

Affect, Rationalization, and Motivation*

Jonathan Cohen[†] and Matthew Fulkerson[‡]

Unreflective cognitive processes — or mere information processing — takes place in ‘the space of causes’ while reflection occurs, instead, in the ‘space of reasons’. This characterization, of course, nicely encapsulates the perspectival shift I have been focusing on. Unreflective processes are naturally viewed from a third-person perspective, and so we see them in mechanistic terms: inhabitants of the space of causes. But when we view reflective processes, instead, from the first-person point of view, we now view our internal states, not as states which are causally related to one another, but, instead, as ones which may serve as reasons for one another: they thereby leave the space of causes and enter the space of reasons. The space of reasons thus becomes a mysterious realm, one which occupies a place devoid of event-causal and nomological connections. There is, of course, no such place. The temptation to believe in it only emerges as a product of the perspectival shift (Kornblith 2012, 160).

Consider poor Lucy, a machinist who’s having a bit of bad luck. While working at the bench, she is momentarily distracted and strikes her thumb hard with a ball-peen hammer. This, as you might expect, results in a great deal of pain. Recent debates about the nature of the unpleasantness of pain have concerned two different roles that pain has been thought to play in a creature like Lucy.

First, pain *motivates*: feeling it typically moves us to act, and often in predictable ways. Thus, for example, we can imagine Lucy acting in various ways to this unfortunate incident. Screaming would be understandable, though not the sort of thing stoic Lucy would do. Still, even she would wince heavily and pull her arm away, immediately taking ameliorative actions on her throbbing thumb — holding it, squeezing it, shaking it. And it seems plausible that Lucy’s pain motivates these actions. Without the pain she wouldn’t have winced or grabbed her thumb, and her performing these actions is a direct consequence of her experience of pain.

In addition, however, pain has been thought to *rationalize*; it provides a reason for our actions — a positive consideration that can justify the actions for the subject

*This work is fully collaborative; the authors are listed alphabetically.

[†]Department of Philosophy, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0119, joncohen@aardvark.ucsd.edu

[‡]Department of Philosophy, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0119, mfulkerson@ucsd.edu

who undertakes them. Thus, the thought is that Lucy's pain (or its unpleasantness) is not only a cause of her pulling her arm away, but also a reason — and a reason for *her* as opposed to some subpersonal mechanism within her — that justifies or makes rational her pulling her arm away. If Lucy had moved her arm in that way in the absence of the pain (and in the absence of some other reason) then her doing so would not have been rational or justified.¹

There is nothing special about our choice of example here. Pain (or more accurately, the painfulness of pain—a qualification we will sometimes drop) is merely a paradigm instance of what we shall call *affect*.² Affective experiences have some quality or character that drives or motivates us to act in various ways. They often seem to accomplish this by being in some sense bad or unpleasant (in the negative cases) or pleasant and enjoyable (in the positive cases). The class of motivating affective states is quite expansive, including drive states like hunger, thirst, itch, tiredness, and sexual desire, along with the full gamut of sensory pains and pleasures from headaches and nausea to the smell of vanilla and the taste of chocolate cake. For ease of exposition, and following established precedent in the literature, we will focus much of what follows on the example of painful experiences. Even so, the points we make carry over wholesale to other affective experiences.

Recently, a number of writers have presented an argument to the effect that the interplay between the two roles of affective experience make trouble for leading (“causal”) theories. Specifically, they have urged that such causal theories make available accounts of affect's motivational role, but at the cost of failing to understand affect's rationalizing role. Moreover, these writers have gone on to argue that these considerations support the adoption of an alternative (“evaluationist”) conception of pleasure and pain that, in their view, successfully explains both the motivational and rationalizing roles of affective experience.

We believe that this argument from rationalization is ineffective in choosing between evaluationist and causal theories, and that the impression to the contrary rests on a serious misunderstanding of the dialectic between the two views. We'll begin by describing general forms of causal and evaluationist theories, setting out the argument that has been deployed by evaluationists against causal theorists, and arguing for the importance of this consideration by showing that it underpins several other of the evaluationist's anti-causal arguments (§1). Having thus emphasized the dialectical significance of the argument from rationalization, we'll show how that argument rests on crucial and highly controversial presuppositions,

¹The question of rational connection arises in several otherwise disjoint areas of philosophy. Its natural home, of course, is in moral psychology and decision theory, where the question specifically concerns the nature of reasons and our motivations for action (see, e.g. Smith 1987; Parfit and Broome 1997; Mele 2003).

²Affect is also sometimes referred to as the *hedonic* or *valenced* component of motivating experiences.

including some that the evaluationist is particularly ill-placed to accept (§2). Finally, we'll conclude (§3).

We want to be clear that, though we aim to show that the argument from rationalization fails in the anti-causal role to which it has been put by evaluationists, our limited goals in this paper do not include that of refuting evaluationism or defending any particular causal theory of affective experience. Instead, our goals are to make clear the underlying theoretical commitments at play in the debate, to show that one central and popular class of arguments against causal theories rests on an insufficient appreciation of the theoretical landscape, and to invite a reconceptualization of the terms of the debate.

