

Telic Perfectionism and the Badness of Pain

Antti Kauppinen (antti.kauppinen@helsinki.fi)

For Mauro Rossi and Christine Tappolet (eds.), *Perspectives on Ill-Being*. OUP.

Really final revisions, August 1, 2023

What if what explains why certain things are in themselves good for us is that they constitute flourishing as the kind of beings we fundamentally are? That, in broad terms, is the basic idea of perfectionism about well-being. As commonly understood, it says that we do well when we develop and exercise the capacities that make us *human* by excelling in activities involving them, such as politics and contemplation. We do badly when we fail to engage in such activities or perform poorly. But, critics say, what about good and bad *experiences* as such – aren't pain and suffering in themselves or robustly bad for us, beyond their possible instrumental impact on our agency? Well, yes. My aim in this chapter is to show that if we reformulate the basic idea of perfectionism in an independently motivated way, it can explain the important role of experiential welfare goods and bads as well. The key move is to understand flourishing in terms of the formal aims that are implicit in the functioning of the capacities that define who we are, including our capacities for positive and negative experience. In terms of such *telic perfectionism*, experiential ill-being turns out to amount to a kind of failure by standards implicit in valenced experience itself, just as poor and unsuccessful exercise of agency is a failure by standards implicit in the very use of reason.

An obvious challenge to this kind of view is that it offers the wrong kind of explanation of the badness of pain. As Gwen Bradford puts it,

Pain is bad, goes the natural thought, because it hurts. But perfectionism must let go of any aspiration to capture this thought. (Bradford 2021, 599)

There's a sense in which this is obviously true: no theory would qualify as perfectionism if it held that the fundamental explanation of why pain is bad is *simply* that it hurts. Again, perfectionists think that something gets to be on the list of basic goods or bads by way of amounting to flourishing or un-flourishing, or success or failure by the internal standards of a fundamental capacity. But what I want to argue for here is that pain *does* amount to failure by the internal standards of our capacity for valenced experience precisely *because of how it feels*. Very roughly, the phenomenal character of negative experience as such directly or indirectly offers for our agency the end of getting rid of such experience. This formal aim will necessarily be frustrated as long as the pain continues. Suffering is a special kind of failure by standards inherent to conscious experience. So for a telic perfectionist, it is in a sense bad because it hurts – the explanation of its badness makes ineliminable reference to its phenomenal character. Consequently, it doesn't require appeal to aims that are external to the experience itself, unlike Bradford's (2021) recent account, and thus avoids the objection from the badness of pain for creatures like infants who plausibly lack such external aims of practical reason (Fletcher 2022).

I'll make the case as follows. In the first section, I'll introduce standard developmental perfectionism and explain in more detail why the badness of pain is so challenging for it. Second, I'll sketch a different way of thinking about flourishing that focuses on the formal aims of the capacities whose operation defines who we are, for short our self-defining capacities, rather than human nature. It is relatively uncontroversial that how we exercise our capacity for *rationality* – what we treat as a reason for an action or attitude – constitutes an important part of our perspective on the world. But our whole self includes more, in particular how we *feel* about things, including what pleases or pains us. I'll label this aspect of us our capacity for *valenced experience*. Here and in the next section, I argue that the operation of both reason and valenced experience contains an implicit formal

aim or telos that is distinct from the contingent aims of the agent. For telic perfectionism, then, someone's life goes well for her to the extent such formal aims are realized, and badly to the extent they are frustrated.

In the third section, I turn to the key issue of how and why valenced experience has a telos that is necessarily frustrated by unpleasant pain and suffering in general. Drawing ecumenically on various recent accounts in the philosophy of mind, I contend that the phenomenal character of negatively valenced experiences more or less directly sets up for our agency the end of reaching a state in which such experiences are eliminated. As long as we remain in pain or otherwise suffer, then, we are necessarily failing to realize the telos of one of our self-defining capacities in virtue of having an experience with a certain phenomenal character. That's why pain, according to telic perfectionism, is robustly bad for us – and its badness is in an important sense due to its hurting, not to contingent person-level aims, because the standard for success and failure is inseparable from the phenomenal feel of the experience. So telic perfectionism is well-poised to give a unified account of both experiential and agential ill-being.

1. Developmental Perfectionism and the Problem of Pain

The label 'perfectionism' is used for a variety of loosely related views in ethics and political philosophy. My focus here is on perfectionism as a theory of well-being (and ill-being), rather than of intrinsic value (unlike Hurka 1993). The common way to characterize what perfectionism is in terms of developing and exercising characteristically human capacities, as in the following textbook definitions:

Perfectionism, broadly speaking, is the view that the development of certain characteristically human capacities is good. (Bradford 2015, 124)

The good life for a human is determined by human nature. Human nature involves a specific set of capacities. The exercise and development of these capacities is good for humans. (Fletcher 2016)

Since this way of understanding perfectionism focuses on developing and exercising capacities, which happens by way of engaging in characteristic activities, let's label it *developmental perfectionism*. There is much to be said for it. It is an inspiring idea that we live our best life when we make the most of what we're capable of. We live life to the fullest, as in the old US Army recruitment slogan "Be all you can be!". We lead an active life, a life fit for a being who is capable of more than lying on the couch and being stimulated for pleasure (cf. NE 1095b). Commonly, the kind of capacities that define human nature include at least theoretical and practical rationality, but some, perhaps most prominently Richard Kraut (2007), include emotional, social, and sensory capacities as well (cf. MacIntyre 2016, 201). Developmental perfectionism captures much of the appeal of an objective list theory while providing an actual theory that unifies the various things that are objectively good for us (though see Sobel 2010 for doubts). While for subjectivists desire-realization or value fulfilment is in itself good for us, perfectionism as a form of objectivism allows for the possibility that some of our desires are misdirected and some of our values mistaken from the perspective of our own good. Instead, it tells us to pursue "valuable functionings" (Sen 1987) like intellectual achievement, political engagement, and friendship, in the context of which we can and indeed must develop and exercise our distinctive capacities in an excellent way. It thus explains why circumstances in which people are deprived of such opportunities are bad for them, even if their desires and values have adapted to them, as Amartya Sen (1987) and Martha Nussbaum (2000) famously argue.

