INTERPRETATION, INTENTIONS, AND RESPONSIBILITY

PETER ALWARD

In this paper, I defend a contextualist account of the role of authors' intentions in interpretation, according to which their role depends on readers' interpretive interests. In light of a general discussion of intentions and responsibility, I argue that insofar as readers are interested in attributing authorial responsibility for interpretations of fictional works, authors' intentions need to play a central role in those interpretations. And I investigate the implications of this account for 'accidental authorship', cases in which interpretations of a work are neither intended nor reasonably foreseen by the author.

The role of intentions in the interpretation of fictional works has long been a matter of dispute. Views range from those that take authorial intentions to be of paramount importance to those that take them to be entirely irrelevant.[1] And it is fairly clear why one might expect this to be a matter of controversy. On the one hand, fictional works are shaped by the intentions of their authors: authors choose the words and sentences that they do largely as a means of achieving their intentions. On the other hand, there is normally a large spatio-temporal gap between authors' acts of producing fictional works and appreciators' acts of interpreting them. And not only does this gap limit appreciators' access to authors' intentions, it also diminishes their relevance to appreciators' interpretive projects.

In this paper, I defend what might be characterized as a contextualist account of the role of authors' intentions in interpretation – an account according to which their role depends on appreciators' interpretive interests. Moreover, I argue that insofar as appreciators are interested in attributing to authors responsibility for interpretations of their works, authors' intentions need to play a central role in those interpretations. And I develop an account of the role played by authors' intentions – and psychological states more generally – in interpretations for which they are responsible. Finally, I investigate the implications of this account for what I propose to call 'accidental authorship',

[1] See E. D. Hirsch, *Validity in Interpretation* (New Haven, CT: Yale University Press, 1967) and Monroe C. Beardsley, 'Intentions and Interpretations: A Fallacy Revived', in *Aesthetics and the Philosophy of Art: The Analytic Tradition*, ed. Peter Lamarque and Stein Haugom Olsen (Oxford: Blackwell, 2004), 189–99. Noël Carroll, 'Interpretation and Intention: The Debate between Hypothetical and Actual Intentionalism', *Metaphilosophy* 31 (2000): 75–95, and Jerrold Levinson, 'Intention and Interpretation in Literature', in *The Pleasures of Aesthetics: Philosophical Essays* (Ithaca, NY: Cornell University Press, 1996), 175–213, among others, defend intermediate positions.

cases in which interpretations of a work are neither intended nor reasonably foreseen by the work's author.

This paper consists of three main parts. In the first part the theoretical background presupposed by the central argument is developed. This includes discussions of interpretation, theories of interpretation, appreciators' interpretive interests, and the various roles authorial intentions might play in interpretation. In the second part responsibility-centred interpretive intentionalism is developed. This includes a general discussion of intentions and responsibility, as well as discussions of what sorts of interpretation are supported by texts, the conditions under which authors are responsible for supported interpretations, and the roles intentions might play within a responsibility-centred account of interpretation. And in the final part, accidental authorship is explored. This includes a taxonomy of various kinds of accidental authorship, as well as the application of responsibility-centred interpretive intentionalism to some actual cases.

I. THEORETICAL BACKGROUND

Questions of interpretation arise for a number of very different kinds of object, ranging from scientific theories to artworks to casual conversations. And although there may be a certain degree of commonality in what the interpretation of these various items involves, the focus here will be on works of fiction. These include literary fictions, works whose content is expressed primarily by means of language, such as novels, novellas, and short stories. But they also include fictional works whose content is, in whole or in part, expressed in other ways, such as films, plays, pictures, songs, and the like. Note: insofar as an artwork counts as non-fiction – or perhaps is neither fiction nor non-fiction – it is not the kind of thing whose interpretation is under consideration here.

I.1. INTERPRETATION

At bottom, interpreting a work of fiction is a way of appreciating or engaging with it; in particular, it consists in appreciating or engaging with it as meaningful. And this involves more than the mere recognition of its meaningfulness; it includes, in addition, understanding it as having a particular meaning or collection of meanings. Insofar as a work consists of conventionally meaningful words or other symbols, interpreting it minimally requires understanding the meanings of those words or symbols, and their concatenations.[2] But it also involves understanding works to have specific descriptive contents (which

[2] Of course, the meanings of words and other symbols can change over time and may differ between different individuals and groups at a given time. As a result, appreciators strictly speaking understand texts relative to times and places. I will largely ignore this complication for present purposes.

characterize various features of the fictional characters and worlds they generate), specific narrative contents (which characterize what happens in the stories they generate), and maybe even argumentative contents, as well as having various themes, morals, allusions, and the like. And such understandings are grounded not only in the conventional meanings of texts, but also in judgements of co-reference of names and other referring expressions, judgements regarding the modes of various sentences – whether they are literal, metaphorical, ironic, and so forth – and judgements regarding the degree of narrative reliability, among other things.