1 Causation and rationalization

1.1 Rationalization and causation

The jumping off point for the present dispute involves the observation that many of the standard accounts of affect's motivational role involve pointing to a causal route — viz., to the idea that the feeling of painfulness motivates Lucy to react in the way she does because it causes those reactions.³ There are, of course, many ways for the causal connection to be realized, and these differences might matter in our explanation of her specific actions (and of course, the routes will likely differ for different forms of affect). Pulling her hand back from the pain might be a purely reflexive action resulting from the activity of spinal interneurons. This sort of reflex cause wouldn't take any route through Lucy's consciousness, and would be more or less automatic in standard conditions. Other causal routes might be more specifically dispositional or statistical in nature. For instance, feeling pain might increase the probability of her acting in various ways. Being in pain makes it much more likely, but need not guarantee, that she will say 'ouch.' And of course, there could be more flexible reactions caused by her propositional attitudes and background cognitive states. For example, she might continue her rubbing because she *believes* that it will soothe her pain, and this belief, along with the painful experience, is the proximate cause of her acting in this way.

The foregoing is, of course, highly schematic: there remain many questions left to be asked about such causal explanations for pain's motivational role. Supposing, however, that those questions could be answered, and that appeals to one or another causal route could provide a satisfactory account of how pain motivates,

³For this reason, we refer to this class of views as "causal" in nature. As we'll describe below, this class is wide, including many individual views, all versions of functionalism, and, importantly for us, imperative views. It is also neutral with respect to one's theory of causation.

we could still ask: would we then be in a position to account for the other role of pain we highlighted above — its role as a rationalizer of action?

Recently, a number of writers (including Helm 2002; O’Sullivan and Schroer 2012; Bain 2012) have argued that we would not. According to these philosophers, no purely causal connection can *rationalize* Lucy’s actions; no causal explanation can make the latter explicable as the deliberate actions of a rational agent or explain the reasons for her actions from her perspective as an agent. Thus, David Bain (2012) writes that,

[U]npleasant pain doesn’t merely cause such behaviour, but rationalises it, providing the reason for which an agent performs it. Unpleasant pain makes available a distinctively perspectival kind of explanation, unavailable in the case of reflexive behaviour, for instance (ms, 1).

Similarly, Bennett Helm (2002) rejects a particular causal account (that pain causes desires for it to stop) because it fails to account for the rationalizing power of pains:

The mere fact that some quale typically causes the desire to eliminate it is insufficient to account for the kind of reason pains provide for action, for such an appeal to desire leaves the question of such rational warrant unanswered: what reason does pain provide for our having such a desire? In the case of the qualia view, there can be no answer (17–18).

These philosophers have used this consideration — henceforth, the ARGUMENT FROM RATIONALIZATION — both to rule out causal accounts of affect’s unpleasantness, and to motivate their preferred alternative (about which see below).⁴

Now, though the argument from rationalization is in principle applicable to a number of distinct, broadly causal accounts of affect, it has been deployed most often against rival views of pain (including dispositional, functional, and psychofunctional accounts). And among these rivals, it has to date been deployed most frequently against the so-called imperative accounts of pain associated with Hall (2008), Klein (2007) and Martínez (2011).⁵

⁴ Though we take the argument from rationalization as our central focus in what follows, we emphasize that evaluationists differ in the extent to which they have rested their cases for the view (and against alternatives) on this argument. Thus, for example, while Bain (2012) directs a version of the argument against imperativists, he joins others who do not (e.g., Cutter and Tye 2011) in offering additional arguments against the imperative view based on the supposed failure of commands to account for pain’s degrees of intensity. Again, we do not see ourselves as intervening here on behalf of any particular causal view, or defending such views against all possible lines of criticism; our focus here is squarely on what we’re calling the argument from rationalization.

⁵ Again, highlighting the generality of this discussion, we note that imperative views have been deployed as an explanation of other states besides pain. Indeed, Hall’s paper is

On Klein's (pleasingly austere) version of the view, which we use for illustration, the painfulness of pains consists entirely in their negative imperative content:

The content of any pain is a negative imperative. The imperative force of pains is thus to proscribe rather than prescribe. What unifies the above pains is the imperative that I stop doing what I am doing; their content is a proscription against action (Klein 2007, 520).

The pain of a broken ankle is an imperative for me not to use my ankle. Give me enough morphine for my pain to be lessened, and you lessen the strength of that imperative. Give me enough to eliminate my pain, and you have eliminated the imperative against action entirely. Nothing more need be postulated to explain my phenomenology, nor to explain the distinctive contribution that phenomenology makes to my behavior. (Klein 2007, 522).

These views have it that the content they posit is imperative, or commanded, in the sense that they are sensory 'urges' that do not require any additional cognitive states such as desire in order to be motivating (Hall 2008). As such, these views build in to the nature of pain (by fiat) an account of the latter's motivational role. In so doing, they place themselves firmly alongside other views like functionalism that see a strong causal (or even constitutive) connection between pains and our reactions to them. So far so good. But, argue the critics, tying pains and what they motivate together in this way cannot account for the second aspect of pains and other motivating states: like other causal accounts, these critics allege, imperativism is left unable to account for the rationalizing role of pains:⁶

primarily an account of itch. Other states that seem amenable to an imperative account include thirst, hunger, and other basic urges. We believe that the points made here about rationalization and motivation carry over wholesale to the related debates about the connection between reason and action for such urges (for instance, one might be tempted to claim that an urge cannot rationalize if it is entirely constituted by a particular disposition or reflexive action).

