Synthese (2025) 205:178
https://doi.org/10.1007/s11229-025-04928-5

ORIGINAL RESEARCH

Reflective equilibrium, aesthetic appreciation, and aesthetic judgement

Murray Smith[1]

Received: 3 March 2023 / Accepted: 16 January 2025

Abstract

I explore the relevance of reflective equilibrium to aesthetic experience, judgement, and appreciation, framing this project in both descriptive and normative terms. I begin with an exposition of reflective equilibrium, construed as a theory of epistemic justification, and of aesthetic appreciation, characterised as an activity with an epistemic goal—namely, achieving an optimal understanding of an artwork or aesthetic object. I stress the broad purview of reflective equilibrium, as well as Rawls' arguments on the importance of aesthetic activity in human experience, and thus its place in a theory of justice. These preliminary considerations set up a more detailed exploration of reflective equilibrium in action in the aesthetic domain. I develop the argument with reference to a range of cases, from mainstream filmmaking to twelve-tone composition, considering the place of reflective equilibrium in aesthetic appreciation in both its individual and collective guises, as well as its "wide" and "narrow" forms. I demonstrate how the application of reflective equilibrium in the context of the critical appreciation of art relates to a range of significant issues in aesthetics, via the examination of arguments from Sontag, Sibley, Walton, Carroll, and Gaut (among others). Finally I assess various objections to, and extensions of, the main argument, including the objection that aesthetic appreciation does not aim at understanding, and the proposal that the embodied, emotional dimension of aesthetic experience, which must be accommodated by reflective equilibrium in the context of aesthetics, may shed light on its nature and operation in other domains.

Keywords Reflective equilibrium · Aesthetic appreciation · Aesthetic experience · Aesthetic judgement · Artistic value · Aesthetic value · Art · Wide and narrow reflective equilibrium · Individual and collective reflective equilibrium · Hard cases · Imagination · Emotion · Embodied cognition · Reflective afterlife · Film · Novel · Song · Music

✉ Murray Smith
mss@kent.ac.uk

[1] Professor of Philosophy, Art, and Film, University of Kent, Canterbury, United Kingdom

Published online: 18 April 2025

> Although in these essays I do talk a great deal about particular works of art and, implicitly, about the tasks of the critic, I am aware that little of what is assembled in this book counts as criticism proper…most of it could perhaps be called meta-criticism…I wanted to expose and clarify the theoretical assumptions underlying specific judgments and tastes.
>
> Susan Sontag, *Against Interpretation*

Is reflective equilibrium (RE) relevant to aesthetic appreciation? That is, to what extent is our engagement with and appraisal of works of art and other objects of aesthetic attention characterised by the dynamic adjustment of and mutual support among our beliefs, "back and forth" (Rawls,1971/1999, p. 18) between our judgements concerning particular events and objects, and the more abstract principles which bear upon them? Running alongside this descriptive question is its normative sibling: to what extent *should* our engagement with and appraisal of artworks and other aesthetic objects be governed by such deliberation? To answer these questions, we will first need to establish working definitions and a fuller understanding of the two phenomena this pair of questions brings together.

1 Reflective equilibrium, aesthetics, and human flourishing

Let's begin with RE. Most of the research conducted on RE has been conducted within the domains of moral and political philosophy, and this, along with the fact that the expression "reflective equilibrium" was coined by John Rawls in his immensely influential *A Theory of Justice*, leads to the widespread idea that RE is primarily or solely pertinent to ethics. But the purview of RE is much broader, as was clear from the moment that Rawls baptized the phenomenon. "The process of mutual adjustment of principles and considered judgments", writes Rawls, "is not peculiar to moral philosophy" (Rawls, 1971/1999, p. 18 n7). In the same passage, Rawls goes on to note "parallel remarks" by Nelson Goodman on the justification of the principles of deductive and inductive inference in his *Fact*, *Fiction*, *and Forecast* (1954).[1] We might also note here that Catherine Elgin, one of Goodman's co-authors, has explored the role of reflective equilibrium in the criticism of the literary, performing, and visual arts, among other areas (Elgin, 1996, 2017). Thus both the history and the substance of debate suggest that, at the most abstract level, the scope of reflective equilibrium is much wider than moral and political deliberation, potentially encompassing everything from deductive logic to aesthetic judgement.

Understanding the broad scope of RE is a first step in motivating and justifying the current enquiry. For if RE is—or is a crucial ingredient in—a general philosophical methodology, as it is according to one influential conception (Lewis, 1983, p. 10;

[1] Rawls' first articulation of reflective equilibrium in "Outline of a Decision Procedure for Ethics" in 1951 appears to have just preceded Goodman's formulation—the lectures which led to *Fact*, *Fiction and Forecast* having been delivered in May 1953, with the book published in 1954. But Rawls does not use the phrase "reflective equilibrium" in the 1951 article; that doesn't appear until the first edition of *A Theory of Justice* in 1971.

Beebee, 2017; see also Walden, 2013), then it will be relevant to any and all philosophical endeavours (or at least, all those endeavours which broadly accord with the account of philosophy RE provides). Given that aesthetics is a branch of philosophy, and aesthetic appreciation is one of the major phenomena explored by aesthetics, we can proceed on the basis that RE should have some kind of pertinence to the domain of aesthetics. And yet there is a striking gap in the research here—in fact, a gap to the power of two, for just as there has been very little attention to aesthetics within the literature on RE, neither is there any sustained and significant body of thinking about RE within aesthetics and the philosophy of art. Elgin's work is very much the exception rather than the rule on both counts. Surveying the existing work on RE, Claus Beisbart and Georg Brun make the related point that there is a need for domain-specific case studies (2024, p. 79). Whatever else it is, the present essay is an attempt to rectify this situation.

A slightly narrower interpretation of the status of RE, though one which still conceives of it as central to the enterprise of philosophy, construes and develops it as a theory of *epistemic justification*—an account of the norms by which epistemic claims (that is, claims to knowledge, truth, or understanding) should be justified. Understood in this way, RE competes with other theories of justification, such as reliabilism and the various forms of foundationalism. Within the theory, the expression "reflective equilibrium" refers both to a state (of optimal coherence among the beliefs of an individual or a group) and a process (through which that optimal state is sought). An agent begins with a set of beliefs about a given phenomenon, which may vary in their level of abstraction, from very specific, "considered judgements" about particular objects or events, to more general principles relevant to these particulars. Through the "process of equilibration" (Beisbart and Brun, 2024, p. 78), the agent reflects on and iteratively adjusts the lower- and higher-order beliefs, seeking to bring them into alignment with each other ("narrow RE"), as well as searching for coherence between these beliefs and relevant background theories ("wide RE") (Rawls, 1974; Daniels, 1979).[2] The resulting state, if achieved, justifies the (now revised) set of beliefs; that is, both the considered judgements about particulars, and the higher-order principles.

The distinction between the state and the process needs to be underscored in that there is no guarantee that the process will eventuate in the state of optimal coherence. Additionally, the process of RE may not (and perhaps typically does not) manifest itself as an explicit methodology. What the process describes is a rational reconstruction of what enquiry meeting the norms of RE should look like, rather than an investigative recipe or method (Beisbart and Brun, 2024, p. 79 n2; Elgin, 2017, p. 64). This point will be important to bear in mind when we come to look at case studies demonstrating RE in action in the domain of aesthetics.

Following Elgin, Beisbart and Brun argue that, of the various epistemic goods, RE aims at understanding, rather than knowledge or truth. They also note that RE is not merely an exercise in seeking coherence and "balance" among claims, as the label

[2] Daniels gives the following description of wide reflective equilibrium: "The method of reflective equilibrium consists in working back and forth among our considered judgments…about particular instances or cases, the principles or rules that we believe govern them, and the theoretical considerations that we believe bear on accepting these considered judgments, principles, or rules, revising any of these elements wherever necessary in order to achieve an acceptable coherence among them" (2016, p. 1).

may imply; the norms of RE additionally require that the "commitments" (beliefs) which an inquirer begins with, and strives to balance and adjust, must also have independent credibility—in other words, we must have prima facie reason to take them seriously, independent of how well they might cohere with the other commitments we seek to bring into RE. Moreover, in seeking RE, we aim not just for coherence among our beliefs, but rather *systematicity*, understood as a commitment to a range of epistemic goals and virtues beyond coherence, both general (e.g. accuracy, scope, parsimony, fruitfulness) and domain-specific (in aesthetics, possible candidates include value maximization, and satisfaction of the acquaintance principle).[3] Relatedly, Norman Daniels argues that the coherence sought within RE is not mere logical consistency, but a stronger condition of mutual support among commitments in a network of beliefs: "An acceptable coherence requires that our beliefs not only be consistent with each other (a weak requirement), but that some of these beliefs provide support or provide a best explanation for others" (2016, p. 3; see also pp. 26–7). Putting this observation together with those from Beisbart and Brun, we might call the goal of RE *systematic coherence*. Given this revised description of the goal of RE, then, considered as a theory of epistemic justification RE is better described as a form of "weak foundationalism" rather than as a species of coherentism (Beisbart and Brun, 2024, pp. 78–9). From the initial quotes and characterisations above, it is easy to understand why RE is often taken to be a form of coherentism, but it is both more complex and more robust.