It is worth noting that there are two very different senses in which an appreciator might be thought to interpret a work of fiction: having or developing a theory of meaning for it; and having contentful appreciative experiences while reading it. To have a theory of meaning for a work is to have a set of beliefs about its various dimensions of meaning or content: its narrative content, its descriptive content, its thematic content, and so on. Such beliefs may be expressed linguistically – perhaps even in the form of an interpretive essay – but need not be so. By appreciative experiences I mean the kinds of experiences appreciators have when they read or watch a work of fiction as a work of fiction, when they are 'caught up in the story'. To say such experiences are contentful is to say they are, or include, propositional attitudes. And differences in their contents correspond to differences in the contents of these experiences. Following Kendall Walton and Gregory Currie, we might take the appreciative experiences to involve a kind of make-believe.[3] And, on this picture, differences in interpretation would correspond to differences in what appreciators make-believe. The discussion in this paper is designed to encompass both senses of interpretation.

I.2. THEORIES OF INTERPRETATION

The main goal of this paper is to develop a theory of interpretation for appreciators with certain sorts of interpretive interests and, in particular, interests in attributing responsibility to authors. As a result, it may prove fruitful to have a preliminary discussion of theories of interpretation in general. The first thing to note is the distinction between descriptive and normative theories of interpretation. A descriptive theory gives an account of how appreciators in fact go about interpreting the fictional works that they read, that is, how they generate their theories of meaning or how they come to have the appreciative

[3] Gregory Currie, *The Nature of Fiction* (Cambridge: Cambridge University Press, 1990) and Kendall Walton, *Mimesis as Make-Believe: On the Foundations of the Representational Arts* (Cambridge, MA: Harvard University Press, 1990).

experiences that they do. A normative theory, by contrast, provides an account, not of how appreciators in fact go about interpreting fictional works, but rather of how they ought to go about doing so. The focus of this paper will be on normative theories of interpretation. Nevertheless, if a normative theory is to count as a theory of interpretation – as opposed to a theory of something else –, it needs to be suitably related to the behaviour that interpreters in fact engage in.

A normative theory of interpretation is a theory of how people ought to go about generating interpretations of the fictional works they read. At its core, this consists of the specification of an interpretive project and a set of criteria for the successful completion of this project. An interpretive project is a procedure that takes an appreciator from the text of the work at issue to a theory of meaning or a series of contentful appreciative experiences. The first element of some such procedure involves a specification of the resources – in addition to the text – that appreciators may draw upon in generating their interpretations. More precisely, what is required is a specification of what extra-textual resources an appreciator must draw upon, what resources she may draw upon, and what resources must not be utilized. So, for example, an interpretive project might require that appreciators utilize any available information about the author's meaning-intentions, prohibit information about the critical response to the work, or permit appreciators to draw upon Marxist or psychoanalytic theory without requiring that they do so. In addition, a theory of interpretation includes some kind of formula from textual and non-textual inputs to an interpretation or a collection of interpretations. This could range from relatively strict rules to relatively loose rules of thumb, and might include the relative weights to be placed on the various inputs, among other things. So, for example, a strict form of interpretive intentionalism might require the identification of the meaning of a work with the author's intentions, whereas a looser form might simply require that they be taken into consideration along with the text and other non-textual inputs.

Interpretive projects are designed to achieve various interpretive goals and are successful to the extent that they in fact do so. Realist interpretive projects, for example, are designed to uncover meanings possessed by their objects prior to, and independently of, interpretations of them. Hence, a criterion of success for a realist project is the generation of an interpretation that corresponds in the right way to these meanings. Non-realist interpretive projects, by contrast, either deny that their objects have interpretation-independent meanings or, more modestly, deny that such meanings are relevant to certain interpretive goals. As a result, a criterion of success for such projects cannot be any kind of correspondence with interpretation-independent meanings. Alternative criteria

are not hard to find, however. Examples might include the quality of the appreciative experience the interpretation yields or the extent to which interpreting the work in that way contributes to certain political goals.