However, there are also significant challenges to developmental perfectionism. I'll set aside for now fundamental questions about the very existence of human nature (e.g. Kitcher 1999) and whether realizing such nature would be in our self-interest (e.g. Sumner 1996), and focus on the extensional adequacy of the thesis that what is basically good for us is developing and exercising perfectible capacities that belong to human nature, such as theoretical and practical rationality. If we set aside subjectivist intuitions (as found in Dorsey 2010), the most serious challenge is the one I started out with: how can perfectionism account for the most glaringly obvious welfare goods and bads, pleasures and (unpleasant) pains, when they don't, on the face of it, amount to exercises of a perfectible capacity that we can develop? (See e.g. Fletcher 2022; Haybron ms) Insofar as we can talk about a capacity to *feel*, or *feel appropriately*, both pleasure and pain seem to constitute the exercise of such capacity to the same extent, though one of them is as a rule good and the other bad for us (Bradford 2017, 348). It is, after all, appropriate to feel pain when injured, say, and more broadly unpleasant emotions when abandoned or denigrated. Developmental perfectionism thus seems to entail that it is in itself *good* for us to have negative experiences in such cases.

So why is unpleasant pain in itself bad for us, as it seems to be? Evidently, perfectionists can explain why pain is *instrumentally* bad, insofar as it prevents us from using our reason, for example. But intense pleasure can hinder rational activity just as much as pain. Yet it seems I'm doing far worse if I'm prevented from contemplating the heavens by a piercing headache than if I'm prevented from doing so by a blissful massage. Pain isn't just the *privation* of a good, but *robustly* bad (see especially Bradford 2021 for arguments).

How can developmental perfectionists respond to such a challenge? It seems insuperable as long as we think of only developing and exercising our capacities as basically good. This motivates reformulating the perfectionist value theory. Gwen Bradford (2021) has recently taken the first steps in this direction. She defends what she calls *tripartite*

perfectionism, which gets its name from a tripartite distinction between capacities, activities, and outputs. Here, *capacities* are the familiar perfectionist capacities like theoretical and practical rationality. Their exercise consists in *activities* like reasoning and deliberating. As in traditional perfectionism, such activities are in themselves good for us in Bradford's scheme. These activities result in *outputs*, such as knowledge or unjustified belief. What is crucial to Bradford's reformulation of the view is that according to it, these outputs are themselves also basically good or bad for us. As she emphasizes, our characteristic capacities are capacities *for* something (2021, 595). Plausibly, theoretical rationality, for example, is a capacity for gaining knowledge and understanding. It is thus successfully exercised only when it results in knowledge or understanding. On the tripartite scheme, knowledge and understanding are thus basic goods, too. Practical rationality, on Bradford's scheme, is a capacity for achievement of goals, so that achievement is another basic good.

What about robust bads, then? On tripartite perfectionism, they consist in outputs that don't fulfil the function of the capacity, such as false beliefs and failure (Bradford 2021, 597). It is in itself bad for us to exercise our capacities in a way that results in their 'malfilment' rather than fulfilment. Since pain isn't an exercise of a capacity we can develop, however, we have to tell a different story of why it's robustly bad. Here's Bradford's suggestion:

[P]ain's badness is a *malfilment of practical rationality* given a near-universal standing end to feel good rather than feel bad. (Bradford 2021, 599, my emphasis)

Practical rationality is the capacity to pursue ends, and the badness of pain is just a special instance of the badness of failure, given that we almost universally have as an end to avoid it.

Bradford is fully aware that her view amounts to a rejection of the commonsense explanation of why pain is bad for us. But I think it has deeper problems as well. In particular, excruciating pain would be bad for a creature who lacked the capacity for practical

rationality – it would be no act of mercy to respond to someone’s pain by removing their ability to reason practically while leaving the quality of their experience intact.¹ Unpleasant pain can also be bad for someone even if they lack the aim to get rid of it, if the aim is absent for bad reasons, like a misguided belief about the moral or aesthetic value of pain (cf. Arneson 1999). This is because avoiding unpleasant pain is not just any end, but an end we typically have *a self-interested reason* to have, which presupposes that it’s bad for us independently of having the end. So there’s excellent reason to look for an explanation of the badness of pain that doesn’t appeal to frustrating independent, contingent ends we may or may not have.

2. Reformulating Perfectionism

I’ve argued that developmental perfectionism, even at its best, fails to yield a satisfactory account of the badness of pain (and the goodness of pleasure). It also has at the very least the burden of proof to show that there is such a thing as human nature, and that doing well *as a human being* is basically good for us ourselves. These challenges motivate reformulating perfectionism while preserving its appealing core idea of well-being as flourishing.

The first step to such reformulation is distinguishing between different elements of perfectionism. Here’s what I regard as its key defining thesis:

Explanatory Perfectionism

What is in itself good for a subject S are the things that constitute flourishing as the kind of being S fundamentally is. What is in itself bad for S are the things that constitute unflourishing as the kind of being S fundamentally is.

¹ In correspondence, Bradford has indicated that she would get around this issue by adopting variabilism, that is to say, the view that a different theory of well-being is true for different kinds of subjects.

Developmental perfectionism is just one take on this general view, one that adopts what I'll call the *activity interpretation* of flourishing as development and exercise of one's fundamental capacities sufficiently well by their internal standards, and the *human nature thesis*, according to which we're fundamentally human beings. The problem with pain and suffering derives from the former, while other, more theoretical problems are linked to the latter. I thus want to sketch an alternative to both, starting from the activity interpretation.