RE is, then, a normative enterprise. But I opened this essay with a pair of questions—one descriptive, bearing on the relevance of RE to aesthetic appreciation as it is actually conducted, and one normative, on the relevance of RE to aesthetic appreciation as it ought ideally be conducted; on its relevance, that is, to a normative theory of appreciation. How do these questions relate to one another, and which has priority? While ultimately my goal is to defend a positive answer to the second, normative question, the descriptive question is both of independent interest, and—on the assumption that our practices constitute an imperfect realisation, and to that extent evidence, of our norms—part of the route to an answer to the normative question. By examining our actual practices (of moral deliberation, scientific enquiry, art criticism, and so forth), we can discover the norms explicitly or implicitly in place in these practices, and consider whether revision or conceptual engineering of the framework of norms is warranted.

So the simple distinction between descriptive and normative questions that I opened with needs to be refined to recognise that part of the enquiry is "descriptive-normative"—it involves unearthing the norms that, as a matter of fact, are embedded in and underpin our existing practices, making explicit what is generally tacit. Here we are engaged in describing and explaining an existing normative framework, which acts as a (perhaps necessary) precursor for revisionist normative enquiry, where we hold up the existing framework we've discovered and consider whether it is robust from a normative point of view. Here we engage in "normative-normative" enquiry.

[3] On value maximization, see Davies, 2006, pp. 122-5; on the acquaintance principle, Wollheim, 1980. A third domain-specific virtue in aethetics might be *perceptual viability*, a topic to which I return in 'Hard Cases, Narrow, and Wide Reflective Equilibrium' below.

So much for an initial sketch of RE; what about aesthetics? We might begin by considering the practice and the phenomenology of artistic appreciation, which (so I will argue) is a process of deliberation in which various considerations concerning the form, meaning, and value of works of art are weighed and assessed in relation to one another.[4] What kind of work is this? In what genre or tradition does it fit? When and where was it made? Who was it made for—children, teenagers, adults? A local audience, or an international one? How do its parts add up to a whole—how does the end relate to the beginning, this character to that character? Does it achieve what it seems to have been designed to achieve? Engaging with an artwork (film, novel, painting, song, and so on) inevitably involves the search for answers to a host of questions of this type. And as Susan Sontag suggests in the passage acting as the epigraph for this essay, answering these questions involves, among other things, an interplay between "specific judgements" about particular works and more general "theoretical assumptions"—precisely the kind of search for holistic, interlevel coherence that Rawls and Elgin seek to identify and theorise under the rubric of reflective equilibrium.[5] This connection should hardly surprise us. For when we engage in an act of appreciation, whether in the context of written criticism or informal conversation, we are not just whistling in the dark, but seeking to shed light on the objects of our attention.[6] "Confronted with an alien dance form, we may initially be bewildered", Elgin notes. "We venture hypotheses and test them to discover whether they make sense of what we are seeing" (2017, p. 218). To the extent that aesthetic appreciation and art criticism aim at increasing our understanding of artworks, then reflective equilibrium—as a theory of epistemic justification—will indeed have relevance for these practices.

Appreciating a work of art is not like making a mathematical calculation, even a complex one. Rather, it is precisely a process of continuous refinement and adjustment, in which our assumptions, inferences, and expectations are iteratively revised, at least up to the point where our understanding and evaluation of the work settles into a relatively stable state. "[A]n understanding of a topic consists in accepting a system of commitments in reflective equilibrium", Elgin argues; and "[a] network of commitments is in reflective equilibrium when each of its elements is reasonable in light of the others and the network as a whole is as reasonable as any available alternative in light of our relevant previous commitments" (Elgin, 2017, pp. 3–4). If we think of an artwork as a very specific topic, then everything in Elgin's description of reflective equilibrium above holds, in terms of what will count as a good—well-justified—understanding of the artwork. And it is important to underline here that

[4] The relationship of the artistic and the aesthetic is the subject of extensive debate in aesthetics (see Gaut, 2007, Chap. 2, and Lopes, 2014, Chap. 5, for discussion). I adopt one standard position in this essay, treating the aesthetic and the artistic as extensively overlapping, if not identical, and adhering to Gaut's "artistic theory of the aesthetic, which holds that the aesthetic properties of artworks are their evaluative properties that give them their value qua artworks" (Gaut, 2007, Oxford Scholarship Online, abstract for Chap. 2). For the most part, as will become clear, my interest is in the bearing of reflective equilibrium on our appreciation of artworks, but by hypothesis my arguments will extend to our appreciation of other objects of aesthetic attention (non-artistic artefacts, natural phenomena, and so forth).

[5] The passage appears in the 'Note to the Paperback Edition' of Sontag's first collection, *Against Interpretation and Other Essays* (1966/1969).

[6] In a later section, I will consider the anti-cognitivist objection that neither art nor its appreciation is correctly characterised in epistemic terms.

the objects of RE—which is to say, the objects of our justified understanding—might well be as specific as an individual work of art, as Daniels notes: "The inquiry might be as specific as the moral question, "What is the right thing to do in this case?" or the logical question, "Is this the correct inference to make?" Alternatively, the inquiry might be much more general, asking which theory or account of justice or right action we should accept, or which principles of inductive reasoning we should use" (Daniels, 2016, p. 1).

The state of equilibrium we reach in appreciating an artwork may or may not coincide with the moment when we complete our initial engagement with that work (finish our first reading of the novel or first viewing of the movie, for example). But it will, in any case, be only a provisional state, since reflective equilibrium is not only holistic, but dynamic. "[T]his equilibrium is not necessarily stable", Rawls notes. "It is liable to be upset by further examination…" (1971/1999, p. 18; see also Elgin, 2022). The resting "state" that we reach—whenever we reach it—is thus never entirely static, but always subject to small or large adjustments. From this perspective too, then, we have good reason to think that aesthetics is a domain in which reflective equilibrium is pertinent, for our judgements about artworks are subject to revision in just this way. We discover something new about some aspect of a work—either through renewed and careful attention to the work, or by acquiring new information about its history or context—and this sets us thinking afresh as to whether we have understood the work optimally.

While there is an abundant literature on RE and another on aesthetics, as I've already noted, there is precious little on the intersection of the two. But it is worth noting that the godfather of RE, Rawls himself, does have some things to say about the aesthetic dimension. Though it is not much remarked upon, Rawls touches upon aesthetic matters and expresses a view about their place in human flourishing in *A Theory of Justice*. Rawls characterises what he terms "the Aristotelian Principle" in the following way:

> …other things being equal, human beings enjoy the exercise of their realized capacities (their innate or trained abilities), and this enjoyment increases the more the capacity is realized, or the greater its complexity. The intuitive idea here is that human beings take more pleasure in doing something as they become more proficient at it, and of two activities they do equally well, they prefer the one calling on a larger repertoire of more subtle and intricate discriminations. (1971/1999, p. 374)

In support of this idea, Rawls draws on the contrast between checkers (draughts) and chess. The two games use the same board, but chess involves a greater array of differentiated pieces, each with distinct powers of movement and capture, and more complex rules (and so it is characterised by "more subtle and intricate discriminations"). Through this example Rawls invokes the domain of play—adjacent if not identical with the domain of art.

A few pages later, Rawls picks up the theme:

> Human beings enjoy the greater variety of experience, they take pleasure in the novelty and surprises and the occasions for ingenuity and invention that such [developed and complex] activities provide. The multiplicity of spontaneous activities is an expression of the delight that we take in imagination and creative fantasy. Thus the Aristotelian Principle characterises human beings as importantly moved not only by the pressure of bodily needs, but also by the desire to do things enjoyed simply for their own sakes, at least when the urgent and pressing wants are satisfied. (1971/1999, pp. 378-9)

Here Rawls draws closer still to the aesthetic domain, as understood according to a broadly Kantian conception—to attend to an object or activity aesthetically is to do so "for its own sake", disinterestedly, rather than instrumentally, for the sake of some practical goal or interest. To attend to the sun setting aesthetically is to appreciate its qualities (the hues and intensity of light comprising it) as an end in itself, rather than, for example, as a means of assessing how much longer there will be sufficient daylight to drive home on the unlit cliff road. Similarly, if more elaborately, to attend to *Moonlight* (Barry Jenkins, 2016) aesthetically is to appreciate its qualities (its narrative structure and kinoaudiovisual design) as an end in itself, rather than to consider how far and in what ways the film might advance the cause of gay rights in the black community, or how the film might be instrumentally useful in the classroom as a prompt for debate on black masculinity. There is an important difference, of course, between the sunset and *Moonlight* insofar as the film possesses moral and political qualities in a way that the sunset doesn't. Nonetheless, on this view, there is still a viable and important contrast between a direct, instrumental engagement with moral and political issues, and the more indirect engagement with them in play when we attend to them aesthetically, "for their own sake"—that is, as part of the overall design of the film (Smith, 2023a).

Rawls is interested in giving us this picture of human motivation as part of what must be considered in arriving at a robust and compelling theory of justice via reflective equilibrium. That is, the facts about human nature captured by "the Aristotelian Principle" are among those that must be accommodated when we contemplate what makes for a just society. So for Rawls, the road to a theory of justice necessarily takes us through the territories of play and the aesthetic, of beauty and the imagination. In this sense we can say that while Rawls' discussion of the aesthetic remains subordinate to his goals in moral and political philosophy, he nonetheless registers its importance to human flourishing. Having brought the aesthetic into view in his exploration of human nature, no matter how briefly, Rawls opens the door to a consideration of the relevance of reflective equilibrium to the aesthetic domain as such. And that is the door through which the present essay walks.