I.3. META-INTERPRETIVE PLURALISM AND INTERPRETIVE INTERESTS

A background assumption of this paper is what may usefully be called 'meta-interpretive pluralism'. According to (object level) interpretive pluralism, there are multiple correct interpretations of a work of fiction. So, for example, an interpretation of a fictional work according to which the protagonist dies at the end and an interpretation according to which she survives could both be correct. The competing view is interpretive monism, according to which a fictional work has a unique correct (complete) interpretation. On this view, insofar as they are incompatible, at most one of the interpretations concerning the mortality of our fictional protagonist could be correct. According to meta-interpretive pluralism, by contrast, there are multiple coherent and legitimate interpretive projects, rather than there being a single project in which all appreciators ought to engage. So, for example, both an intentionalist interpretive project (in which evidence of the author's narrative intentions is required among extra-textual resources) and an anti-intentionalist project (in which such evidence is prohibited) could be coherent, and there might be nothing illegitimate or inappropriate about engaging in either of them.

Moreover, this view is, in a sense, contextualist. If multiple interpretive projects are coherent and legitimate, the question remains exactly which project an appreciator should engage in. And that depends on the appreciator's interpretive interests. Different appreciators – and even individual appreciators at different times – have differing interpretive goals; that is, they hope to achieve different outcomes by means of their interpretations of the fictional works they read. Some appreciators merely hope to have an enjoyable reading experience. Others hope to come up with an interpretation that will convince others to read the work as well.[4] And a third category of appreciators aim for interpretations that will enable them to use a work to achieve certain political goals. Which interpretive project an appreciator should engage in depends on which project is best suited to the achievement of these goals. One project might be well suited to the generation of interpretations that yield enjoyable reading experiences whereas another might be better suited to the generation of interpretations that facilitate political goals. A full defence of meta-interpretive pluralism in part includes establishing that a variety of interpretive projects are coherent and

[4] See Peter Alward, 'Rogues or Lovers: Value Maximizing Interpretations of *Withnail and I*', *Projections* 12 (2018): 39–54.

legitimate.[5] But it also includes showing what the various points of those projects are, that is, what interpretive goals they (are designed to) enable appreciators to achieve. My paper can be viewed as one part of this overall argumentative strategy.

It is worth noting the contrast between meta-interpretive pluralism and the kind of interpretive pluralism defended by Robert Stecker.[6] On Stecker's view, different interpreters have different interpretive aims, and different interpretive projects are more or less well suited to these aims, which is generally the picture on offer here. But in addition, Stecker identifies the meaning of a work with 'what the artist does in the work that is artistically significant'.[7] As a result, Stecker singles out projects that aim to provide an accurate account of what an artist does in a work that is artistically significant as making claims about the meanings of artworks. Other projects – to the extent that they can be understood as making assertions at all, rather than making recommendations, suggestions, or the like – make claims about what artworks could mean rather than what they do.[8] According to my theory, by contrast, what Stecker identifies with the meaning of a work is just one aim among many that interpreters sometimes share, having no special status. And many, if not most, interpretive projects can be understood to yield interest-relative claims about what artworks in fact mean rather than claims about what they could mean. In most other respects, however, Stecker's picture is one I find quite congenial.

Many people who are familiar with Stecker's work find his suggestion that interpretive projects which aim towards uncovering a work's (interpretation-independent) meaning have a special status to be very persuasive. Although a thorough analysis of Stecker's thesis would be out of place in the present context, there is one comment worth making here. According to Stecker, only interpretive projects aimed at uncovering what the artist does in a work that is aesthetically significant yield assertions about what the work in fact means; other interpretive projects yield only assertions about what the work could mean, if they yield assertions at all. But this would require substantially recasting the linguistic outputs of these latter projects – which often look, at least

5 For example, Peter Alward, 'Butter Knives and Screwdrivers: An Intentionalist Defense of Radical Constructivism', *Journal of Aesthetics and Art Criticism* 72 (2014): 247–60, is designed to defend constructivist interpretive projects against Stecker's charge that they are incoherent. Compare Robert Stecker, 'The Constructivist's Dilemma', *Journal of Aesthetics and Art Criticism* 55 (1997): 43–51.

6 Robert Stecker, *Interpretation and Construction: Art, Speech, and the Law* (Oxford: Blackwell, 2003), 57–63. This view is also endorsed by Ted Nannicelli, 'Ethical Criticism and the Interpretation of Art', *Journal of Aesthetics and Art Criticism* 75 (2017): 406.