As Bradford (2021) already observed, it is a somewhat surprising feature of standard developmental perfectionism that what she calls the "outputs" of exercising our fundamental capacities don't count for well-being. What this means is that if two people reason just as excellently in pursuit of the valuable aim of manufacturing a room-temperature superconductor, the project contributes the same amount to their well-being, even if only one of them is successful, while the other fails due to impurities in the available source materials. This doesn't seem right. However, while I thus agree with Bradford that exercising capacities well doesn't suffice for well-being, talking of things like knowledge and achievement as the "outputs" of capacities of theoretical and practical reason is misleading, since the output of reason is more naturally understood as a belief or intention. It is better to think of knowledge (or better yet, a kind of understanding) and valuable practical achievement as the *formal aims* implicit in the exercise of our rational capacities. Such aims set the *standard of success and failure* for beliefs and intentions (O'Brien 2019). After all, what makes capacities rational is that they are more or less sensitive to reasons for belief and reasons for actions and attitudes – when using reason, we are at least implicitly trying to get it right, aiming to base our actions and thoughts on considerations that genuinely support them, and to end up acting or believing as we should as a result of this.² Whatever the particular purpose we're using our reason for,

² For this type of argument, see, for example, Sher 1997 and Enoch 2011. Sher puts the point in a very congenial fashion: "basing one's decision on one's weightiest combination of reasons is the generic aim of all practical deliberation" (1997, 205). See also Sosa 2007 on apt performances.

we will fail in one way if we don't respond correctly to the relevant reasons. This is what I mean when I say that reason has an implicit formal aim or telos.

We make use of our reason for two broad kinds of purposes: to change the world and to comprehend the world. The difference between them yields the distinction between practical and theoretical reason. Practical reason is concerned with action (and related attitudes). Any intentional action has three elements: the (first-order) end, the means we take to the end, and the resulting state of affairs. By the internal standards of practical reason, we are successful, roughly speaking, when a) the end is sufficiently valuable to warrant the investment we make, b) we pursue the end competently (for example, the means we take are efficient in the light of our evidence and don't undermine our other pursuits), and c) the world is friendly enough for the end to be actually realized, in part through our competent efforts (which is to say that part of the formal aim of practical reason is that one doesn't do what one should merely by luck). In brief, success in the formal aim consists in competently realizing valuable ends that are in harmony with each other. In the case of theoretical reason, the direction of fit is different: instead of changing the world to match what we have in mind, we aim at grasping the structure of the world in important respects. Very briefly, we're successful when we competently (and thus non-accidentally) form a correct unified explanatory conception about some sufficiently important subject matter on the basis of evidence that does not turn out to be misleading (see Kauppinen 2022).

Focusing on the formal aims implicit in the way our capacities function yields a natural way to understand flourishing. As George Sher says, on this picture "what has inherent value is not the mere exercise of a fundamental capacity, but rather its successful exercise as measured by the achievement of its defining goal" (1997, 202). I part ways with Sher when it comes to understanding the notions of fundamentality and defining goal, and would add that for an agential capacity, the success must be *because* of competent exercise

(cf. Sosa 2007), but on the basic point I'm in agreement. So the first part of my revision of traditional perfectionism is replacing the activity interpretation with the following:

The Telic Interpretation of Flourishing and Unflourishing

Flourishing consists in successfully realizing the formal aims implicit in the functioning of our fundamental capacities to a sufficient degree. Unflourishing in some respect consists in frustrating a formal aim, or realizing it to an insufficient degree.

So far, I've talked about success and flourishing, but what about unflourishing? As Shelly Kagan (2014) rightly emphasizes, when success requires many elements, there will be many ways to fail, which may not all be equally bad. To take the paradigm agential capacity of practical reason, the first-order end might not be worth pursuing, we might pursue it incompetently, and we might not realize it. And several of these might happen at the same time – I might incompetently pursue an end not worth realizing, for example, and fail to realize it. (Think of a dictator sending a bunch of undermotivated soldiers to conquer a neighbouring country and failing.)

Not every kind of failure to flourish, however, is robustly bad rather than merely the absence of a good. Most obviously, it seems that failing to exercise essential agential capacities is an absence of a welfare good, and thus only privatively bad, or absence of flourishing.³ What if we do exercise a capacity? Let's say that a pursuit *fails* if a rational capacity is exercised but the end doesn't come about. After all, as Lilian O'Brien (2019) emphasizes, it is plausibly constitutive of intending, as opposed to merely desiring, that one *regards oneself* as an agential failure if one doesn't perform the act one set out to perform, in

³ As Jason Raibley (2022) has recently observed, getting a zero on the scale of well-being can nevertheless amount to *ill-being* in the case of some goods – a person who never exercises reason fares *badly*.

spite of believing it was feasible. (That's why Kagan's (2014) observation that not getting what one wants is just the absence of a good doesn't apply to intentions.) Let's further say that a pursuit *predictably fails* if the failure results from incompetent exercise. I would suggest that while failure may be only privatively bad in the case of rational capacities, predictable failure is robustly bad. That's because predictable failure non-accidentally frustrates not only the agent's particular end, but also the formal aim of reason, since it results from poor use of reason. So, for example, if you try to shoot an arrow at a target and make a good effort, but a gust of wind blows it off, you've failed (possibly through no fault of your own). If you perform the same way, but the target turns out to have been a projection that vanishes as soon as you shoot, you arguably didn't fail, since the conditions for exercising your capacity were not in place. In these cases, you miss out on a good, so what happens is only privatively bad for you. But finally, if you predictably fail because you just don't know how to shoot but nevertheless go ahead with the pursuit, your failure *both* by the standards imposed by your end and the internal standards of practical reason is robustly bad for you.⁴