⁶Note that this criticism does not assume that the commands in question are merely reflexive. For as imperativists make clear, the commands do not compel us by literally forcing us to react in various ways:

All these imperative bodily sensations can occur with different degrees of intensity, and the more intense, the harder to resist. But even at their most irresistible, these imperative sensations and their resultant actions are completely different from true reflexes, like the knee jerk. Reflexes should be viewed either as purely mechanical linkages not involving any sensory messages with intentional content at all, or else, if involving intentional content, then as imperatives that go to some lower level action centre (physically in the spinal cord, perhaps), not accessible to, and not under the

Why should receiving that command be a good reason or even appear to you to be a good reason — for lifting your foot? Both imperativism and the desire view invoke inclinations; and while imperativism explains the inclinations in terms of the subject's receipt of certain commands, my worry is that this explanation does nothing to illuminate what the desire view left unexplained: how the inclinations rationalise or justify action (Bain 2012, 7).

1.2 Rationalization and evaluation

But proponents of the argument from rationalization have not been content to use that argument as reason to reject imperativism and other causal theories of affective experience. On the contrary, they have also appealed to the argument to motivate a positive view that, in their estimation, provides a more satisfying account of affect's rationalizing force. Roughly, they contend that we can repair what they take to be the crucial shortcoming of causal theories (viz., their lack of any element with rationalizing force) by adding into affect's nature something that does have rationalizing force.

Thus, Helm (2002); O'Sullivan and Schroer (2012); Bain (2012) argue for what we'll call an EVALUATIVE account — one according to which Lucy behaves as she does because the pain in her thumb consists of a special evaluative representation that the pain *is bad*.⁷ Importantly, 'bad' in this context is an evaluative predicate, assigned as a result of the particular way in which Lucy's pain is represented.⁸

This is taken to be something of a broad normative or evaluative property (in contrast to the descriptive contents found in most garden variety perceptual experiences). Her immediate reflexive responses to the pain do not seem to

control of, the beliefs and desires of the Higher Cognitive Centre (Hall 2008, 534–535).

⁷We reiterate that the argument from rationalization is not the only motivation evaluationists have put forward for their view, and that there are evaluationists (e.g., Cutter and Tye 2011) who don't endorse the argument at all (cf. note 4).

⁸The evaluative view is thus importantly different from the related 'attitudinal' views usually applied to sensory pleasure. According to attitudinal theories, for an experience to be pleasant (or painful) is for the agent to take a positive (or negative) attitude toward that experience. The evaluationist does not explicitly posit a separate mental attitude as constitutive of the pleasant or painful aspects of the experience. Instead, the pain or pleasure state consists entirely of a special kind of evaluative content. Nevertheless, there are important points of overlap, especially in the versions defended by Helm and O'Sullivan & Schroer, since they make use of broad notions like *concern* that seem to go beyond the content of any single state. The "attitudinal" terminology is found in Feldman (2002); a recent defense of such views can be found in Heathwood (2006) — see Smuts (2011) and Bramble (2013) for critical discussion.

have such evaluative content; the protective action was initiated without any representation of the state's badness (since, by hypothesis, the reaction occurred prior to the formation of any evaluative contents). Rather than a rationalized action, Lucy's reflexive response was just something that happened to her (her arm was flung back). Her later actions, however, can be understood as having the required contents: her rubbing and seeking an ice pack is presumably done because her experience tells her that something about her bodily condition (specifically, something with her arm) is bad, and being a rational agent, she seeks however she can to rid herself of this bodily badness.⁹

We can explain this behavior as stemming from her judging, based on the content of her experience, that the pain in her thumb is bad and that she would much prefer it to cease (combined with her belief that rubbing may bring about that cessation). Crucially, the thought goes that the evaluative representation of the pain as bad (and the ensuing judgments and beliefs) are suited to rationalize in a way that the elements in a causal chain are not — hence that an account of pain's unpleasantness that essentially includes such representations will succeed in accounting for pain's rationalizing role (and so will meet the demands of the argument from rationalization) where causal accounts failed. Problem solved.

Or is it?

1.3 The argument from rationalization

Before we attempt to answer this question (negatively), and to show that the argument from rationalization fails to advance the dispute between evaluationists and causal theorists, it will be helpful to state the argument in a way that reveals its underlying assumptions. Thus, here is the generalized negative form of the argument for pain, as we understand it:¹⁰

- (1) Any theory of pain should account for pain's rational role.
- (2) *In general*, causal views can't account for pain's rational role.

⁹Evaluationists differ in how, exactly, they cash out such evaluative contents (and some of the accounts are quite involved). While we won't catalog the options here, the details do matter; as we'll see, it will be almost impossible for the evaluationist to account for these sorts of contents in a way that preserves for them the force of the argument from rationalization.