7 Stecker, *Interpretation and Construction*, 60.

8 Ibid., 43.

superficially, like assertions of work meaning – and perhaps even rejecting interpreters' own judgements of what they mean. This hardly counts as a decisive objection to Stecker's view, but it does point to a troubling implication that a full defence of Stecker would ultimately need to address.

I.4. INTERPRETIVE INTERESTS AND AUTHORIAL INTENTIONS

The question I want to address in this paper is exactly what sorts of interpretive goals would be served by intentionalist interpretive projects. And there are a number of candidate goals out there. One possibility is the goal of respecting an author's proprietary rights. The idea here is that interpretive projects in which the author's intentions play no role run the risk of violating property rights over a work she has written. And so anyone whose interpretive goals include respecting the author's rights ought to engage only in projects in which those intentions play the requisite role. The trouble with this strategy is that it is not clear how interpreting a work in a way that disregards the author's intentions violates her rights over it. Destroying or damaging an artwork – even one you have purchased – may well violate an artist's rights; but interpreting a fictional work in a way the author did not intend neither destroys it nor prevents others from interpreting it as intended. Moreover, one might violate an author's intellectual property rights over a fictional work by falsely taking credit for it or failing to properly compensate the author for one's use of it; but again, ignoring an author's intentions by itself does not involve doing either of these things. Of course, it might frustrate an author's desire that her work be interpreted as she intended, but it is far from clear that she has any right to have it interpreted (only) in this way.

Another possibility is the goal of communication with the author. The idea here is that interpretive projects in which the author's intentions play no role run the risk of a failure of communication. And so anyone whose interpretive goals include communicating with the author ought to restrict themselves to intentionalist projects. Now, it is true that speakers' intentions play a central role in certain forms of communication; one simple example is communication designed for the coordination of action.[9] In order to aid you in your enterprises, I need to know what you are going to do and what you want me to do. And this requires that I discern what you mean by your words. Otherwise the result might be actions on my part which frustrate your enterprises rather than furthering

[9] In some cases, however, the nature of the communicative exchange does not require intention-directed interpretation. Consider, for example, a conversation designed to produce humour by means of interpreting speakers' utterances in a way that is counter to their intentions.

them. The trouble is that, normally at least, the kind of communication that occurs between authors of fiction and their appreciators is not geared towards the coordination of their actions. This is not to say that there is some kind of standard approach towards works of art which is privileged among other approaches; rather, it is just to point out that in most ordinary novels and films, authors do not intend for appreciators to do anything beyond reading or watching them. As a result, appreciators who ignore authors' intentions in interpreting their works are unlikely to frustrate authors' enterprises in this way.[10] Of course, some authors may intend their work to have a certain political content and to achieve corresponding political goals by means of communicating this content. In some such case, it would be incumbent upon an appreciator who desired to engage in collective political action with the author to utilize an intentionalist approach to interpretation.

In this section, two proposals regarding the sorts of interpretive goals that might be served by intentionalist interpretive projects were presented and found lacking. This leaves us with an open question as to how we might go about finding a rationale for such projects. In the next part of this essay, we will explore the idea that some such rationale is to be found in an account of authorial responsibility for the works they produce.

II. INTERPRETATION AND RESPONSIBILITY

In what follows, the relation between attributions of authorial responsibility and intentionalist interpretive projects will be explored. Some appreciators want to hold authors responsible for the fictional works they compose, praising them for the rich and rewarding stories they have told or blaming them for the moral or artistic defects of their works. Now, whether a fictional work warrants praise or blame depends, at least in part, on what it means, which, given the kind of pluralism on offer here, amounts to how it is interpreted. As a result, in order for the justification for praise and blame to be transferred from the work to the author, the interpretation relative to which the work is evaluated needs to be suitably related to the author. In particular, it has to be the product of the right kind of intentionalist interpretive project. It should be noted that even if certain interpretations of a work count as correct in some interest-independent sense, there is no presupposition that all of the interpretations under consideration here are correct in this sense. Although incorrect – or unsupported – interpretations

[10] See Noël Carroll, 'Art, Intention, and Conversation', in *Intention and Interpretation*, ed. Gary Iseminger (Philadelphia: Temple University Press, 1992), 97–131, for a more sophisticated attempt to defend interpretive intentionalism by appeal to communication.

will be primarily considered in the section on accidental authorship, an author could nevertheless be responsible for an incorrect interpretation of her work insofar as she ought to have foreseen it would be interpreted in that way.