Suppose we accept this view of flourishing in terms of the formal aims of fundamental capacities. The next natural question is: which of our capacities are fundamental to us, if we don't want to get entangled in debates about human nature? At this point Sher appeals to goals that are "near universal and near unavoidable" (1997, 236–239). While his account avoids the problem with assuming a human nature in the classical sense, it remains a challenge to explain why successfully realizing near universal and near unavoidable goals would be in itself good *for* a subject, when realizing other goals isn't. Still, there is definitely

⁴ Bradford (this volume) appeals to what she calls 'contrastive ends', like running a race faster *rather than* slower than last year, to account for failure. On her view, if you have such an end and run more slowly than last year, you've attained the 'negative contrastive' of your end, which you aimed against, and thus failed. As noted in the main text, I think having an end automatically involves being set against falling short of it, so that there's automatically a negative contrastive. One difference between the views is that I hold that if you failed in spite of making an effort that might easily have resulted in a better time, you didn't *predictably* fail, in which case your failure isn't robustly bad for you.

something to Sher's idea. In emphasizing a certain class of goals that agents have, he captures something about the idea that welfare goods and bads are *subject-relative*. As subjectivists about well-being have rightly emphasized, for something to be good for *me*, it has to *fit* or *suit* me. It doesn't suffice for it to be good from the point of view of the universe, or simply make me a good member of a kind. As Connie Rosati puts it in the course of defending a subjectivist view, "the good of a creature must *suit its own nature*" (Rosati 1996, 323 my emphasis).

For subjectivists, it's the *attitudes* of a subject – typically, their values or desires – in suitable circumstances that determine our nature in the relevant sense, and thus what fits us. But perfectionists can give a broader, and in my view more plausible, account of our nature. They can say that who I am, my own nature in the relevant sense, is determined by *how my self-defining capacities function*. It's not the attitudes that we end up with but how and why we get there that shows what we're made of. After all, it makes perfect sense to ask if our values or desires – or even systems of values or desires – reflect who we really are. For example, if I'm manipulated into valuing something, it isn't going to be something that suits *me*. And what is it that makes instilling values an instance of manipulation? The standard answer, given by people like Alfred Mele (1995), is that it's bypassing the agent's rational capacities. This immediately suggests that it is how we use our rational capacities that at least in part determines who we fundamentally are. Whether or not rationality is a part of human nature, there's a very strong case to be made for it being a self-defining capacity for you and I, as I have tacitly assumed in defending the telic interpretation of flourishing.

Of course, there are agents, such as (most) other animals, whose selves, insofar as they have them, are defined by instincts that both tell them what to pursue and how to pursue it (Korsgaard 2018). But in our case, even the way in which our most basic instincts are expressed is shaped by our own take on reasons and the relative importance of different

goods. As Kantians like Christine Korsgaard rightly emphasize, we are faced with practical questions we simply have to answer for ourselves, and in so doing make use of our rational capacities (which should not be confused with intelligence). Similarly, our beliefs are our own doing insofar as they amount to our own answer to a question about whether something is the case – as Pamela Hieronymi (2008) puts it, we exercise evaluative control over our beliefs by figuring out what is true. These familiar considerations support the idea that how we use our rational capacities in part defines who we really are, and thus what kind of things fit or suit us. Reason is thus fundamental to our nature in the sense relevant to flourishing, even if we think of our nature in terms of self-defining capacities rather than what makes us human.⁵

At the same time, there is more to our kind of subjectivity than the way we exercise our reason. After all, it is a striking fact about you and me that we're not *zombies*. There is something it's like to be me, and that may be different from what it's like to be you. And it's not just that I have *some* sort of conscious experiences that determines, for its part, who I am. My experiences can have *valence*, whether positive or negative. We have pleasures and pains, positive and negative emotions, felt desires and aversions. Like our take on reasons, these valenced responses constitute a part of our own perspective on the world – what we enjoy and what we suffer from says a lot about us, and our values and desires can fail to match it. Importantly, our affective tendencies speak for us even when we don't reflectively endorse them. As Nomy Arpaly and Timothy Schroeder (1999) observed, there is more to our 'whole self' than the attitudes we would like ourselves to have. Even if I can't help enjoying something I regard as bad, my pleasure isn't alien to me in the way a manipulated desire might be, but just the way *I* happen to be. So I take it the capacity for *valenced experience* is

⁵ Small children are plausibly only developing this capacity, or possess it to a lesser degree. It follows that for telic perfectionists, what is basically good for a small child is not what is basically good for an adult.

another self-defining capacity for us (and other sentient creatures), and thus fundamental to who we are. It's not an agential capacity, a capacity that we intentionally exercise, but rather a capacity that *functions* or *operates* when we have experiences that feel good or bad.

Bearing these considerations in mind, my proposed replacement for the Human Nature Thesis is the following:

The Subjective Nature Thesis

Our fundamental capacities in the sense relevant for well-being are those whose functioning defines who we are. In the adult human case, they include at least the practical and theoretical rationality and the capacity for valenced experience.⁶

Putting this thesis together with telic perfectionism yields a view of well-being that could be labelled Subjective Nature Telic Perfectionism: roughly, we flourish when we are successful by the internal standards of rationality and the internal standards of valenced experience. I've already discussed what kind of standards are implicit in the use of reason. But what about valenced experience?

3. The Nature and Telos of Valenced Experience

On the face of it, valenced experiences *happen* to us. They're not exercises of agency, so they're not goal-directed in the way that intentional actions are. In lifting a bale I'm aiming to feed a horse, but in feeling pain in my ear I'm not thereby aiming at anything. However, I'm going to argue that we can nevertheless attribute an overall telos to valenced experience on the basis of its nature and function. Specifically, I'll argue that the formal aim implicit in the functioning of our capacity for negatively valenced experience is not having negatively

⁶ There may be other capacities that fundamentally define us as the kind of subjects we are. For example, a good case to be made for our capacity to relate to other subjects as subjects and not just as worldly objects. Arto Laitinen has pressed me on this for years, and he's probably right. See Laitinen forthcoming.

valenced experiences, or self-elimination for short. The exact basis for attributing this formal aim depends on which theory of valenced phenomenal character is correct.