¹⁰Though we present here a version of the argument targeting imperativism for pain (which is standardly understood as a species of representationalism about its target), the argument can be easily adapted to any other causal view of affect, including especially functional and dispositional accounts of pain (which are not). Given that this is so, and that our discussion below abstracts from issues about representationalism, it will be useful to confine ourselves to the version targeting imperativism for reasons of specificity and simplicity.

- (3) The imperative view of pain is a causal view.
- (4) The imperative view cannot account for pain's rational role.
- (5) **Therefore**, imperative views are inadequate accounts of pain.

We should note that there is another possible way of stating the argument. Instead of (2) and (3), the evaluationist might have the following in mind:

- (2*) Causes of type Z cannot rationalize.
- (3*) The imperative view involves causes of type Z.

The rest of the argument is as before; but on this reading the evaluationist is trying to suggest that the particular causal story invoked by the target view is the sort of thing that cannot rationalize. There are hints of this strategy when, for instance, the evaluationist suggests that *commands* can't rationalize (for instance, because if they are always obeyed then they are like reflexes, which can't rationalize; and if not always obeyed we require an additional explanation beyond the command itself). While we will address this more narrow version in what follows, we focus most of our efforts on the general version of the argument. As we see it, if the evaluationist has (2*) or (3*) in mind, then we require considerably more elaboration and defense of the claim that only certain causes can rationalize. In addition, as we'll see, it becomes very difficult to see how the evaluationist can sustain (2*) without undermining the rationalizing force of her own view.

2 Reasons and causes revisited

The argument from rationalization is a centerpiece of evaluationist criticisms of causal theories (including, but not limited to imperativism about pain). However, we now want to urge that this argument is unpersuasive.

2.1 The role of rationalization

The first observation we want to make is that much in these waters seems to hang on our assessment of premise (2) in particular.

First, we hope it is clear that (2) is the crucial premise in the argument from rationalization, as we have presented it. We can imagine someone denying (1) for various reasons, perhaps because of a general skeptical attitude toward reason-talk of this kind. And perhaps someone taken with an austere causal account of affect might also be more inclined toward such skepticism about reasons. But most imperativists, indeed most causal theorists, don't take such a stand. Instead, they all seem to take their view to provide the best account of affect's rationalizing

role. So we don't want any criticism of the argument to rely on forcing the causal theorists into a radical rejection of all reason-talk. That leaves premise (3); but since the various views in question are, in the very broad sense in which we use the term, clearly causal, there should be no debate on this point. As such, it would seem that all the work of the argument is being done by (2).

Indeed, we believe that (2) underlies not only the argument from rationalization, but several (ostensibly) other arguments typically deployed against causal theories as well.

For example, one such additional argument against causal accounts rests on the thought that (i) such views can only understand the connection between the reaction and the hurtfulness of pain as a relation of identity or a necessary causal connection, and that either relation is too intimate to provide for the kind of space required for one to be a reason for the other (Helm (2002, 2–3)). Or, again, often with reference to Plato's *Euthyphro*, critics have sometimes contended that (ii) causal accounts get the order of explanation between motivation and the painfulness of pain backwards: the claim is that we should explain the motivation inherent in pain experiences by appeal to its badness, whereas such accounts can only explain its badness by appeal to our reactions (O'Sullivan and Schroer (2012, 750); (Bain 2011, 169); Bain (2012, 12); used against motivational theories of pleasure: Smuts (2011, 242–243)). Finally, one occasionally sees the crucial premise in the argument from rationalization (viz., (2)) run together with the claim that (iii) causes lack a perspectival quality that attaches to reasons.¹¹

But we now want to suggest that each of (i)–(iii) is ultimately dependent on something like premise (2) in the general argument from rationalization.

For instance, if, *pace* premise (2) of that argument, causes could rationalize, then (i), by itself, would be rendered innocuous: it would amount more or less to the claim that some causes (viz., those with too close a relation between cause and effect) don't rationalize. But the typical causal theorist should be fine with this; the view after all isn't that every causal relation counts as a reason, but only that some of them do.

As for (ii), if it were the case that some causes could rationalize, then we could legitimately set the causal constituents of our sensory reasons alongside the person-level considerations in a way that we couldn't if we held fast to the rational impotence of causes. That is, the causal elements of our sensory systems would be allowed to play in the reason game, and the explanatory asymmetry could be easily averted by noting that we can still be motivated to avoid pain (at the person level) because it hurts (at the sensory level), even when the sensory level is cashed out in purely causal terms. So again, the essential premise (2) of the argument is doing all of the work. We'd add that, as a general matter, order of explanation

¹¹This claim about perspectives is mentioned somewhat offhandedly in Bain (2012); it is not clear whether he takes it as an unrelated argument against causal accounts or a piece of support for (2).

arguments are very hard to make in a way that convinces the other side, and this instance is no different. As usual, it's plausible that we lack theory-independent access to the explanatory order, so claims about explanatory priority are more reasonably treated as spoils to the victor rather than as independently established premises that can be assumed before the other issues are settled. Moreover, and in particular, if we don't assume that causes are rationally impotent, then issues about the explanatory priority between pain's badness and its motivational power will seem especially unobvious. Once again, therefore, argument (ii) seems to depend crucially on the claim we take as our main target below.