II.1. INTENTIONS AND RESPONSIBILITY

Before discussing the connection between authorial intentions and authorial responsibility, it may prove fruitful to discuss the relationship between intentions and responsibility more generally. Typically what we hold people responsible for are their actions, things they do intentionally. And these actions can be positive, negative, or neutral in a number of different dimensions. For simplicity, I am going to focus only on moral and aesthetic value here. Now, it is commonplace to note that by means of a single behaviour, an agent can perform a plurality of actions. So, for example, by a certain movement of her finger, an agent can perform the actions of moving her finger, pulling the trigger of a gun she is holding, firing the gun, and shooting someone. Moreover, even though they are all actions of hers, the agent performs some of them by performing others: in the case at hand, she pulls the trigger by moving her finger, she fires the gun by pulling the trigger, and she shoots someone by firing the gun.[11] Finally, there is a stronger and a weaker sense in which an action can be intentional. An action is intentional in a strong sense if it is an action the agent was trying to perform: the agent had the goal of performing an action of that kind, attempted to do so, and succeeded in her attempt. An agent's act of shooting someone is intentional in this sense if a goal of hers was to shoot someone by means of what she did. An action is intentional in a weaker sense just in case it is performed by means of an action that is intentional in a strong sense, whether or not it itself is intentional in a strong sense.[12] Suppose, for example, that one of my goals is to fire a gun and I do so; and suppose that by firing the gun I shoot someone. My action of shooting that person would be intentional in a weak sense in such circumstances whether or not I had the goal of doing so. In what follows, we will focus on actions that are intentional in this weak sense.

For simplicity, I am going to focus on the following somewhat artificial case. Suppose that Fred is being held by a robotic arm over a raging fire, and suppose that some distance away is a lever which controls the robotic arm: if the lever is

[11] Of course, some of the things one does by means of one's actions are not further actions but rather consequences of those actions. The distinction between actions and their consequences should make no difference to the discussion here.

[12] For more on the individuation of action, see Donald Davidson, 'Actions, Reasons, and Causes', *Journal of Philosophy* 60 (1963): 685–700.

pushed forward, there is a high probability that the robotic hand will open and drop Fred into the fire; if it is pulled back, there is a high probability that the arm will move away from the fire and release Fred to safety. Now, suppose Mary intentionally moves the lever forward and thereby drops Fred into the fire. Broadly speaking, there are three cases of interest here. First, Mary intentionally dropped Fred into the fire in a strong sense. Her goal was to drop him into the fire and she did so by means of pushing the lever forward. Second, Mary's dropping Fred into the fire was entirely accidental. Dropping Fred into the fire was not a goal of hers, and she did not realize that by pushing the lever forward she would do so, nor is there any reason to believe that she should have realized this, general admonitions against moving random levers aside. And third, Mary's action was reckless or negligent. Dropping Fred into the fire was not a goal of hers, but she either realized – or should have realized – that by moving the lever forward she would likely do so. In the first case, in which Mary intentionally dropped Fred into the fire, she is wholly responsible for what she did. And since what she did was highly morally negatively valuable, she deserves a proportionally high degree of blame for her action. In the second case, in which dropping Fred into the fire was entirely accidental, she is not responsible at all for what she did. As a result, any blame she receives for her action is entirely unwarranted despite how morally disvaluable it was. And in the third case, in which her action was reckless or negligent, Mary is partly responsible for what she did. As a result, blaming her for her action is warranted, but the degree of blame that is appropriate to attribute to her is less than it would be had her action been intentional.[13]

Similar results ensue if an agent's actions are morally valuable rather than disvaluable. Suppose that rather than moving the lever forward, Mary intentionally moves the lever backwards and thereby moves Fred to safety. Again there are three cases of interest. First, Mary intentionally, in the strong sense, moved Fred to safety. Second, her act of moving him to safety was accidental: she lacked the goal of so moving him and did not realize that by moving the lever backwards she would do so. And third, she knew, or should have known, she would likely move Fred to safety by moving the lever backwards, but this was not a goal of hers. In the first case, Mary is wholly responsible for what she did, and since what she did was highly morally valuable, she deserves a proportionally high degree of praise for her action. And in the second case, she is again not at all responsible for what she did and any praise she receives for her action is entirely unwarranted. After all, her moving of Fred to safety was pure

[13] Much of the discussion here draws on Richard Parker, 'Blame, Punishment, and the Role of Result', *American Philosophical Quarterly* 21 (1984): 269–76.

happenstance, not by design. In the third case, although it is reasonable to suppose that Mary is partly responsible for what she did, I am not entirely convinced that this suffices to warrant praising her for her action. Even though moving Fred to safety is highly morally valuable, her failure to have aimed to rescue Fred strikes me as a serious moral failing, perhaps serious enough to remove any justification for praise there might otherwise be.