2 Reflective equilibrium in the appreciation of a work of art

"[Reflective equilibrium] requires", writes Alison Hills, "that we start with some moral judgements, of any level of generality (that is, some may be general principles, others about specific cases). Then we formulate principles that 'account for' these specific cases, altering the principles or modifying our judgements of particular cases until there is a good fit between all our moral judgements. This outcome is the state of reflective equilibrium, and the process is the method of reflective equilibrium" (Hills, 2021, p. 1402). In the context of our appreciation of artworks, is there an equivalent to this interlevel interplay between general principles and particular cases, or intralevel comparison between principles, motivated by the search for systematic coherence among these elements characteristic of reflective equilibrium? Here I will explore the hypothesis that an important dimension of our engagement with artworks is indeed characterised by intra- and interlevel interplay of this type, which lies at the core of RE as a process. I begin with the case of narrow RE here—clearly the target of Hills' description above—before turning to wide RE in the next section.

Appreciating a work—that is, coming to a reasoned understanding and appraisal of it—is in part a matter of situating it against a range of background schemas, concerning relevant artistic genres, traditions, and categories. While individual works may conform to established artistic norms to different degrees, no work is entirely *sui generis*; even the most avant-garde works invoke certain norms by (often ostentatiously) violating or disdaining them. On the other hand, every work is unique and distinctive to some degree, and to that extent cannot be exhaustively understood by reference to a genre or category of which the work is held to be a member. For this reason, the mere application of a principle—generic or categorical schema—is insufficient for appreciation. Insofar as arriving at an aesthetic judgement of a work involves this to-and-fro between attention to the particular details of the work and the artistic categories that it (implicitly or explicitly) evokes—a process of mutual adjustment between the assumptions embedded in these categories, and the (minor or major) departures from those assumptions articulated by the work—there are good reasons to think that at least a core aspect of engaging with an artwork is characterised by (a species of) reflective equilibrium.

Let's take, as a vivid recent example, Martin McDonagh's *The Banshees of Inisherin* (2022), which tells the story of longstanding friends Colm Doherty (Brendan Gleeson) and Pádraic Súilleabháin (Colin Farrell), a fiddler and a farmer respectively. The film explores Colm's abrupt and somewhat opaque decision to bring the friendship to an end, and the psychologically and eventually physically violent consequences of this action—Colm threatening to slice off one of his fingers each time Pádraic attempts to talk to him, thereby maiming himself to spite his erstwhile drinking companion. Consider the opening lines of two reviews of the film, by *The Guardian*'s Pete Bradshaw and *The New York Times*' A. O. Scott, respectively:

> Martin McDonagh's new film is a **macabre black comedy** of toxic male pride and wounded male feelings, a **shaggy-dog story** of wretchedness and a **dance of death** between aggression and self-harm, set on an imaginary island called Inisherin off the Irish coast. It's happening in 1923 during the civil war; the

> additional symbolic acrimony is offered to us on a take-it-or-leave-it basis. / As with so many of **McDonagh's works**, the glint of the unburied hatchet is all too visible in the murk, and the setting is a stylised and ironised Irish rural scene **not so very far from John Millington Synge**. Mutilation is a familiar motif. There are plenty of genuine laughs in this movie, but each of them seems to dovetail into a banshee-wail of pain. (Bradshaw, 2022)

> The island of Inisherin, a rustic windswept rock off the coast of Ireland, does not appear on any real-world maps, but its geography is unmistakable. Not only because the sweaters and the sheep, the pints of Guinness and the thatched roofs bespeak a carefully curated Irish authenticity, but also because what happens on this island locates it firmly in **an imaginary region that might be called County McDonagh**. / This is a place, governed by the playful and perverse sensibility of the **dramatist and filmmaker Martin McDonagh**, where the picturesque and the profane intermingle, where jaunty humor keeps company with gruesome violence. (Scott, 2022)

Both reviews position the film squarely within the category "Martin McDonagh films", each offering sketches of that category through salient features of the film which connect it with the category. Bradshaw's *Guardian* review also situates *Banshees* within a range of other artistic categories of different levels of specificity, suggesting how the "McDonaghesque" weaves together aspects of these various older and broader traditions; Bradshaw also notes a particular aesthetic kinship with the work of Irish playwright J. M. Synge. The picture this suggests is that any individual work can be positioned within a criss-crossing network of relations among different categories of varying orders of breadth. The first job of an appreciator—a critic being nothing other or more than a professional, or at least formal, appreciator—is to discern exactly where in this large, complex, and heterogenous web of relations a particular work is to be situated.

And it is here that we see the "back-and-forth" movement characteristic of reflective equilibrium—that "mutual adjustment" among our judgements and principles. Starting at the level of the principle describing a category, we ask: is this a comedy? Moving down to the level of our judgement concerning the specific work, we answer: well, yes, but a black comedy; so black, in fact, that it contains hints of horror (the film is a "dance of death", "where jaunty humor keeps company with gruesome violence").[7] But it is also characterised by a certain stylisation ("stylised and ironised"), and a knowing digressiveness and implausibility ("shaggy-dog story"). That's what makes it just this kind of thing: a work by Martin McDonagh, a director with a signature style.[8]

[7] Later in his review, Scott writes: "[*Banshees* is] a good place to start if you're new to [McDonagh's] work, and cozily—which is also to say horrifically—familiar if you're already a fan". Bradshaw also returns to the theme (as we see in the passage quoted in the main text): "There are plenty of genuine laughs in this movie, but each of them seems to dovetail into a banshee-wail of pain".

[8] Or, to be more precise, a work in the style we associate with McDonagh—for nothing in the account here rules out another director making a work in this style, or guarantees that McDonagh himself was in fact

This first task of the appreciator and critic is descriptive; it answers the intimately related questions, "what is this work?" and "what kind of work is it?" But criticism is a kind of "thick discourse" which addresses not only these descriptive questions, but the evaluative questions "how good is this work?" and "how good is it as an instance of the kind of work that it is?"—these descriptive and normative enquiries being intertwined. Here we see the normativity of RE in play: the fact that we adduce reasons for judging how valuable a given work is, in various respects and as a whole, exemplifies the normativity of RE as a theory of epistemic justification, foregrounding the nature of the reasoning involved in making epistemic claims (to understanding or knowledge) in general. The back-and-forth movement between different possible judgements, noted above in relation to the descriptive and analytic work of the critic, is also evident when critics turn to these evaluative questions. Scott concludes his review with a "verdictive judgement" (Zangwill, 2001, Chap. 1)—a final and overall judgement or verdict on the aesthetic value of a work. This is the kind of judgement that we might regard as expressive of the *state* of reflective equilibrium; the (provisional) endpoint of the *method* (or in the more circumspect terms noted in the introduction, the *process*) of reflective equilibrium. Scott writes:

> *The Banshees of Inisherin* might feel a little thin if you hold it to conventional standards of comedy or drama. It's better thought of as a piece of village gossip, given a bit of literary polish and a handsome pastoral finish. Inisherin may not be a real place, but its eccentric characters, rugged vistas and vivid local legends make it an attractive tourist destination all the same. (Scott, 2022)

In other words: you might think that this film is a really substantial comic or dramatic achievement (the "conventional standards" Scott alludes to forming the aesthetic equivalent of the moral principles we invoke in making or explaining particular moral judgements). But that would be a mistake, says Scott; it would be more apt to think of this work as a piece of highly-crafted gossip or a sophisticated tourist memento (like a good many heritage films). It's a decidedly double-edged evaluation, quietly skewering the film with faint praise; and it works dialectically, by invoking and then rejecting alternative evaluative judgements of the film (that the film is a good drama, or a good comedy, by conventional canons of judgement). So a verdictive judgement need not be, and very often will not be, "thin"—the mere assertion that a work is good, or bad, or mediocre. More typically, such judgements will be part of the "thick" fabric of appreciative discourse, describing as they evaluate.

Although I've drawn on a case study in journalistic film criticism as an example of how reflective equilibrium works in the context of aesthetic appreciation, in many ways such criticism is not the ideal form of evidence. As will be clear from my discussion of the reviews above, journalistic criticism is typically highly-compressed, so that the rather complex judgements often arrived at by critics—Scott's claim that *Banshees* doesn't fare well when judged by the conventional norms of drama and comedy, for example—are sketched rather than fully detailed, mortgaged rather than

responsible for the work. What the account does provide is prima facie evidence that *Banshees* is a film directed by McDonagh.

paid in full. The distilled nature of such criticism makes it of limited value in tracking the importance of reflective equilibrium in aesthetic appreciation in two further ways. To take the first of these points: as we've seen, reflective equilibrium can be thought of as both the method or process by which we weigh up and make adjustments between different judgements and principles, and the "final", stable state that that we aim at through this method. Journalistic criticism tends to express that state, as we saw in the concluding lines of Scott's review, with a rather limited view of the process and method that led to it.