Finally, (iii) is, similarly, tied up intimately with the idea of the rational impotence of causes, as articulated by (2). As far as we know, no one in the literature has offered independent support for (iii). On the contrary, (iii) appears to be accepted by all and only those already persuaded that causes are rationally impotent — and, we suspect, for the very same reasons. So again, if causes could rationalize, it would become a much more open possibility that a causal relation (of a certain kind) could be appropriately perspectival.

These considerations suggest, then, that much depends on our assessment of (2).

On reflection, however, it appears that the evaluationist needs yet more. For the argument from rationalization won't function in its positive role in motivating evaluationism without the assumption (for which no argument is offered) that the separation between causes and reasons underlying (2) works only in one direction: while causes are unable to rationalize, evaluative contents have no problem causing or motivating our typical behavioral responses to painful stimuli. We think therefore that the evaluationist, in deploying the argument from rationalization, is actually committed to two distinct but interrelated assumptions: that causes cannot rationalize, and also that evaluative contents (broadly construed) can motivate. Alas, we now want to suggest that the evaluationist cannot have it both ways. Indeed, we do not think the evaluationist can have it *either* way. In the following sections we present arguments against both assumptions — first against the assumption that causes cannot rationalize (§§2.2–2.3), and then against the claim that evaluations can motivate (§2.4).

2.2 Causes and reasons

We've already urged that the pivotal premise in the argument from rationalization — and, therefore, additionally, in the several other arguments that have been used against causal theorists — is (2), the claim that causes are incapable of rationalizing actions. But if so much is hanging on that premise, then we should ask: What reasons are available to the evaluationist in support of (2)? We want to answer that there are very few — both because (2) is extremely controversial and, because,

even worse, (2) turns out to be *inconsistent* with positive aspects of the evaluationist view.

Unfortunately, evaluationists themselves give us little to go on: they have said little in defense of (2) (beyond asserting claims that entail/presuppose its truth). That said, the very insouciance of standard presentations of the argument from rationalization, and the absence of specific discussions of the requirements on rationalizing actions (as opposed to other things that can be rationalized, such as beliefs), suggests that proponents are thinking of the premise as a consequence of a more general shortcoming of causes — viz., their inability to rationalize anything at all.¹² If so, then the premise under consideration should be regarded as a special case of the more general thesis that causes are incapable of supplying rational warrant. So, then: should we believe *that* thesis?

That depends.

To be sure, the thesis of the general rational impotence of causes can be seen as stemming from a set of views about epistemic justification that boasts a distinguished pedigree (defenders include Sellars 1956; Davidson 1986; Alston 1989). Roughly, these views have it that beliefs possess, and causes lack, the normative properties necessary for the justification of belief, and therefore that there's a species of rational motivational force that accrues to beliefs (denizens of the "space of reasons") but not causes (denizens of the entirely separate "space of causes"). On the other hand, such views are, to put it mildly, not universally accepted. On the contrary, they are rejected by a very wide range of theorists, including (but not limited to) those sympathetic to more externalist conceptions of justification (prominent representative examples include Goldman 1967; Dretske 1970; Plantinga 1993; Kornblith 2002)). Indeed, it is hard to think of a more controversial thesis in epistemology (or, therefore, in philosophy).

On the other hand, while the debate has been slightly less heated, it is widely thought to be problematic to assume that evaluative contents can have causal efficacy without some story for how these causal powers are realized (especially for those sympathetic to various forms of naturalism about such contents). This can be thought of as a special case of the general worries about mental causation.¹³ The

¹²Our reconstruction of the support for the premise is, of necessity, largely conjectural, and hence possibly inaccurate. We certainly don't mean to close off the possibility that someone might attempt to argue for (2) by considerations stemming from more specific requirements of rationalizing action in particular (while allowing that causes might have what it takes to provide rational warrant for things other than actions). (This would be the possibility alluded to above as (2*).) We find it difficult to imagine just how such an argument would go: offhand, it's hard to see why action should impose further, special requirements on its rationalizers that go beyond the requirements on rationalization in general. But if someone has an argument of this kind to offer, we're all ears.

¹³For instance, how can high-level states like beliefs have causal efficacy if they are realized by the components of relatively low-level neural systems? (See e.g., Kim 1992, 1998).

details of the story one might tell do not matter for our purposes here; but, as we'll see, assuming that evaluative content can have causal efficacy has consequences for the evaluationist. In particular, it becomes difficult to maintain the separation of causes from reasons in one direction while heartily endorsing the connection going the other way.¹⁴

Of course, we cannot hope to settle the debates over the nature of epistemic justification or mental causation here. But the complexity and controversy attending these debates just underlines our point. Namely, it suggests that these crucial presuppositions of the argument from rationalization — that causes cannot rationalize, but that evaluative content *can* have motivating causal powers — should not be allowed to pass without defense.

We believe that this observation, by itself, reduces considerably the force of both the evaluationist's negative use of the argument from rationalization against causal theorists and her positive use of the argument to motivate evaluationism. However, we now want to argue that this understates the badness of the situation for evaluationists — that (given plausible assumptions they themselves accept), evaluationists can only endorse the needed presuppositions (controversial though they are) at the cost of undercutting their own account of the rationalizing role of affect.