Let's start with *externalist* views. They try to explain valence in terms of attitudes toward the sensation or experience in question – for example, a sensation or experience is unpleasant because we intrinsically desire not to have it when we have it (Heathwood 2007). It would not be difficult to argue on this basis that unpleasant experience contains a self-eliminative *telos* given by the partially constitutive desire. However, the traditional stumbling block for externalist views is that such desires seem secondary to the phenomenal character of the experience itself – it's because the experience is *already* pleasant or unpleasant that we have intrinsic desires concerning it. Moreover, we have *reason* to get rid of unpleasant pain, but the mere fact that we want something not to be the case doesn't generate a reason to do away with it (Bain 2017, 468–470).

Internalist views, in contrast, hold that some experiences inherently possess an *aversive* phenomenal character, while others are inherently *attractive*. As these very terms suggest, valenced experiences thus appear to give direction to our agency in virtue of their feel. While this might be a brute fact, within the philosophy of mind, *intentionalist* accounts of valence have recently become popular. Generally speaking, what makes an account intentionalist is appealing to the *content* of the experiences to account for the positive or negative phenomenal character. (A minority phenomenalist view explains content by phenomenal character (see Kriegel (ed.) 2013); what I say in the following can also be modulated into the phenomenalist key.) Such views originate in debates about explaining the qualitative character of consciousness in general, in which people like Michael Tye (1995) have long argued that what it's like to have an experience is determined by (if not identical with) its intentional content.

Some experiential contents, like the content of visual experience, represent non-evaluative features. But most agree that *valenced* phenomenal character requires a different type of content. The two main schools here are *evaluativist* and *imperativist*. The evaluativist approach holds that experiences feel good in virtue of representing something as good, and bad in virtue of representing something as bad (Cutter and Tye 2011, Carruthers 2018). Let's focus here on pain, since it is our main concern. Because of the phenomenon of pain asymbolia, in which people on certain drugs report feeling pain but deny that it is unpleasant, it is now common to distinguish between the *sensory* phenomenology of pain and its usual *affective* phenomenology of unpleasantness. (In the following, when I talk about pain without qualification, I will always mean unpleasant pain.) Here how David Bain formulates the evaluativist view:

A subject's being in unpleasant pain consists in his (i) undergoing an experience (the pain) that represents a disturbance of a certain sort, and (ii) that same experience additionally representing the disturbance as bad for him in the bodily sense. (Bain 2013, 582)

Here, the subject's experience *feels* bad (and consequently *is* bad) in crucial part in virtue of *representing* a bodily disturbance as bad. Now, there are certainly ways of representing something as bad that don't themselves feel bad at all, so evaluativists insist that what is at issue is some sort of non-conceptual representation of badness (Carruthers 2018, 663–4; Bain 2019, 483–4). This proposal is somewhat difficult to evaluate, since it's not clear what it is to represent something non-conceptually *as* bad for me – indeed, I've argued that the best explanation of what is consists in appeals to non-representational content (Kauppinen 2021b).

While evaluativist views are popular, they face some fundamental questions. One key challenge is that valenced experiences have what Luca Barlassina and Max Hayward call

intrinsic motivational force: they can motivate us independently of desires or other conative states (2019, 1016).⁷ The question for evaluativists is: why would a state that non-conceptually represents something as good or bad have such intrinsic motivational force? Analogously to non-cognitivists in ethics, who account for the motivational role of moral beliefs by appealing at least in part to non-representational content, imperativists maintain that valenced experiences intrinsically motivate, because they have non-representational *imperative* content.⁸ As Barlassina and Hayward (2019, 1018) summarize it, representational content has the function of carrying information and truth conditions that determine which information is carried. It is taken up by an audience when it forms a belief with the same content. Imperative content, in turn, has the function of directing the addressee to do something which is determined by its satisfaction conditions, and is taken up by an audience when it is motivated accordingly.

What kind of imperative content might valenced experiences have? In the case of pain, Manolo Martinez (2011) has defended first-order imperativism, according to which an unpleasant pain sensation has two kinds of (nonconceptual) content, indicative (which could be paraphrased as “There’s damage to your ankle”) and imperative (“See to it that the damage to your ankle goes away!”), which together account for its sensory and affective phenomenology, respectively. Accepting the command that pain gives us will then on its own motivate us to attend to the bodily damage. More recently, Barlassina and Hayward have argued that what is distinctive of valenced experiences is their intrinsic *reflexive* motivational force: unpleasant pains motivate us to get rid of themselves and pleasures motivate us to get more of themselves. To capture this, they propose, in a nutshell, that “an experience is

⁷ As Mauro Rossi pointed out, this motivational claim is less plausible for moods. An imperativist might here emphasize the “can” part of “can intrinsically motivate”, since it does seem that moods can intrinsically motivate at least mental action, such as attending to something, or letting one’s mind wander.

⁸ Alternatively, the experience presents standard representational content with a mental analogue of imperative or directive (rather than assertoric) illocutionary *force*. For my purposes this would work as well.

pleasant/unpleasant in virtue of commanding us: More/less of me!” (2019, 1032; cf. Prinz 2004). This reflexive imperativist hypothesis would readily explain why unpleasant pain intrinsically motivates us to take painkillers, for example.

While my own view is closest to Martinez’s, for my purposes here, I want to be neutral between the different intentionalist accounts. The key question at this point is how we get from such views to my thesis about the *telos* of negatively valenced experience. Here, we should think of the formal aim as the state in which negatively valenced experience as a whole comes to rest – where it is no longer functional, because there’s no more job for it to do by its own lights.⁹ And that, I claim, is the state in which we no longer have negatively valenced experience. What I want to argue for next is that each of the competing accounts of the phenomenal character of experience I’ve introduced supports this thesis.

