions or consequences, or of any facts that come directly within the cognisance and judgment of others besides the sentient individual.³⁶

Similarly, Baier, as described by Hare,

thinks that the fact that we dislike pains is not a contingent fact; 'whatever sorts of sensations we like and dislike, we only call pains those which we dislike. And if there are sensations which we ordinarily dislike but on some occasions like having, then we do not call them pains on those occasions on which we like having them.'³⁷

Thus on these stance views, it is impossible to like a pain kernel. Pains are therefore intrinsically bad because they are necessarily disliked.

On another sort of dislike theory, it is perfectly possible to fail to dislike the pain kernel. Ryle seems to hold this with his claim that

Pain is a sensation of a special sort, which we ordinarily dislike having³⁸ where 'ordinarily' implies contingency. More strikingly, in *Pain and Evil* Hare holds that

If I were flogged with a cat of nine tails, I should certainly dislike it, constituted as I am. But all this is not enough to establish...[a] logical, and not merely a psychological connexion, however inescapable, between experiencing the sensation ϕ and disliking it....If I have the sensation called

³⁶ Sidgwick (1884), 131. I have replaced 'pleasure' with 'pain' and 'desirable' with 'undesirable'.

³⁷ Hare (1972), 77.

³⁸ Ryle (1949), 109.

' ϕ ', all that I can be compelled logically to admit is that I have the sensation called ' ϕ '. Logic cannot make me suffer.³⁹

While this distinction raises a host of interesting issues, we can set it aside and treat all dislike views together.

2.6.2 Stance view motivation theories

Finally, as I mentioned above, there are stance view versions of the motivation theory. Korsgaard initially seems to hold a containment view when she writes that

The painfulness of pain consists in the fact that these are sensations which we are inclined to fight.⁴⁰

But the motivation and thus the intrinsic value of the pain do not lie in the pain kernel. Instead it is in the sufferer's disposition to be moved when a pain occurs.

She writes that

Pain is not the condition that is a reason to change your condition, the condition in which the natural and the normative are one. It is your *perception* that you have a reason to change your condition.⁴¹

and, that

pain is nearly always bad —because the creatures who suffer from it object to it. But it is important to see that this does not show that pain is an intrinsically bad sensation. For one thing, we don't always object to pain⁴²

Given that the intrinsic value of pain depends on an essential contribution by the sufferer, Korsgaard's motivation theory is a stance view.

³⁹ Hare (1972), 90.

⁴⁰ Korsgaard (1996), 147.

⁴¹ Korsgaard (1996), 148. *Italics original.*

⁴² Korsgaard (1996), 154.

§2.7

Conclusion

In chapter five, we will see that focusing narrowly on Q1 yields some important conclusions about the nature of pain's intrinsic value. In the next two chapters I shall argue for some very surprising answers to Q2 and Q3. These answers will be surprising and fresh in part because they tend to be off the radar screen when we follow the standard coherence method.