2.3 Can evaluations rationalize?

To see why, recall that, for the evaluationist, affect rationalizes because it constitutively involves evaluative contents (as opposed to mere causes). Of course, for this form of explanation to count as an advance over that offered by causal theorists, it had better turn out that evaluative contents are better suited to rationalize than the mere causes that evaluationists disparage as rationally impotent. So, are they?

Not if, as extant evaluationists have exclusively held, the relevant bits of evaluative content are to be understood naturalistically.¹⁵ Thus, for example, Cutter and Tye (2011) hold that the represented evaluative badness at issue “is just the property of being apt to harm you” (99); and they understand aptness to harm in terms of a naturalistic notion of proper function (for pleasures, they replace ‘apt to harm’ with ‘apt to benefit’). Helm (2002, 23) proposes a related view on which

¹⁴Moreover, as we'll argue below (§2.4) even if the asymmetry itself turns out to be unproblematic, the evaluationist *still* has to buy into at least one additional substantive controversial thesis — motivational internalism — to get her view on the playing field. (Cf. Helm 2002, note 3).

¹⁵Extant evaluationists have also, as far as we know, typically endorsed naturalistic theories (e.g., some version of a tracking theory) of the representation relation that connects subjects to such contents. Helm (2002) is a bit more circumspect about how he plans to cash out the notions of ‘import’ and ‘concern’ used in his theory, but nothing he says indicates a non-naturalized psychosemantics.

badness is aptness to harm, in a context in which the subject cares about his or her proper functioning. Similarly, if less specifically, O'Sullivan and Schroer (2012) defend a version of the view on which the source of the rationalizing force of pain's unpleasantness is a "representational content that concerns one's own body" (740); and they tell us that the view they are defending "is 'reductive' in that it analyses the notion of 'having a particular representational content in a certain manner' in entirely physicalist-friendly terms" (739).¹⁶

Despite their (important) differences, these forms of evaluationism are united in taking the rationalizing evaluative component of affect to be naturalizable, viz., reducible to something causal.¹⁷ But if so, then the evaluationist account of affect, and of the the rationalizing role of pains in particular — just like the causal theorist's account — will only succeed in rationalizing on the condition that mere causes *can* amount to reasons. We saw that evaluationists were unwilling to accept that condition when it came to the assessment of causal theories. Our present point is that a (naturalistically inclined) evaluationist explanation of the rationalizing role of affect is in just as much need of that condition as a causal theory's explanation of the rationalizing role of affect.

In effect, given the evaluationist's commitment to naturalization, her view turns out to be just one more species of causal account — one on which the *because* in the claim 'Lucy winces because her experience contains evaluative content' must be read causally. This species is no better or worse situated to account for the rationalizing role of pain's unpleasantness than any other. If causes can provide reasons, then causal theories (of whatever type) can answer that demand; if not, not. In neither case (and much literature to the contrary notwithstanding) does the naturalizing evaluationist view carry any advantage over the imperativist or other causal views concerning the possibility of accounting for the rationalizing role of affect.

We hasten to add that there is a way out of this situation for evaluationists: they can give up their commitment to naturalism. But, to our minds, this seems a steep price to pay (and one that is at odds with the explicitly stated purposes of the proponents of the view quoted above).

¹⁶Needless to say, many other naturalistic accounts of the evaluationist's crucial evaluative content are possible. We take this list to be representative but not exhaustive.

¹⁷Recall that the notion of 'cause' here is very broadly understood to include dispositional, functional, psychofunctional, and statistical forms of interaction. The reduction at issue thus does not require reduction to particular physical mechanisms. Given the broad notion of cause used here, we don't think it's controversial to assume that the realization base of our best scientific explanations (that is, naturalism) must consist of some combination of causal elements.

2.4 Can evaluations motivate?

We have been arguing that the motivation for (2) in the evaluationist's argument from rationalization is both highly controversial (2.2) and unavailable to those evaluationists who endorse naturalism about evaluative contents — which, in practice, means all evaluationists (§2.3). But, as we noted in §2.1, the evaluationist needs to argue for more than just (2). She needs also to show that the separation between causes and reasons is asymmetric — that causes are rationally impotent, but that evaluative contents are causally/motivationally potent. Unfortunately, it would appear that the evaluationist can only secure the latter position on the assumption of yet another controversial and unargued theoretical commitment, viz., a kind of practical motivational internalism.

To see this, recall that the evaluationist is committed to holding that her understanding of affect explains both its rational and its motivational roles. Now, the elements evaluationists invoke to explain affect's rational role are, as we have seen, evaluative. Hence, evaluationists are committed to the idea that those very same evaluative elements are motivating.