To begin with, the reflexive imperativist view says that pain directly tells us to get rid of the experience it’s a part of and thus itself, and possibly indirectly guides us to see to bodily disturbance as an effective means to this. By itself, this doesn’t yield my thesis about the formal aim. After all, if your headache just tells you to get rid of the experience it’s a part of, this command is satisfied if I electrocute you, so that you have a new, more painful experience. But if we think of valenced experience as a whole, for reflexive imperativism, it will have done its job only when we’ve got rid of *all* the experiences that tell us to get rid of themselves. Less directly, the same goes for first-order imperativism and evaluativism. According to them, pain and other bad feelings either present things in our body or world as bad or tell us to change their current state, perhaps in specific ways. As I said, it’s not quite clear what it is to non-conceptually represent something as bad, but evaluativists wouldn’t deny that at least part of what is distinctive of *evaluative* representation is to guide action –

⁹ Pain might be dysfunctional when there’s nothing we can do about it, but this is clearly different from lacking a further function in virtue of having fulfilled the function of the capacity.

very crudely put, to replace the bad with the good. First-order imperativists unambiguously hold the unpleasant experiences tell us to change the non-evaluative states of affairs they represent. On either view, we can attribute a world-directed aim to individual experiences. But if we think of negatively valenced experience as a whole, there's no particular way the world needs to be for it to come to fruition. What we can say, however, is that the world needs to be such that it does not evoke negatively valenced experience – that is, we must have reached the state in which experiences of this type are absent. In this indirect fashion, the functioning of this capacity implicitly contains the formal aim of getting rid of this type of experience even according to accounts that emphasize world-directed content.

It is worth briefly noting that this conception of the formal aim of unpleasant pain meshes well with current accounts of the adaptive *function* of pain. For example, according to Seymour's reinforcement learning model, "pain must be understood in the context of behavioral control to minimize current and future harm" (2019, 1030). Its key function, beyond motivating reflexive avoidance behaviour, is to serve as an internal reinforcement signal for learning, leading agents to update the expected value of potential actions in a way that results in less harm (presumably understood as bodily dysfunction or disturbance of homeostasis) in the future. Insofar as acute pain, then, is a conscious teaching signal whose job is to guide us away from behaviours and circumstances that give rise to it, and which has fulfilled its function when the stimulus changes and the signal itself goes away, it looks like we can attribute the self-eliminative aim to it on a functional basis, too.

I'll summarize these observations as follows in the following thesis that is compatible with almost all current accounts of the phenomenal character of pain and other unpleasant experiences:

The Telic Character of Pain

In virtue of the phenomenal character of unpleasant pain and associated intentional content, the formal aim implicit in the operation of the capacity for negatively valenced experience is elimination of all such experiences.

As I said at the beginning of this section, we shouldn't think of the telos of unpleasant pain in the same way as the telos of intentional action. Rather, it offers an end for activity, which we may or may not take up in motivation or deliberation – in some cases, we might not want to reduce unpleasant pain for some independent reason, and in other cases might not be able to do anything about it. When we're not able to change a situation our experience tells us to change, we may experience *suffering*, which, I've argued, goes beyond physical pain in involving pervasive negative affective construal of our situation (Kauppinen 2019). Suffering can obviously have other sources, such as the loss of a loved one, but as an experience, it will share the same self-referential telos as unpleasant pain in addition to its world-directed content.

Does pain always direct us towards self-elimination? A seeming counterexample is provided by what Bradford (2020) calls 'hurts-so-good experiences', such as eating very spicy food or jumping in a cold lake. On her view, we find unpleasantness enjoyable and welcome it when we have such experiences (2020, 239). On the face of it, this is in tension with my claim that unpleasantness calls for self-elimination. But I think a more plausible view of the phenomenology of hurts-so-good experiences is that when we have them, we take higher-order pleasure in having a *mix* of attractive (or neutral) and aversive qualia, so that the presence of the latter raises the experience as a whole above a more 'vanilla' version. It's not that the 'burning tingle' (Bradford 2020, 240) of spicy food in itself tells us to get more of it, but that the *combination* of the 'more, more!' directive of (say) the sweetness and the 'no more!' directive of the burning tingle is more appealing than sweetness alone. I agree with

Bradford that the burning tingle isn't "merely tolerated" and that the whole experience is pursued in part because of it, but this is compatible with that unpleasant aspect of the experience by itself telling you to get rid of itself. After all, if you *only* had the burning tingle, the experience wouldn't tell you to get more of the same. (Note that if someone experiences the burning tingle itself as pleasant, then it's not a *hurts-so-good* experience in the first place!)

4. Experiential Unflourishing

We now have the materials we need for a perfectionist explanation of the badness of pain.

The argument can be simply stated:

The Perfectionist Badness of Pain

1. In virtue of the phenomenal character of unpleasant pain and associated intentional content, the formal aim implicit in the operation of the capacity for negatively valenced experience is elimination of all such experiences. (The Telic Character of Pain)
2. Our capacity for negatively valenced experience is one of our self-defining capacities. (From the Subjective Nature Thesis)
3. So, whenever we are in unpleasant pain, the formal aim of one of our self-defining capacities is frustrated. (From 1 and 2)
4. Unflourishing as the kind of beings we fundamentally are is constituted by frustrating the formal aims implicit in the operation of our fundamental capacities. (From The Telic Interpretation of Flourishing and Unflourishing)
5. So, being in unpleasant pain is a form of unflourishing as the kind of beings we fundamentally are. (From 3 and 4)

6. Things that constitute unflourishing as the kind of beings we fundamentally are are in themselves bad for us. (From Explanatory Perfectionism)
7. So, being in unpleasant pain is in itself bad for us. (From 5 and 6)

Does this explanation yield the result that unpleasant pain is robustly bad for us precisely because of how it feels? As I said in the introduction, there's a clear sense in which it does. If pain felt different – say, if one just had the pain sensation but didn't find it unpleasant, as some patients on morphine report – it wouldn't be the kind of experience that sets for us an aim that is frustrated as long as it is ongoing. Further, while all unpleasant pains frustrate this aim, more intense pains frustrate it to a greater extent, since they're further away from realizing it on a scale established by those experiences themselves – ecumenically put, things are changed in the direction of matching the content of pain (or, on externalist views, the content of desires constitutive of it) when it becomes milder, but not when it becomes more intense. That's why it follows from the theory that more intense pains are worse than less intense ones, other things being equal.