There is one final kind of case worth mentioning: that of unsuccessful action. Suppose, for example, that Mary's goal is to drop Fred into the fire and she attempts to do by means of pushing the lever forward, but, for whatever reason, the robotic hand does not open and drop Fred. Or suppose that Mary's goal is to move Fred safely away from the fire and she attempts to do so by means of pulling the lever backwards, but, for whatever reason, the robotic arm does not move Fred away from the fire. Now, it might be true that in both cases Mary deserves the same degree of praise or blame as she would have if her actions had been successful. After all, the argument goes, the only difference between successful and unsuccessful action is luck, and luck is not morally significant.[14] But even if this is right, what is important to note is that whatever praise or blame Mary deserves does not stem from her being responsible for dropping or rescuing Fred simply because she did not perform either of these actions. This distinction will prove important in what follows below.

II.2. INTERPRETED WORKS AND SUPPORTED INTERPRETATIONS

Our central concern here is the evaluation not of actions but of fictional works. Now, it is commonplace to suppose that the aesthetic and moral value of a work depends on what it means. But insofar as one endorses the kind of pluralism on offer here – which eschews the idea that fictional works have interest-independent meanings – it is better to say that the value of a work varies with its interpretation. If, however, one resists this kind of pluralism, one should understand this claim to be that value judgements about a work vary with its interpretation and not the value of the work per se. More precisely, given that the aesthetic and moral value of a work can vary with its interpretation, the focus here is on interpreted-works: fictional works as interpreted in a particular way. And the question is which interpreted-works an author is responsible for, in the sense that the value or disvalue of the interpreted-work justifies praise or blame for the author.

One might worry that taking interpreted-works – rather than works *simpliciter* – to be the objects of evaluation here is both unmotivated and ontologically

[14] Ibid.

profligate. After all, the matter can be more simply formulated in terms of the meanings in fact possessed by works. This worry, however, is misplaced. First, the motivation for the appeal to interpreted-works is a formulation that yields a common object of evaluation, which can serve as an intentional product of authorial activity, a foreseen but unintended product of authorial activity, and an accidental product of authorial activity. And just as a weaker sense of intentional action was required to distinguish between cases in which a given action was intentional, negligent, or accidental, the appeal to interpreted-works is required to distinguish between cases in which a given object of evaluation is the product of the corresponding kinds of authorial activity. Second, the appeal to interpreted-works does not come with any suspect ontological commitments. They can be viewed as collections (or pairs) whose members include a work and an interpretation. As a result, insofar as one endorses a principle to the effect that for any two objects that exist a collection whose members include just those two objects exists as well, one is already committed to the existence of interpreted-works.

In some cases, the work and the interpretation which together constitute an interpreted-work are related in the sense that the interpretation is supported by – or embodied in – the work. The idea here is that some interpretations of a given work are grounded in the text of the work – are reasonable or justified given the text – and some are not. Although a theory of supported interpretations is beyond the scope of this paper, there are a number of comments worth making at this point. First, whether or not an interpretation is supported by a work depends, at least in part, on textual evidence; as a result, since evidence comes in degrees, so too does support. Second, given that the concern is with whether an interpretation is supported by a work and not a text, a supported interpretation needs to be compatible with the criteria of identity for the work; insofar as works 'take their identity from the circumstances of their creation', this requires interpreting works in light of the linguistic/symbolic and art-historical contexts of their creation.[15]

One might worry that talk of supported interpretations is incommensurate with the kind of pluralism on offer here; after all, it seems to favour interpretive projects that yield supported interpretations over those that do not. But even if it does favour such projects, in cases in which multiple interpretations are supported by a work, different projects can still identify different supported interpretations as correct: an intentionalist project, for example, would identify any supported interpretation compatible with the creator's intentions as