Moreover, and importantly, evaluationists insist that pains themselves — rather than any further desires or other conative states typically but inessentially connected with pains — are practically motivating. This is what leads Bain (2011), for instance, to open his discussion this way: “Two obvious truths can seem to be neglected by the ‘perceptualist’ view that pains are perceptual experiences: first, that pains are typically unpleasant; second, that they drive us to act, e.g. to remove our fingers from a flame” (164). Bain and other evaluationists take it to be an obvious truth that pain itself is practically motivating, that it drives us to act; and they use this obvious fact as a starting point from which to run the argument from rationalization. But in saying that the evaluational states (as opposed to further inessentially connected states) motivate, and do so in a way that is not contingent on other considerations like desires, the ability to act, and the absence of other overriding concerns, evaluationists commit themselves to a controversial (and undefended) form of internalism about practical motivation. This form of internalism requires that there be a necessary connection between our evaluative states and our overall motivation to act. That is, if one is in the appropriate evaluative state, then, necessarily, one is motivated to act with respect to that evaluative state. This is a strong requirement.

Might evaluationists simply dispense with the commitment to motivational internalism?¹⁸ Not if they want to continue to press the argument from rationalization against causal theorists. For an evaluationist who allowed that pains are externally practically motivating would open the door to a causal theorist who

¹⁸Besides avoiding the controversial commitment, allowing that pains motivate only in the presence of further conative elements might also carry the additional benefit of providing a story about the otherwise problematic morphine pain and asymbolia cases.

claims that pains are, similarly, only externally rationalizing (contra premise (2) of the argument from rationalization). This causal theorist would hold that, it is not pain itself, but rather some numerically distinct evaluative state typically but inessentially connected with pain, that accounts for pain's rationalizing role. Since allowing for this response would undercut the argument from rationalization, and therefore would leave evaluationism unmotivated, we take it that evaluationists will not want to go down this road.

It would seem, then, that evaluationists have the sort of account of the motivating role of pains that they take themselves to have only if internalism about practical motivation is true. Now, for all we know, that sort of motivational internalism might be true. On the other hand, there are also reasons for doubt. One such reason comes from the apparently coherent case of the amoralist offered by Brink (1997) as a challenge to the idea that evaluative judgments are internally morally motivating. The amoralist is a subject who, after undergoing frontal lobe trauma, makes all of the same moral evaluative judgments as before the brain damage, but no longer feels motivated to act appropriately on the basis of these judgments. Such cases, if indeed possible, seem to provide a serious impediment to any full-throated internalism about moral motivation. As Brink notes, "Where there is such physical and psychological interference, practical judgment does not produce motivation. If so, we must deny that judgments of practical reason entail motivation" (Brink 1997, 17). But if Brink is right in thinking that evaluative content might supply moral motivation only externally, then it's hard to see without further argument why the same might not be true about practical motivation — i.e., why there could not be an "apracticalist" who, after brain trauma, forms all of the same practical evaluations of her bodily condition as before the injury, but who no longer feels motivated to act appropriately on the basis of these evaluative states. And this is just to say that we are owed reasons for thinking that the kind of internalism about practical motivation required by the evaluationist is mandatory.

Of course, what we've said is by no means the end of the story, and we can't hope to settle the disputes about either moral or practical motivational internalism here.¹⁹ We do hope to have made clear, however, that in requiring the truth of internalism about practical motivation, the evaluationist is holding her view hostage to premises at least as vexed and controversial as the view that causal relations can provide justifying reasons.

Before closing this section we'd like to add that, while the seemingly required assumption of motivational internalism here is telling, there are further reasons for worrying that the evaluationist's conception of pain's motivational role is at odds with her larger picture of the relation between causes and reasons. For suppose

¹⁹Good entries into the debates about moral and practical motivation, with a particular focus on the debate between internalists and externalists, include Railton (1986); Smith (1994); Brink (1997); Svavarsdottir (1999); Shafer-Landau (2000); Wallace (1990); Pettit and Smith (1990); Helm (2001).

the evaluationist is right in saying that evaluative elements constituting pain's unpleasantness are internally practically motivating. Given the standard view that motivation runs through a causal route (discussed above), this would mean that the relevant evaluative bits are internally motivating only by being causally potent. But this conclusion, once again, conflicts with the evaluationist's assumption of a divorce between the space of causes and the space of reasons (as discussed in §2.2).

3 Conclusion

We opened with the observation that the debate between evaluationists and causal theorists involves the interplay between the rationalizing and motivational roles of affect. Proponents of the argument from rationalization have put forward that argument as a way of establishing what they take to be a significant asymmetry with respect to these two desiderata depending on whether we construe affect in causal or evaluational terms. Namely, they have taken the argument to show that pain as construed by causal theorists is motivationally potent but rationally impotent, while pain as construed by evaluationists is both motivationally and rationally potent. What we have seen in the foregoing is that every part of this description of the situation is up for grabs (or worse).

First, we argued, the idea that the causal theorist's understanding of affect is rationally impotent is deeply worrisome. This idea is extremely controversial among contemporary philosophers, and, in the present context, borders on question-begging. Moreover, and perhaps even more telling, the idea is only available to evaluationists willing to give up their naturalistic aspirations (as extant evaluationists have not been). Second, we noted that the evaluationist's account of affect's motivational power depends on supposing, again controversially, both that reasons can be causally efficacious and that practical motivational internalism is true.