Moreover, to turn to the second point, while reflective equilibrium can be conceived in individual or collective terms—that is, as a matter of an individual, or a group, moving through the process to attain the state of RE—it is arguably in its collective form that the back-and-forth, mutual adjustment of judgements central to reflective equilibrium is most visible. In one of the few writings in aesthetics to invoke reflective equilibrium, Noël Carroll makes the following observation about informal discussion of narrative artworks:

> As we compare and contrast what we have observed about the characters [in a narrative] with the views of others, and as we weigh our judgments against what others have to say, we are, of course, engaging in a process of reflective equilibrium… Often informal conversation and debate of this sort explicitly interrogates disputants' use of this or that virtue in describing this or that character. Is so-and-so correct in calling such-and-such a character shy rather than aloof, arrogant rather than magnanimous, reckless rather than courageous? Does the ingenue have character, properly so called, or merely personality? This then can quickly escalate into questions about what it means to call a behavior or a character truly courageous, or cowardly, or reckless. (Carroll, 2002, p. 15)

The relatively compressed, static, and individualistic nature of journalistic criticism tends to marginalise this kind of collective, dialogical reflection, making it a less than optimal form of evidence for the workings of reflective equilibrium in aesthetic appreciation.

The economical conjuring up and appraisal of a work is, of course, part of the art of the journalistic critic, so it is hardly an indictment of such criticism to note that it works in this way. Nonetheless it does suggest that longer and "slower" forms of criticism, and dialectical exchanges between critics, are more apt to bring out the significance of reflective equilibrium for appreciation. And in the context of journalistic reviewing, high-profile, controversial cases are more likely to generate such dialogue. In his review of *Oppenheimer* (2023), for example, *LA Times* critic Justin Chang takes other reviewers to task for some of the negative commentary on the film, charging them with misunderstanding the film and failing to appreciate the rigour of its focus (Chang, 2023). He thereby challenges them to adjust their beliefs—both descriptive and evaluative—and revise their judgement of the film. To the extent that this intervention forms part of a dialogue among critics and appreciators aiming to arrive at an optimal—most illuminating and best justified—understanding of the film, here we see RE in action.

As will be evident, I am treating understanding the work of art *as a whole* as the appropriate target of reflective equilibrium, and critical appreciation as the arena in which the method or process is staged. It is also possible to "apply" reflective equilibrium to judgements about *aspects* of works rather than works as wholes. One might attempt to assess a work in terms of its *moral and political implications*, with little or no regard for its form or style—that is, for its aesthetic features—treating it simply as a vehicle for moral and political content (as if it were no different to a political manifesto or a treatise on ethics). And appreciators do weigh up—using reflective equilibrium—the deeds and motivations of *characters*, as if they were independent entities (as Carroll suggests, in the second part of the quote above).[9] Scott wonders whether we can fathom, and justify, Colm's actions: "it's hard for Padraic or the audience to accept what's happening, let alone understand it". Is Colm on morally firm ground in abruptly breaking off his friendship with Pádraic? "Colm simply decides he doesn't want to be friends with Padraic any more", Bradshaw writes. Scott continues: "'I just don't like you no more', [Colm] tells Padraic, who responds with wounded incredulity". Is Colm justified in his actions of self-mutilation as a protest against Pádraic's incomprehension, and refusal to be dropped as a friend? The other characters weigh in, of course; of Colm, Dominic asks, "What is he, 12?" (Scott, 2022; Bradshaw, 2022). But we appreciators are the final arbiters in this process of reflective assessment, and our judgements of the actions and motivations of characters, or about the morality and politics of works, are ultimately made—or at least, should be made—in the context of our appraisal of works as *art*works in their entirety, that is, as wholes.

Aside from its intrinsic interest, the argument I advance here may be of additional significance to the extent that it has implications for recognised theories, problems, and debates. One such case concerns Frank Sibley's influential arguments on the nature of aesthetic concepts and their attribution, in particular the thesis that the possession of an aesthetic property by an object or event cannot be inferred from any set of (logically or causally) necessary and sufficient conditions—in other words, that we cannot infer that a given aesthetic property will be present through deduction or induction.[10] Does the thesis advanced in the present paper that aesthetic appreciation is a "thick" business, involving an intricate (inter- and intralevel) and iterative movement between descriptive and evaluative claims about an artwork, framed by an effort after reflective equilibrium, comport with or shed light on Sibley's theory?

Sibley seems to write within a broadly naturalistic vein (this is particularly clear in the final pages of 1959/2001, pp. 20–3). So how one might reconstruct his view about aesthetic properties in naturalistic terms? Aesthetic properties (for example, "graceful") are complex, high-order properties which emerge from the interaction of more basic properties (for example, "curved" and "regular"), Sibley suggests; but these interactions are of such complexity that it is (practically) impossible to predict their emergence from the presence of simpler, lower-order properties. This is the sense in which aesthetic properties are not "condition-governed", in spite of their robust con-

[9] I discuss the ontology of character, and the sense in which characters can come to exist independent of the narratives in which they first appear, in Smith 2023b.

[10] Sibley (1959/2002) puts the emphasis throughout on logical conditions, but he also considers and rules out "inductive procedures" (pp. 9–10); he rejects deductive and inductive inference together on p. 15.

nection with non-aesthetic features: "Aesthetic concepts, all of them, carry with them attachments and in one way or another are tethered to or parasitic upon non-aesthetic features" (p. 17). Additionally, detecting aesthetic properties requires some expertise ("taste"), and so in many cases will be subject to disagreement; and that disagreement may be hard or (practically) impossible to settle, because of the complex, emergent nature of aesthetic properties, and the expertise their discernment requires. But there is nonetheless a fact of the matter about whether a given object or phenomenon possesses a certain aesthetic property.[11]

How might reflective equilibrium fit in here? To the extent that Sibleyean appreciation (the discovery of aesthetic properties) doesn't rely on either straightforward deductive or inductive logic, but rather on a holistic "back and forth" akin to that described by Goodman and Rawls, there is indeed an important kinship. Note in this connection that Sibley's descriptions of critical appreciation typically also involve a more localised back-and-forth movement—between grasping the details of the artwork under discussion, and understanding the overall work; between fathoming parts of the work and apprehending the whole. This seems very much like reflective equilibrium in operation at a smaller scale: we test our understanding of the whole by looking at its parts, and reciprocally we test our understanding of the parts by seeing if those construals make sense in light of the whole. Putting these points together, I suggest that RE provides a solution to the problem identified by Sibley: we may not be able to infer aesthetic properties from (logically or causally) necessary and sufficient conditions, but rather, in order to discover and attribute such properties, we must engage—and we do engage—in the back-and-forth, dynamic and holistic, movement within and between levels described by RE.

Note, however, that Sibley's distinction between non-aesthetic and aesthetic attributions doesn't quite line up with the distinction between descriptive and evaluative judgements in the present essay. In one paper, for example, Sibley writes that he is "concerned with aesthetic descriptions. I have in mind assertions and disagreements about whether, for instance, an artwork (or where appropriate, a person or thing) is graceful or dainty, moving or plaintive, balanced or lacking in unity. I deliberately ignore, by a partly artificial distinction, questions about evaluation, though many assertions of the sorts I discuss are relevant to whether a work has merits or defects" (1968/2001, p. 71; see also 1959/2001, p. 5 n3 and p. 12 n7). So on the one hand, Sibley holds that the descriptive and the evaluative dimensions of appreciation can be teased apart and analysed independently of one another (at least to some extent, or with respect to certain questions). The identification and attribution of aesthetic properties, and the relationship between non-aesthetic features and aesthetic properties, on which Sibley focussed, is but one aspect of aesthetic appreciation. On the other hand, Sibley's admission that his separation of aesthetic description and evaluation is "partly artificial" suggests that there is further work to be done here, and that there is some prospect of a deeper integration between the account of the role of reflec-

[11] See Levinson (2006) for a related discussion. Walton builds on but also revises the picture given by Sibley, on the one hand indicating with greater confidence how the presence of certain non-aesthetic features might determine the presence of some aesthetic properties, while on the other hand insisting that "a work's aesthetic properties depend not only on its nonaesthetic ones, but also on which of its nonaesthetic properties are "standard", which "variable", and which "contra-standard"" (1970, p. 338).

tive equilibrium in the aesthetic domain advanced here, with a Sibleyean account of aesthetics.

3 Hard cases, narrow, and wide reflective equilibrium

Whether in the form of a brief individual review, a more sustained critical investigation, or a series of exchanges among multiple critics, criticism of most works generally manifests "narrow" rather than "wide" reflective equilibrium. The conception of wide reflective equilibrium (in its individual mode) has been concisely described by Daniels as "an attempt to produce coherence in an ordered triple of sets of beliefs held by a particular person, namely, (a) a set of considered moral judgments, (b) a set of moral principles, and (c) a set of relevant [scientific and philosophical] background theories" (Daniels, 1979, p. 258). He elaborates on the contrast between narrow and wide reflective equilibrium in the following way:

> To the extent that we focus solely on particular cases and a group of principles that apply to them, and to the extent that we are not subjecting the views we encounter to extensive criticism from alternative moral perspectives, we are seeking only narrow reflective equilibrium. Presumably, the principles we arrive at in narrow equilibrium best "account for" the cases examined. Others, however, may arrive at different narrow reflective equilibria, containing different principles and judgments about justice. Indeed, one such narrow equilibrium might be characterized as typically utilitarian, while another is, we may suppose, Kantian or perhaps Libertarian. (Daniels, 2016, p. 11)

Thus the particular case of *Banshees* is explored in relation to a group of principles (concerning comedy, drama, horror, black comedy, shaggy dog stories, and so on) that might apply to it. Although the appraisal our two critics arrive at is at least subtlety different—Scott implying that the film is ultimately less substantial than it might appear to be—they are to a large extent in agreement with one another; to that extent they arrive at the same or very similar equilibria.