To be sure, as a telic perfectionist account, the explanation I've given of pain's badness fundamentally appeals to frustrating formal aims. This may still trouble some people, perhaps for reasons analogous with some familiar objections to rule consequentialism or Kantianism in normative ethics. For example, some might say in the latter context that we should keep our promises just because we've made them, not because a general practice of sticking to them yields optimal results or because we somehow contradict ourselves otherwise. It's not as if being inconsistent is somehow worse than failing to keep your word! But this type of objection rests on a confusion of levels of explanation. In normative ethics, we can say that an act's wrongness is explained by certain wrong-making features, like the fact that you promised to do otherwise, while deep ethical theories offer competing

explanations of why certain features are wrong-making. The idea isn't that what such explanations appeal to is the wrongest of wrong things. Rather, they appeal to something that unifies wrong things and illuminates their status, given certain assumptions about the nature of morality. Just the same goes for explanatory perfectionism. It's not committed to claiming that frustrating our formal aims is worse for us than agony without hope, but just that the former provides a unifying explanation that is illuminating, given certain independently plausible assumptions about well-being, for example that it is subject-relative and that different kinds of things are welfare goods and bads for different kinds of subjects. For short, formal aims have *explanatory priority*, not *evaluative priority*.

It should be clear at this point how my account diverges from Bradford's (2021) tripartite perfectionist explanation. On her view, it is our contingent if nearly universal *personal-level aims* that get frustrated by being in pain, and this, in turn, amounts to a failure of *practical rationality* – not in terms of exercising it, as for traditional developmental perfectionism, but in terms of its functional output. In contrast, for the telic perfectionist account, pain and suffering frustrate the formal aim implicit in negatively valenced *experiences themselves*, regardless of our contingent personal-level aims. For this reason, too, pain and suffering amount to failure by the internal standards of the *capacity for valenced experience*, not some other capacity. This is why it avoids Guy Fletcher's (2022) objection to Bradford, which is based on the plausible idea that pain is bad for infants, even though they plausibly lack the standing person-level aim to avoid feeling pain. The pain *itself* is telling the infant to get rid of it, as it were, and that's part of what makes it a pain. It's also worth noting that pain need not be failure in terms of the *function* of the capacity, since having experiences that guide us away from bodily disturbance or danger may well be functional. Indeed, pain typically functions well when it offers the *aim* of bringing about a pain-free state for our

agency to take up, even though that aim remains unrealized while the pain is ongoing. This is why it matters that telic perfectionism focuses on formal aims rather than functions.

A further nice feature of the present view is that it yields a nuanced explanation of the prudential disvalue of *negative emotions*. The key to it is that emotions involve *both* the exercise of a form of rational agency and valenced experience. Of course, we don't 'form' emotions in the way we sometimes form plans, for example, after deliberating on pros and cons. But emotional responses can manifest responsiveness to reasons and be fitting or unfitting depending on the situation (e.g. Tappolet 2016). For example, if you lose someone you love, you have reason to think about them and the changes in the meaning of activities and things that results from the loss. On my view of fittingness (Kauppinen 2014), grief is fitting, since it inherently motivates such mental activity (indeed, grief might be the only way of properly registering the significance of the loss for you). At the same time, grief is, of course, an unpleasant feeling.

So suppose that I have solid evidence that a good friend of mine died, and I'm sad. Is having this emotion good or bad for me? Well, it manifests responsiveness to a genuine reason, and thus realizes the telos of reason to some degree. In this respect, it is good for me that I grieve to something like the appropriate extent (if I grieve too much or too little, my exercise of rationality has failed to some extent). At the same time, qua unpleasant experience, grief frustrates the telos of valenced experience to some degree. In this respect, it is bad for me that I grieve. The contribution of negative fitting emotions to well-being is thus ambivalent. In contrast, there's nothing to be said for feeling bad without good evidence that things are going badly – such emotions amount to frustrating both the telos of rational agency and valenced experience. Similarly, feeling good for bad reasons has an ambivalent status – whether it's all-things-considered good for me depends on the strength of the reasons and the pleasure, and the correct way to weigh these prudential goods against each other, which I

haven't attempted to offer here. And finally, fitting positive emotions are doubly good for us, as it were. Perhaps that's why it's better for us to love and experience the pleasures of love than just to experience equally intense purely sensory pleasure.

Conclusion

To sum up, in this chapter, I've focused on the most serious challenge to the extensional adequacy of perfectionism about well-being: how can we make sense of the badness of experiences, when they don't constitute anything like a failure of a rational capacity that we can develop and exercise? My answer has two key elements. First, perfectionists should say that what is basically good or bad for us is not just engaging in activities in which we exercise our fundamental capacities, but rather realizations or frustrations of aims that are implicit in the functioning of our fundamental capacities. Second, while some first-order ends are set by and realized in and through rational activity, others are offered up to us by the phenomenal character of valenced experience and realized when that experience is positive and frustrated when it is negative. That's why telic perfectionism says that the best kind of life for us – a life in which we harmoniously realize the various formal aims of our self-defining capacities and thus flourish as the kind of beings we fundamentally are – is a life in which we successfully exercise our rational capacities in pursuit of freely chosen, objectively good particular ends on the basis of a comprehensive understanding of the world, and have an excellent time doing so.¹⁰ The worst kind of life, in turn, consists of predictably failed or worthless pursuits, ignorance, and suffering. That seems pretty compelling to me.¹¹

¹⁰ I first proposed an enumeratively dualistic view of well-being as consisting in agential success and positive experience in Kauppinen 2012.