Finally, the very asymmetry underlying the argument from rationalization is itself suspect. Though we are not ourselves much sympathetic to a rupture between the space of reasons and a space of causes, at least we know what a view of that sort amounts to. What is much odder, it seems to us, is the conjunctive view, required by the argument, that (i) causes are unable to penetrate the space of reasons, but (ii) reasons are nonetheless able to penetrate the space of causes. If we are right about this, then we are owed a broader defense of this picture before taking the argument from rationalization, in anything like its current form, as compelling.

We do not take ourselves to have shown that causal theories of affect are true, or that evaluative theories are false. But we do take ourselves to have shown that

the central argument in the area — one that has often been taken to choose between the main theories in contention — is unpersuasive and dialectically ineffective.²⁰

References

- Alston, W. (1989). *Epistemic Justification*. Cornell University Press, Ithaca.
- Bain, D. (2011). The imperative view of pain. *Journal of Consciousness Studies*, **18**(9-10), 164–85.
- Bain, D. (2012). What makes pains unpleasant? *Philosophical Studies*.
- Bramble, B. (2013). The distinctive feeling theory of pleasure. *Philosophical Studies*, **162**(2), 201–217.
- Brink, D. O. (1997). Moral motivation. *Ethics*, **108**(1), 4–32.
- Cutter, B. and Tye, M. (2011). Tracking representationalism and the painfulness of pain. *Philosophical Issues*, **21**(1), 90–109.
- Davidson, D. (1986). A coherence theory of truth and knowledge. In E. Lepore, editor, *Truth and Interpretation. Perspectives on the Philosophy of Donald Davidson*, pages 307–319. Basil Blackwell.
- Dretske, F. (1970). Epistemic operators. *The Journal of Philosophy*, **67**, 1007–1023.
- Feldman, F. (2002). The Good Life: A defense of attitudinal hedonism. *Philosophy and Phenomenological Research*, **65**(3), 604–628.
- Goldman, A. I. (1967). A causal theory of knowing. *Journal of Philosophy*, **64**(12), 357–372.
- Hall, R. J. (2008). If it itches, scratch! *Australasian Journal of Philosophy*, **86**(4), 525–535.
- Heathwood, C. (2006). The reduction of sensory pleasure to desire. *Philosophical Studies*, **133**(1), 23–44.
- Helm, B. W. (2001). Emotions and practical reason: Rethinking evaluation and motivation. *Noûs*, **35**(2), 190–213.
- Helm, B. W. (2002). Felt evaluations: A theory of pleasure and pain. *American Philosophical Quarterly*, **39**(1), 13–30.

²⁰We are grateful to Murat Aydede, David Brink, and an anonymous referee for extremely helpful comments.

- Kim, J. (1992). Multiple realization and the metaphysics of reduction. *Philosophy and Phenomenological Research*, **52**, 1–26. Reprinted in Kim (1993), 309–335.
- Kim, J. (1993). *Supervenience and Mind: Selected Philosophical Essays*. Cambridge University Press, New York.
- Kim, J. (1998). *Mind in a Physical World: An Essay on the Mind–Body Problem and Mental Causation*. MIT Press, Cambridge, Massachusetts. Originally given as the 1996 Townsend Lectures.
- Klein, C. (2007). An imperative theory of pain. *The Journal of Philosophy*, **104**(10), 517–532.
- Kornblith, H. (2002). *Knowledge and its Place in Nature*. Oxford University Press.
- Kornblith, H. (2012). *On Reflection*. Oxford University Press, New York.
- Martínez, M. (2011). Imperative content and the painfulness of pain. *Phenomenology and the Cognitive Sciences*, **10**(1), 67–90.
- Mele, A. R. (2003). *Motivation and Agency*. Oxford University Press.
- O’Sullivan, B. and Schroer, R. (2012). Painful reasons: Representationalism as a theory of pain. *Philosophical Quarterly*, **62**(249), 737–758.
- Parfit, D. and Broome, J. (1997). Reasons and motivation. In *Proceedings of the Aristotelian Society*.
- Pettit, P. and Smith, M. (1990). Backgrounding desire. *Philosophical Review*, **99**(4), 565–592.
- Plantinga, A. (1993). *Warrant and Proper Function*. Oxford University Press.
- Railton, P. (1986). Moral realism. *Philosophical Review*, **95**(2), 163–207.
- Sellars, W. (1956). Empiricism and the philosophy of mind. In H. Feigl and M. Scriven, editors, *Minnesota Studies in the Philosophy of Science*, volume 1, pages 253–329. University of Minnesota Press, Minneapolis.
- Shafer-Landau, R. (2000). A defense of motivational externalism. *Philosophical Studies*, **97**, 267–291.
- Smith, M. (1987). The Humean Theory of Motivation. *Mind, New Series*, **96**(381), 36–61.
- Smith, M. (1994). *The Moral Problem*. Blackwell, Oxford.

Smuts, A. (2011). The feels good theory of pleasure. *Philosophical Studies*, **155**(2), 241–265.

Svavarsdottir, S. (1999). Moral Cognitivism and Motivation. *The Philosophical Review*, **108**(2), 161–219.

Wallace, R. J. (1990). How to argue about practical reason. *Mind*, **99**(395), 355–385.