Neither Bradshaw nor Scott, however, pushes beyond the set of principles, related to the various artistic categories enumerated above, according to which the film is best understood. Neither of them sets the film against the backdrop of more abstract theories according to which both certain artistic categories and individual works will count as aesthetically valuable—be those theories drawn from Hume, Kant, Hegel, Adorno, Sontag, Danto, Argle, or Bargle. "Art-critical evaluative practices are subject to reflective equilibrium and are ultimately responsive to the value of art qua art", Berys Gaut argues (2007, p. 40); and different theories of art cash that value out differently (see footnote 4). It is at this level that we reach the equivalent to Daniels' utilitarian, Kantian, and Libertarian conceptions of moral and political life, and the prospect that each would generate distinct narrow reflective equilibria in considering particular cases. In an exercise of wide RE, each of them would be considered and compared against relevant background theories. But we don't typically see justification pushed to this level of abstraction in everyday appreciative discourse, including

journalistic criticism, any more than we see its moral equivalent in everyday moral deliberation.

Are there places where we do encounter something akin to wide reflective equilibrium in aesthetic discourse? There are indeed: we find such reflection in philosophically self-conscious criticism, particularly in relation to "hard", contentious cases. But consider first a "standard" or "soft" case, where there is little or no doubt about the status of the work in question as an artwork. Writing in *The Journal of Aesthetics and Art Criticism*, James Harold offers the following comment on Jonathan Gilmore's discussion (in his 2020) of Dickens' *Oliver Twist* (1838):

> The worry…is that our aesthetic reasons for thinking that Dickens' description of Fagin's greasy clothing do not warrant our moral disgust might arise from more general norms about how to emotionally engage with artworks. Why should we agree with Gilmore that the best way to experience fiction is to be manipulated by it? (Harold, 2021, p. 274)

A particular case is at stake—Dickens' novel, and more specifically what we are to make of the character Fagin. By contrast with the reviews of *Banshees*, however, Harold directs our attention to a much more abstract background question, concerning "general norms about how to emotionally engage with artworks". It's not an accident that we find such a remark in a publication called *The Journal of Aesthetics and Art Criticism* rather than one called *The New York Times*. The clue is in the title of the journal, which tells us that its purview is both art criticism and philosophical aesthetics. As in the case of morality, we are much more likely to encounter this third tier of reflection, concerned with background theories and the way these theories justify the principles and the judgements of the first two tiers—more abstract than the first tier bearing on considered judgements, and the second tier picking out principles of judgement—in the context of philosophy.

What about "hard cases"? A "hard case" is a work of art "whose status as art is controversial from a theoretical perspective" (Lopes, 2014, p. 6). Paradigm cases would include Marcel Duchamp's readymades, most famously his *Fountain* (1917) (just an inverted urinal relocated to an art gallery?); Andy Warhol's *Brillo Box (Soap Pads)* (1964) (no different to a box of Brillo pads one might encounter at the supermarket?);[12] and John Cage's *4'33"* (1952) ("I came here to listen to some music, not to endure an awkward silence"). Another such case from the world of music is twelve-tone serialism, the compositional tradition initiated by Josef Matthias Hauer and (more influentially) Arnold Schoenberg which systematically confounded the principles of tonal music, replacing those principles with a complex set of alternative norms concerning the selection and ordering of notes. Unsurprisingly, serialism was (and to the extent that it still exists as a living tradition, remains) a controversial if influential practice, attracting a small audience of partisans and aesthetic adventurers.

[12] Warhol's Brillo boxes are actually readily distinguished from their supermarket siblings; but the carefully-crafted, ersatz beer cans featuring in Alexandre Lavet's *All the Good Times We Spent Together* (2016) were mistaken for the real thing by a lift technician working at the LAM Museum in Lisse in the Netherlands, and tossed in the garbage.

The philosopher Diana Raffman kicked up some dust when she described serialism as artistically "defective", on the grounds that "twelve-tone pitch structure is not perceptually real" (Raffman, 2003, p. 86). Serialism is "fraudulent", Raffman argues, because humans don't possess the perceptual or cognitive architecture necessary to perceive—to experience in listening—the harmonic structures created by these works; they thus cannot achieve their primary ends. Note that Raffman does not deny that these structures can be discovered and understood through analysis. Her point is that dodecaphonic harmony can't be discovered or appreciated in the act of listening, in contrast to tonal (and other non-Western) harmonic systems.

Once again, it's not a coincidence that we find Raffman making this argument in a lengthy, theoretical piece in *Midwest Studies in Philosophy*, rather than in a short-form piece of criticism. Note too that Raffman's argument makes a descriptive claim about an aspect of human nature—the fact that there are limits to the sonic structural complexity that humans are able to perceive and cognize—which then grounds a normative aesthetic principle, and a particular aesthetic judgement:

theory: humans cannot perceive twelve-tone harmonic structure; such structure is "not perceptually real" (p. 72) and "aurally inaccessible" (p. 86) to human beings.
principle: works of music should be *perceptually viable*, that is they should respect what is perceptually possible for appreciators; "a compositional technique can be justified only by its audible results, that is, by how the resulting music sounds" (p. 73).
judgement: dodecaphony fails to meet this norm, thus any work adopting this system of harmony will be "defective" (passim).

Raffman's method, then, exemplifies wide reflective equilibrium in that it seeks to explore (i) the justificatory grounds for (ii) the principles to which we appeal in appreciating (iii) individual works (or, as in this case, a genre, class, or practice—recall Daniels' point that the objects of RE range from the very specific to the much more general). Here we are again in the territory limned by Sontag, exposing and clarifying "the theoretical assumptions underlying specific judgments and tastes". Moreover, just as Rawls considers what he takes to be facts about human nature ("human beings enjoy the exercise of their realized capacities") in following the method of reflective equilibrium in his pursuit of a theory of justice, so does Raffman ("in virtue of human psychological design, a composer cannot intend to communicate pitch-related musical meaning by writing twelve-tone music", 86) in her pursuit of a theory of the appreciation of music. In other words, just as in the moral and political case, wide reflective equilibrium in the context of aesthetics takes into account "relevant background theories" and "relevant moral and non-moral beliefs (including general social theory)" (Daniels, 1979, p. 258, and 2016, p. 13) in its search for the most coherent and systematic—and in that sense, best justified—stance towards our principles and judgements.

Reviewing our two case studies, focussing respectively on narrow and wide RE in the context of aesthetics, we see the familiar shape of RE in the critical appreciation of *The Banshees* and serialism: the movement back-and-forth between "considered judgements" regarding particular items and higher-order judgements regarding prin-

ciples and background theories, aiming at a settled (though revisable) state in which (to revisit a key passage from Elgin) "each of [the elements in a network of commitments] is reasonable in light of the others and the network as a whole is as reasonable as any available alternative in light of our relevant previous commitments" (Elgin, 2017, pp. 3–4). By the same token, though, we shouldn't expect RE to look identical in surface appearance across different domains, given the broad scope of RE and the inevitable contrasts in language, concepts, and argumentative styles across (for example) logic, moral and political philosophy, and aesthetics. Nor should we expect that the presence of RE will be immediately discernible, prior to the kind of analysis undertaken here to reconstruct and reveal it. What we should expect to see, as a mark of the presence of RE, and what we do see in the exemplary cases examined here, is that search for systematic coherence across judgements at different levels that lies at the core of RE.

4 Objections, refutations, conjectures

We can now address two objections which might be raised in relation to what I've argued. The first of these objections challenges the very idea that artistic appreciation counts as a kind of reasoning, contending that such a characterisation overly intellectualizes our engagement with artworks and other aesthetic objects. The second objection protests against the invocation of a single, unified category of "art"—or worse still, a unified field of aesthetic experience, encompassing art—capable of explaining our engagement with everything from sunsets to symphonies, sonnets to suspense films. The two objections thus share a suspicion that the account being developed is overly reductive, and mischaracterizes at least certain kinds of art. Let me take each objection in turn.

In "Against interpretation", Sontag famously argued that the interpretation of artworks was a kind of "revenge" against their powerful sensory and aesthetic character (1966/1969, p. 17). In expressing this view, Sontag placed herself in the tradition reaching back to Plato which conceives of art as a phenomenon opposed to—or escaping, subverting, or in some more elusive way acting as the "other" to—reason.[13] Interpretation functions to contain and master the sensory potency and sometimes explosive, affective power of art, converting its bodily charge into the conceptual currency of the mind. Sontag is surely right to object to the overweening character of the contemporary institution of interpretation; but her account does not demonstrate that artworks engage us only in sensory and emotional terms, divorced from any kind of reasoning. Of course it is true that we must begin by looking at the painting, listening to the song, or imagining the contents of the novel we read, and experiencing the perceptions, imaginings, and emotions to which these give rise, but our minds do not—and should not—stop at this point.

[13] There is an irony here in that, in the setup of her essay, Sontag treats Plato as one of the architects of the mimetic theory of art, going on to argue that it is that theory that creates the mandate and the demand that works of art be interpreted. Sontag in effect asks us to celebrate and savour what Plato regards as a problem in need of a solution. The irony, though, does not eliminate the connection.