¹¹ This paper benefited greatly from generous written comments not only by Mauro Rossi and Christine Tappolet but also Gwen Bradford, Lorenza D'Angelo, and Lilian O'Brien, to whom I owe special gratitude.

References

- Arneson, Richard J. (1999). Human Flourishing Versus Desire Satisfaction. *Social Philosophy and Policy* 16 (1): 113–142.
- Arpaly, Nomy and Schroeder, Timothy (1999). Praise, Blame, and the Whole Self. *Philosophical Studies* 93 (2): 161–188.
- Bain, David (2013). What Makes Pains Unpleasant? *Philosophical Studies* 166 (1): 69–89.
- Bain, David (2019). Why Take Painkillers? *Noûs* 2017 (2): 462–490.
- Barlassina, Luca & Hayward, Max Khan (2019). More of me! Less of me!: Reflexive Imperativism about Affective Phenomenal Character. *Mind* 128 (512):1013–1044.
- Bradford, Gwen (2015). Perfectionism. In Guy Fletcher (ed.), *The Routledge Handbook of Philosophy of Well-Being*.
- Bradford, Gwen (2017). Problems for Perfectionism. *Utilitas* 29 (3):344-364.
- Bradford, Gwen (2020). The Badness of Pain. *Utilitas* 32: 236–252.
- Bradford, Gwen (2021). Perfectionist Bads. *Philosophical Quarterly* 71 (3):586-604.
- Bradford, Gwen (this volume). Failure.
- Carruthers, Peter (2018). Valence and Value. *Philosophy and Phenomenological Research* 97 (3):658-680.
- Cutter, Brian & Tye, Michael (2011). Tracking Representationalism and the Painfulness of Pain. *Philosophical Issues* 21 (1):90-109.
- Dorsey, Dale (2010). Three Arguments for Perfectionism. *Noûs* 44 (1): 59–79.
- Enoch, David (2011). *Taking Morality Seriously: A Defense of Robust Realism*. Oxford University Press.
- Fletcher, Guy (2016). *The Philosophy of Well-Being: An Introduction*. Routledge.
- Fletcher, Guy (2022). A Painful End for Perfectionism? *Midwest Studies in Philosophy* 46: 233–250.

- Haybron, Dan (ms). *The Lives We Should Want*.
- Heathwood, Chris (2007). The reduction of sensory pleasure to desire. *Philosophical Studies* 133 (1):23-44.
- Hieronymi, Pamela (2008). Responsibility for Believing. *Synthese* 161 (3):357-373.
- Hurka, Thomas (1993). *Perfectionism*. Oxford University Press.
- Kagan, Shelly (2014). An Introduction to Ill-Being. *Oxford Studies in Normative Ethics* 4: 261-88.
- Kauppinen, Antti (2012). Meaningfulness and Time. *Philosophy and Phenomenological Research* 84 (2), 345–377.
- Kauppinen, Antti (2014). Fittingness and Idealization. *Ethics* 124 (3): 572–588.
- Kauppinen, Antti (2019). The World According to Suffering. In David Bain, Michael Brady, and Jennifer Corns (eds.) *Philosophy of Suffering*. Routledge.
- Kauppinen, Antti (2021a). Rationality as the Rule of Reason. *Noûs* 55 (3), 538–559.
- Kauppinen, Antti (2021b). Relational Imperativism about Affective Valence. *Oxford Studies in Philosophy of Mind* 1:341-371
- Kauppinen, Antti (2022). Epistemic Welfare Bads and Other Failures of Reason. *Midwest Studies in Philosophy* 46: 251–279.
- Kitcher, Philip (1999). Essence and Perfection. *Ethics* 110 (1):59-83.
- Korsgaard, Christine (2018). *Fellow Creatures: Our Obligations to the Other Animals*. Oxford University Press.
- Kraut, Richard (1994). Desire and the Human Good. *Proceedings and Addresses of the American Philosophical Association* 68 (2):315
- Kraut, Richard (2007). *What is Good and Why: The Ethics of Well-Being*. Harvard University Press.
- Kriegel, Uriah (ed.) 2013. *Phenomenal Intentionality*. New York: Oxford University Press.

- Laitinen, Arto (forthcoming). Mutual Recognition and Well-Being: What Is It for Relational Selves to Thrive? In Onni Hirvonen and Heikki Koskinen (eds.) *Theory and Practice of Recognition*. London: Routledge.
- Martínez, Manolo (2011). Imperative Content and the Painfulness of Pain. *Phenomenology and the Cognitive Sciences* 10 (1):67-90.
- Nussbaum, Martha C. (2000). *Women and Human Development: The Capabilities Approach*. Cambridge University Press.
- O'Brien, Lilian (2019). The Subjective Authority of Intention. *Philosophical Quarterly* 69 (275):354-373.
- Prinz, Jesse (2004). *Gut Reactions*. New York: Oxford University Press.
- Raibley, Jason (2022). Less than Zero. *Midwest Studies in Philosophy* 46: 193–232.
- Rosati, Connie (1996). Internalism and the Good for a Person. *Ethics* 106 (2): 297-326.
- Sher, George (1997). *Beyond Neutrality: Perfectionism and Politics*. Cambridge: Cambridge University Press.
- Sen, Amartya (1987). *On Ethics and Economics*. Blackwell, Oxford.
- Seymour, Ben (2019). Pain: A Precision Signal for Reinforcement Learning and Control. *Neuron* 101: 1029–1041.
- Sosa, Ernest (2007). *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Vol. 1*. Clarendon Press, Oxford.
- Sumner, L. W. (1996). *Welfare, Happiness, and Ethics*. Oxford University Press.
- Tappolet, Christine (2016). *Emotions, Value, and Agency*. Oxford University Press.
- Tye, Michael (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. MIT Press.