Instead, our minds actively seek to understand the work—and to anticipate how it might develop—by pursuing answers to the kinds of questions I note above, both internal (how do the parts of the work relate to one another?) and external (in what genre or category does this work fit?). The fundamental problem with the kind of irrational (or at least arational) conception of art given by Sontag is that it is rooted in an implausibly dualistic picture of human agency, which sees body and mind, art and philosophy, and reason and emotion as necessarily pitted against one another rather than thoroughly entwined. To argue, as I have, that aesthetic appreciation involves a kind of reasoning, and that that reasoning in turn involves the pursuit of reflective equilibrium, is in no way to deny or erase the richly sensory, imaginative, and emotional character of (most) art. It is rather to argue that we bring to bear our powers of reasoning on and within our aesthetic experiences, and that there is no *special* tension between aesthetic experience and reasoning. Our reasoning is certainly subject to all manner of pushes, pulls, and failures in the context of aesthetic appreciation, but no more so than when we exercise our powers of reasoning on our experiences as human subjects outside of the aesthetic domain.

A more moderate version of this first objection might be taken to be the "anti-cognitivism" defended by Peter Lamarque and Stein Haugom Olsen (1994; see also Lamarque, 2009, Chap. 6), who hold that the value of literature (and art more generally) does not obtain in the knowledge about the world afforded by artworks (or in any related epistemic good, such as understanding or insight). On this view, as Jerome Stolnitz (1992) puts it, works of art are "cognitively trivial". But this is really a very different kind of claim to Sontag's, and a claim that is in fact quite compatible with the idea that works of art typically engender, and should engender, reasoning activities of the type described by the concept of reflective equilibrium. The key here is to underline that "cognitive" in this context refers to "knowledge" rather than reasoning per se. Anti-cognitivists hold that works of art do not afford knowledge about the world beyond the work, and thus do not invite us to reason about what we may learn from them about the world, but there is no denial that they engage our reasoning powers in a host of other ways—most importantly in relation to how we understand, assess, and appreciate them on their own (non-cognitive, on this theory) terms. In other words, even for the anti-cognitivist, criticism is still an epistemic activity, insofar as its goal is to improve our understanding of artworks (and where, on this view, part of the improved understanding involves recognizing that we do not value such works primarily in terms of their ability to create knowledge about the world). Thus an anti-cognitive stance towards the value of artworks in no way precludes the thesis that appreciating artworks typically involves reflective equilibrium, such activity being sanctioned by the norms of the artworld.

What of the second objection—that it may not be plausible to make a single argument pertaining to the putative role of reflective equilibrium in the aesthetic domain at the level of "art" or "the aesthetic" as a whole? One line of argument in contemporary aesthetics has indeed pressed the significance of the differences among media, artforms, and the varied landscape of "appreciative practices" (Kivy, 1997; Lopes 2014). And there is certainly scope to accommodate this point within the account under development in the present essay. There can be no doubt regarding the significance of these differences—indeed it is part of the picture of our initial response to a

work of art, given above, that among our first reactions will be an attempt to situate the work in the appropriate artistic category. What *kind* of a painting or film is this—from where in the world does it hail? Who was it made for and what conventions does it observe?

But determining what kind of work we are in the presence of—what category or categories it fit(s) into—generally involves more than meets the eye. Perception alone is insufficient; further reasoning and inference are necessary for proper understanding. Gus Van Sant's remake of *Psycho* (1998) looks like a feature film, but perhaps it makes more sense—is more fully understood and appreciated—as a work of conceptual art; Picasso's painting *Guernica* (1937) looks just like one of the *guernicas* in Kendall Walton's extensive collection, but precisely because one is a painting and the other a *guernica*, the aesthetic properties of the two works will be rather different.[14] Picasso's work "seems violent, dynamic, vital, disturbing to us", perceiving and appreciating it as a painting. "But I imagine it would strike [those seeing it as a *guernica*] as cold, stark, lifeless, or serene and restful, or perhaps bland, dull, boring—but in any case not violent, dynamic, and vital" (Walton, 1970, p. 347). Among other things, then, Walton's thought experiment shows that our perception of artworks is (to some extent, at least) cognitively penetrated.

Thus it doesn't follow from recognizing the importance of variation in the norms among different types of artwork that the kind of reasoning examined here, marked by the search for reflective equilibrium, does not apply generally across artistic types and traditions. On one view, because of its dependence on language rather than depiction, literature will be somewhat closer to the space of reasons, and thus of reflective equilibrium, than will painting. Perhaps it is sufficient to appreciate a Rothko canvas to savour its powerful sensory and aesthetic properties, and accept that—as Sontag would have it—to seek to interpret the work is to diminish it. Perhaps we should heed her call for an "erotics of art", displacing the "hermeneutics" with which we are all so familiar (1966/1969, p. 23). But just because a Rothko painting, or a film by Stan Brakhage, engages us exclusively in visual terms—Brakhage famously eschewed all sound in almost all of his films—and moreover largely in terms of abstract rather than depictive forms, does not mean that we surrender to a narrowly perceptual experience void of all cognition.

Nor should we. Instead our minds will—entirely appropriately—wonder about the design of these works (in terms of composition, colour, texture, editing rhythm, and so on), the immediate context of their display and wider context of their history—including the intentions of the artist—and what items in our stock of previous appreciative experiences might be relevant to our appreciation of them. Rothko himself certainly thought that it was worth arguing about what sense should be made of his

[14] Walton imagines "a society which does not have an established medium of painting, but does produce a kind of work of art called *guernicas*. *Guernicas* are like versions of Picasso's "Guernica" done in various bas-relief dimensions. All of them are surfaces with the colors and shapes of Picasso's "Guernica", but the surfaces are molded to protrude from the wall like relief maps of different kinds of terrain. Some *guernicas* have rolling surfaces, others are sharp and jagged, still others contain several relatively flat planes at various angles to each other, and so forth. Picasso's "Guernica" would be counted as a *guernica* in this society—a perfectly flat one—rather than as a painting" (1970, p. 347).

paintings, and how they should be experienced.[15] Whatever else we do, we will—and should—move "back and forth" between judgements about what it is that this particular work affords us perceptually, and the broader principles we invoke, regarding relevant artistic categories, in order to make sense of it (adjusting both as needed, as in the analysis of *Banshees* above). The point holds for other forms of abstract, non-representational art, such as "absolute" music: even where a musical composition represents nothing, we will still seek to understand how its parts fit together as a whole, and how it fits into one or more musical category or tradition, by working our way through a succession of hypotheses about what it is that we are listening to.

5 Appreciation, judgement, and embodiment

Two more clarifications regarding the "space of aesthetic reasons" (Currie, 2004, p. 243) are in order. To take the first of these: when we think about the bearing of reflective equilibrium on the aesthetic domain, a pair of questions arises, each related to a distinct dimension of the process of aesthetic appreciation with a distinct epistemic focus. The first concerns our understanding and appreciation *of artworks themselves*; the second, our understanding and appreciation *of the world* that we derive from artworks (that is, the world beyond the artwork itself, which is of course part of the world, but a very specific, tiny sliver of it). It is tempting to think of this dyad in temporal terms, as a contrast between our immediate, direct engagement with a work, and the longer "reflective afterlife" (Kivy, 1997; Carroll, 2002)[16] of a work in the wake of our reading, viewing, or listening to it. But this temporal contrast is at best only roughly indicative of the distinction I am making, and may be actively misleading with respect to it.

On the one hand, the task of understanding *a work* may extend well beyond the duration of our direct engagement with it, continuing well into the afterlife of that initial engagement. It can take days, weeks, months—even years or decades—for our understanding of a work to crystallise; and as I stress above, even an apparently settled "take" on a work can be disturbed and revised. On the other hand, attention to *what a work implies about the world (beyond the work)* may begin while we are engaged with it, rather than being deferred to and contained within later reflection on it, away from the heat of battle. So the temporal distinction doesn't line up with the epistemic one. In the present essay I have only been concerned with the first of these questions, concerning the potential role of reflective equilibrium in our understanding and appreciation of artworks in and of themselves. The second question, concerning the role of RE in relation to what we learn from artworks and how this relates to our knowledge and understanding of the world more generally, must wait for another

[15] For an overview, see Lewis, 1998. Note too Rothko's efforts to create the ideal context for the appreciation of his canvases through the construction of the Rothko Chapel in Houston. Through this example, we can see that an artist's efforts to shape decisively—if not fully determine—the appreciator's experience and understanding extends from the work in the narrowest, most concrete sense (the canvas, in the case of a painting) through the context of its exhibition (the space in which it is to be viewed) to the context of its critical discussion.

[16] The phrase "reflective afterlife" is first used by Carroll in his 2002, referring to the idea in Kivy (1997).

occasion for substantial treatment (though see my remarks on Lamarque and Olsen's anti-cognitivism above).

A second clarification: I have made reference to both "aesthetic appreciation" and "aesthetic judgement". Aesthetic judgement may seem to be more directly connected with reflective equilibrium, insofar as our "considered judgements" of particulars—to use Rawls' expression—are a crucial ingredient in the process of seeking reflective equilibrium. Our aesthetic judgements are important in the aesthetic case, but it is important to grasp that such evaluative judgements—on the positive or negative aesthetic value of works of art—are but part of the process of appreciating a work, which must begin with an effort after understanding it. As we have seen, appreciation is a "thick" affair: before we can confidently judge a work, we first have to gauge it—to figure out what it is that we are dealing with, which partly means determining what *kind* of work we are encountering. In some important and basic sense, we need to describe before we can evaluate, even if it is true that once appreciation is in train—once we are engaged with a work of art—we will usually move seamlessly between descriptive and evaluative judgements (and in this way, the back-and-forth movement between judgements of different types, characteristic of reflective equilibrium, can be seen to be already in operation). Description and appraisal are closely bound up with one another in the phenomenology of aesthetic appreciation.

In addition to pointing to the thick character of our engagement with artworks, the concept of appreciation also encompasses the perceptual and emotional dimensions of such engagement. Tragedies move us, to pity and fear; music induces chills; those Rothko canvases command our attention; and just about all artworks prompt imaginings. Artworks engage our senses and emotions, and arguably more exotic flora and fauna of the mind, such as "aliefs" (Gendler, 2010), as much as our beliefs; and it is this whole array of embodied states that we seek to bring into equilibrium as we make sense and take stock of a work. This observation takes us back to Sibley, who notes the role of expressive, bodily gestures in critical appreciation: "We accompany our talk with appropriate tones of voice, expression, nods, looks, and gestures. A critic may sometimes do more with a sweep of the arm than by talking. An appropriate gesture may make us see the violence in a painting or the character of a melodic line" (1959/2001, p. 19).

Elgin also recognises the embodied nature of art and its appreciation, noting how the non-propositional and non-argumentative form of much art nonetheless functions epistemically, by exemplifying and thereby highlighting an endless variety of properties and phenomena. "The Judson Theater dancers put us in a context where we attend to the physical intelligence that goes into such mundane activities [as walking, running, and carrying loads]. We notice and attune ourselves to the minute, intricate, muscular adjustments involved in keeping one's balance while schlepping a mattress. We notice the rise and fall, the small and large physical adjustments that it takes to walk or run across the floor. The dances thus exemplify features that mundane motion instantiates but that we, either makers or observers of that motion, routinely ignore" (2017, p. 210). By at once exemplifying and defamiliarizing such ordinary physical actions, the Judson Theater dances make the character of such actions salient and enhance our understanding of them.

But, it might be asked, is this expansion of the kind of states posited within the process of RE legitimate? The orthodox conception of reflective equilibrium is strictly doxastic, concerning itself with beliefs of various degrees of abstraction and generality, but only with beliefs. Does the process remain coherent if we open its doors to the wider network of mental states which bear upon our beliefs, but can't be reduced to them?

My positive answer to this question cuts in two directions. On the one hand, the expansion is necessitated and justified by the nature of aesthetic appreciation, which cannot be adequately captured by belief alone; perhaps this expansion is a feature of RE specific to the domain of aesthetics. It is possible to accommodate a wider array of types of state within the framework of reflective equilibrium to the extent that we are still dealing with belief-like states, that is, states which *represent* the artwork as being a certain way—as, for example, representing tragic or humorous events when we respond to a work with sadness or laughter, or bearing racist or sexist attitudes when we respond with upset or anger. On the other hand, and cutting in the other direction, it might be that reflective equilibrium as it is deployed in some other domains can and should be expanded in a parallel fashion, in line with the widespread recognition across many areas of philosophy and other disciplines of the importance of embodied cognition (Shapiro, 2019). It is not only in aesthetic contexts that our beliefs are entangled with other states which might, so to speak, be legitimate stakeholders in the process of reflective equilibrium.

Our understanding of reflective equilibrium in the context of moral and political deliberation, for example, might be illuminated by the aesthetic case, for there are some important points of contact here—our moral judgements are often accompanied by emotions, and a great many of our emotions are morally saturated. Rawls argued that judgements made under the sway of certain emotions, like fear and distress, should be excluded from RE (1971/1999, p. 42), but in the final chapters of *A Theory of Justice*, he recognizes the cognitive value of moral emotions like guilt, which can reveal rather than obscure and distort what we value. Elgin makes the case in more detail, arguing that emotions, no less than beliefs, are fully paid-up members in the quest for understanding and in the process of reflective equilibrium as an important dimension of that quest. "Only by fostering reflective equilibrium does the insight [accessed through an emotion] earn itself a place…Deliverances of emotion, like other initially tenable commitments, are subject to compensation, revision, or rejection in the construction of a tenable system of thought" (Elgin, 1996, p. 150, 152).

Meanwhile, in his discussion of the role of imagination in moral deliberation, which he notes operates "in accordance with the general framework of reflective equilibrium", Gaut writes:

> Up to this point I have talked of the judge 'judging' what ought to be done to him were he in the target's position. This way of putting the matter should not be taken, however, to rule out the role of feelings and emotions in making the judgement. The judge may feel pity towards himself in the target's position, or feel anger at the way he as target has been treated. (Gaut, 2007, p. 159)

So there is an acknowledgement here of the role of both emotion and imagination in moral deliberation under the aegis of RE. Moreover, as Tamar Gendler notes, Rawls himself draws on imaginative techniques within the framework of reflective equilibrium in order to build his theory of justice. In particular, she notes the critical role of Rawls' "original position" thought experiment, arguing that through this device Rawls makes "the moral stance cognitively available at a moment of moral decision-making" (2010, p. 132; see also Elgin, 1996, p. 191). In other words, through the thought experiment, Rawls invites us to imagine ourselves deliberating and acting in the "original position", and what we learn from this exercise then becomes an input in the process of reflective equilibrium. It would be a mistake, then, to hold that acts of imagining cannot be accommodated by RE, and a further mistake to overlook the use of imaginative and emotional techniques by philosophers—including analytic philosophers—even if we associate such techniques in the first instance with the arts. Imaginative practices—thought experiments and other fictions, metaphors, exemplifications—are critical to the sciences as well as the arts, and emotions are "deeply and intricately interwoven into our fabric of commitments" (Elgin, 1996, p. 163).[17]

This doesn't make moral situations identical with artworks, of course. But it does suggest that in neither domain can reflective equilibrium be quite as dispassionate and affect-free as the phrase implies; and given the actual constitution of humans, nor should it be. (Ought implies can, so if some state of affairs cannot be realised, we ought not make it a normative requirement.) The fact that we respond to morally-charged situations with anger and outrage, compassion and sympathy, disgust and pride, and not merely with the "cold cognition" of belief, is relevant to the judgements we make about these situations. The emotions tells us how we value the objects, events, and agents we encounter, but the story doesn't end with these responses as they initially arise—rather they are weighed, adjusted, and (as far as possible) brought into relations of alignment and mutual support. In other words, we seek to bring our responses into a state of reflective equilibrium. On this view, then, reflective equilibrium is not only holistic and dynamic, but emotional and embodied as well—certainly in the context of aesthetic appreciation, and arguably in other domains too.

Note that this view is related to Sontag's anti-hermeneutic stance towards art appreciation, discussed above. For in both cases, there is an accent on the way art engages us through perception, imagination, and emotion. In this way our experience of art goes beyond questions of belief and propositional understanding. But the present claim is more moderate and nuanced: where Sontag wants to jettison interpretation and reflection—replacing the hermeneutics of art with an "erotics" of art—on the argument advanced here, the two are inextricably connected. While we can distinguish the cognitive dimension of artworks—the ideas that they bear, their explicit or implicit claims about the world—we cannot extract them from their perceptual, imaginative, and emotional vehicles without travesty. *Oppenheimer* doesn't just state a set of propositions about the history that it narrates, nor does it simply present a lot of kinetic imagery and thunderous sounds for our "erotic" delectation. Christopher Nolan's stylistics dramatize a set of ideas, a perspective on, the story of

[17] Elgin explores the contribution of fiction, metaphor, and exemplification to understanding in science and art in Chap. 6 of her 1996.

Robert Oppenheimer and the development of the atomic bomb in the US; whether we find the design of the film bold or bombastic, we fail to appreciate it unless we grasp both the cognitive and the emotional sides of the equation.

6 Conclusion: equilibrium achieved, provisionally

I began this essay by asking: is RE relevant to aesthetic appreciation? I have answered in the affirmative. But note that I have not been asking whether RE provides a comprehensive account of aesthetic appreciation, and nor do I claim that it does. On the interpretation of RE tendered here, it isn't quite a comprehensive philosophical methodology, and it certainly doesn't furnish a complete theory of aesthetic experience, appreciation, and judgement. "To say that mathematicians or biochemists or historians strive to bring their opinions into equilibrium", avers Timothy Williamson, "would be sadly inadequate as even a summary description of their method of research" (Willliamson, 2007, p. 246; see also Kelly and McGrath, 2010). The same goes for aesthetics. RE is an important part of the fabric of aesthetic appreciation, but not the whole cloth: a necessary but not sufficient condition for the generation of optimally-justified aesthetic judgements, and a typical characteristic—common if not strictly necessary—of more ordinary aesthetic judgements, the back-and-forth of RE following quickly in the wake of our initial aesthetic responses.[18]

Notwithstanding his scepticism about the adequacy of RE as a description of scholarly and scientific methods, Williamson is insistent on the epistemic role of literary and art criticism. As he notes: "Every normal reader has [a] sort of elementary literary critical knowledge [manifest in their understanding of basic narrative facts, such as who the protagonist of a work is]. Those who know far more about the historical context in which literary works were produced, read them many times with unusual attention, carefully analyse their structure, and so on, naturally have far more knowledge than casual readers do" (Williamson, 2014, p. 37). It is in virtue of the epistemic function of criticism that RE earns its place as an aspect of criticism, while theorizing reflective equilibrium helps us identify and sharpen our understanding of a process which must be central to any enterprise with epistemic goals. And by such identification and sharpening, we are also in a position to ameliorate it.

I've made the case that reflective equilibrium, as both a process of reasoning and an embodied state of mind arising from that process, plays a significant role in aes-

[18] Sarah McGrath holds that, in the moral domain, RE is neither necessary nor sufficient for the generation of moral knowledge (McGrath, 2019, p. 19). But there is common ground between our positions. McGrath is concerned to stress that, if we can be said to possess moral knowledge at all, then we possess (some) moral knowledge prior to the exercise of RE. Similarly we might say that an appreciator is likely to have some knowledge of an aesthetic work in advance of RE, just by dint of minimally engaging with the work against the backdrop of some relevant prior experience of the type of aesthetic object in question. The point I am concerned to stress, and the burden of the argument of the present essay, however, is that RE is necessary for *properly justified* aesthetic judgements (but not for initial reactions, first takes, shots-from-the-hip, and so forth). In McGrath's terms, I am committed to a *modest* rather than *ambitious* version of the method of RE, insofar as the independent epistemic standing of the inputs, at the outset of the process, plays an important role. This modest conception of RE nonetheless plays a significant role in the making of aesthetic judgements.

thetic appreciation, drawing on a variety of argumentative strategies and forms of evidence in support of this claim. At the most abstract level, far from being restricted to the moral and political spheres, reflective equilibrium is pertinent to reasoning in a wide range of contexts and domains (as was evident in the framing and exposition of the concept by Rawls, and through its origins with Goodman in the justification of the rules of logical inference). At the most concrete level, we have seen how the "back and forth" movement between particular judgements (thin and thick) about individual works of art, and the principles and assumptions lying behind those judgements, is characteristic of the critical appreciation of art. Moreover, in the context of more extended and philosophically self-conscious criticism, the three-way interplay among specific judgements, principles, and background theories characteristic of "wide" reflective equilibrium is similarly apparent. In between the pincer movement formed by these two arguments, I've noted Rawls' own recognition of the place of aesthetic activity and experience in human flourishing ("human beings [are] importantly moved not only by the pressure of bodily needs, but also by the desire to do things enjoyed simply for their own sakes"), and thus its pertinence to justice. That recognition is also registered in Rawls' use of imaginative and affective techniques, such as the "original position" thought experiment, in building the case for his theory of justice. This in turn leads to the additional hypothesis that reflective equilibrium as it is manifest in the aesthetic domain illuminates its nature in human deliberation more generally, which cannot—and should not—be entirely divorced from its embodied and emotional facets. That is why, in this conclusion, I speak of an "embodied state of mind" arising from the process of RE rather than, more narrowly, a mutually supporting "set of beliefs". Within the debates on reflective equilibrium, as in so many areas of philosophy, a proper reckoning of the significance of aesthetic experience is long overdue. If the present paper creates some momentum in that direction, it will have served its purpose.

Acknowledgements My thanks to Claus Beisbart, Georg Brun, and Gregor Betz for their feedback and patience; the two anonymous journal referees for pushing me to develop and improve this paper along several dimensions; and to Berys Gaut, and the participants at the conference "Reflective equilibrium 51 years after *A Theory of Justice*", Bern (12–14 May 2022), where I presented an initial version of the paper, for helpful discussion.

Funding The research and writing of this paper was undertaken as part of the *Character Engagement and Moral Understanding in Screen Stories* project, funded by the Templeton Religion Trust (TRT-2021-10445).

References

Bradshaw, P. (2022, September 5). A Guinness-black comedy of male pain. Review of *The Banshees of Inisherin*. *The Guardian*. https://www.theguardian.com/film/2022/sep/05/the-banshees-of-inisherin-review-martin-mcdonagh

Beebee, H. (2017). Philosophical scepticism and the aims of philosophy. *Proceedings of the Aristotelian Society*, *CXVIII*(Part 1), 1–24.

Beisbart, C., & Brun, G. (2024). Is there a defensible conception of reflective equilibrium? *Synthese, 203*(79), 1–26.

Carroll, N. (2002). The wheel of virtue: Art, literature, and moral knowledge. *The Journal of Aesthetics and Art Criticism, 60*(1), 3–26.

Chang, J. (2023, August 11). 'Oppenheimer' doesn't show us Hiroshima or Nagasaki. That's an act of rigor, not erasure. *The Los Angeles Times*. https://www.latimes.com/entertainment-arts/movies/story/2023-08-11/oppenheimer-atomic-bomb-hiroshima-nagasaki-christopher-nolan

Currie, G. (2004). Aesthetic explanation. In *Arts and minds* (pp. 241–254). Clarendon Press.

Daniels, N. (1979). Wide reflective equilibrium and theory acceptance in ethics. *The Journal of Philosophy, 76*(5), 256–282.

Daniels, N. (2016). Reflective equilibrium. *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/sum2020/entries/reflective-equilibrium/

Davies, S. (2006). *The philosophy of art*. Blackwell.

Elgin, C. Z. (1996). *Considered judgment*. Princeton University Press.

Elgin, C. Z. (2017). *True enough*. MIT Press.

Elgin, C. Z. (2022). Dynamic epistemic equilibrium. [Conference paper presentation] Reflective equilibrium: 51 Years After *A Theory of Justice* conference, Bern, 12–14 May 2022.

Gaut, B. (2007). *Art, Emotion and Ethics.* Oxford University Press.

Gendler, T. S. (2010). *Intuition, imagination, and philosophical methodology*. Oxford University Press.

Gilmore, J. (2020). *Apt feelings: Feelings for fictions and other creatures of the mind*. Oxford University Press.

Goodman, N. (1954). *Fact, fiction, and forecast*. University of London/Athlone Press.

Harold, J. (2021). Review of Jonathan Gilmore, *Apt feelings*. *The Journal of Aesthetics and Art Criticism, 79*(2), 272–275.

Hills, A. (2021). Review of Sarah McGrath, *Moral knowledge. Mind, 130*(520), 1401–1407.

Kelly, T., & McGrath, S. (2010). Is reflective equilibrium enough? *Philosophical Perspectives, 24*, 325–359.

Kivy, P. (1997). The laboratory of fictional truth. In *Philosophies of art: An essay in differences* (pp. 120–139). Cambridge University Press.

Lamarque, P. (2009). *The philosophy of literature*. Blackwell.

Lamarque, P., & Olsen, S. H. (1994). *Truth, fiction, and literature: A philosophical perspective*. Clarendon Press.

Levinson, J. (2006). V. Aesthetic Properties. In *Contemplating art: Essays in aesthetics* (pp. 315–351). Clarendon Press.

Lewis, D. (1983). *Philosophical papers* (Vol. I). Oxford University Press.

Lewis, J. A. (1998, May 2). Rothko: Not nothing. *The Washington Post*. https://www.washingtonpost.com/archive/lifestyle/style/1998/05/03/rothko-not-nothing/74d80839-7281-40bd-8e23-500d2a160ded/

Lopes, D. M. (2014). *Beyond art*. Oxford University Press.

McGrath, S. (2019). *Moral knowledge*. Oxford University Press.

Raffman, D. (2003). Is twelve-tone music artistically defective? *Midwest Studies in Philosophy, 27*, 69–87.

Rawls, J. (1951). Outline of a decision procedure for ethics. *The Philosophical Review, 60*(2), 177–197.

Rawls, J. (1971/1999). *A theory of justice*. Revised ed. Harvard University Press.

Rawls, J. (1974). The independence of moral theory. *Proceedings and Addresses of the American Philosophical Association, 1974–1975*(48), 5–22.

Scott, A. O. (2022, October 20). Giving your friend the finger. Review of *The Banshees of Inisherin*. *The New York Times*. https://www.nytimes.com/2022/10/20/movies/the-banshees-of-inisherin-review.html

Shapiro, L. (2019). *Embodied cognition* (2nd ed.). Routledge.

Sibley, F. (1959/2001). Aesthetic concepts. In J. Benson, B. Redfern, & J. R. Cox (Eds.), *Approach to aesthetics: Collected papers on philosophical aesthetics*. Clarendon Press. Originally published in *The Philosophical Review,* 68.

Sibley, F. (1968/2001). Objectivity and aesthetics. In J. Benson, B. Redfern, & J. R. Cox (Eds.), *Approach to aesthetics: Collected papers on philosophical aesthetics*. Clarendon Press. Originally published in *Proceedings of the Aristotelian Society,* 42.

Smith, M. (2023a). Human flourishing, philosophical naturalism, and aesthetic value. In T. Corrigan (Ed.), *Cinema, media, and human flourishing.* Oxford University Press.

Smith, M. (2023b). The reality of (screen) characters. In: R. Davies, P. Russo, & C. Tieber (Eds.), *The Palgrave Handbook of Screenwriting Studies* (pp. 111–130). Palgrave Macmillan.

Sontag, S. (1966/1969). *Against interpretation and other essays*. Laurel/Dell.

Stolnitz, J. (1992). On the cognitive triviality of art. *British Journal of Aesthetics, 32*(3), 191–200.

Walden, K. (2013). In defence of reflective equilibrium. *Philosophical Studies, 166*, 243–256.

Walton, K. (1970). Categories of art. *The Philosophical Review, 79*(3), 334–367.

Williamson, T. (2007). *The philosophy of philosophy*. Blackwell.

Williamson, T. (2014). The unclarity of naturalism. In M. C. Haug (Ed.), *Philosophical methodology: The Armchair or the laboratory?* Routledge.

Wollheim, R. (1980). *Art and its objects* (2nd ed.). Cambridge University Press.

Zangwill, N. (2001). *The metaphysics of beauty*. Cornell University Press.