[15] Stephen Davies, 'Authors' Intentions, Literary Interpretation, and Literary Value', *British Journal of Aesthetics* 46 (2006): 224.

the correct one, whereas a value-maximizing project would identify the supported interpretation that makes the work most valuable as the correct one.[16] As a result, it is worth emphasizing that the notion of a supported interpretation is distinct from that of a correct interpretation. One might, after all, endorse a picture according to which there are multiple supported interpretations of a work, only one of which is correct. So, for example, one might argue that interpretations of *Withnail and I* according to which the title character, Withnail, is in love with the narrator Marwood and interpretations according to which he is not are both supported by the film.[17] But insofar as one endorsed a value-maximizing approach to interpretation, one might insist that the former is correct; and insofar as one endorsed an intentionalist approach to interpretation, one might insist that the latter is correct.[18] (And, of course, it would be incumbent upon the advocate of each approach to provide reasons for supporting one interpretive project over the other.) Alternatively, one might accept the notion of a supported interpretation while eschewing the notion of correctness altogether.

The point of introducing the distinction between supported and unsupported interpretations is to make room for the analogue of unsuccessful action in the discussion of authorial responsibility. When an action has been successfully performed, we can evaluate the agent in terms of her degree of responsibility for it. But if an action is unsuccessful, we can evaluate an agent only for having tried to perform the action and not for in fact having done so. Similarly, if an author produces a work which supports a given interpretation, we can evaluate her degree of responsibility for the corresponding interpreted-work. But if she produces a work that does not support her intended interpretation of it, then we can again determine her degree of responsibility only for the work she tried to produce and not the work she in fact produced.

II.3. INTERPRETATIONS AND RESPONSIBILITY

Whether or not fictional works are themselves in fact created by a process of interpretation, as certain interpretive constructivists would have it, the objects of evaluation under consideration here – interpreted-works – are so created.[19] And our question is what are the conditions under which the author of the work that has been interpreted is responsible for the interpreted-work (which, let us recall, is distinct from the work per se), in the sense that the value or disvalue of

16 Carroll, 'Interpretation and Intention', 76.

17 Bruce Robinson, dir., *Withnail and I* (1987; Los Angeles: Anchor Bay, 2001), DVD.

18 Alward, 'Rogue or Lover'.

19 See Michael Krausz, 'Interpretation and Its "Metaphysical" Entanglements', *Metaphilosophy* 31 (2000): 125–47.

the interpreted-work justifies praise or blame for the author herself. For present purposes, we will focus on moral and aesthetic value. So, for example, an interpreted-work might be morally valuable or disvaluable in virtue of containing praiseworthy or reprehensible themes. And an interpreted-work might be aesthetically valuable or disvaluable in virtue of containing elegant or crude characterizations or storylines. In addition, the focus is going to be interpreted-works whose generating interpretation is highly supported by the text. At the end of this section we will briefly consider interpreted-works generated by poorly supported or unsupported interpretations.

Let's consider a particular interpreted-work, *I-W*, generated by a highly supported interpretation, *I*. As above, there are three cases of interest: first, the author of the work (or text, or pattern of inscriptions, and so forth), *W*, intended that it be interpreted in this way;[20] second, the author neither intended that *W* be interpreted in this way nor was she in any position to expect that it would be; and third, although the author did not intend that *W* be interpreted as per *I*, she foresaw – or should have foreseen – that there was a good chance it would be. In the first case, the author is fully responsible for *I-W*. As a result, if *I-W* is positively morally or aesthetically valuable, then praising her for it would be justified. And if it is morally or aesthetically of negative value, then blaming her for it would be justified. In the second case, from the point of view of the author's compositional actions, the existence of *I-W* is entirely accidental and, hence, she has no degree of responsibility for it. As a result, praising the author for it, if it is positively valuable, or blaming her for it, if it is negatively valuable, is entirely unjustified. Finally, in the third case, the author is partially responsible for *I-W*. As a result, praising the author for it, if it is positively valuable, or blaming her for it, if it is negatively valuable, is justified but only insofar as the degree of praise or blame is proportional to her degree of responsibility. Suppose, for example, a fictional work supports an interpretation according to which it endorses racist themes. If the author intended her work to have racist themes, she is wholly blameworthy for this moral flaw in her work when interpreted in this way. But if the author did not foresee that her work would support such an interpretation – nor could she have reasonably been expected to have done so – then she is blameless for its moral failing so interpreted. And if she foresaw this interpretation, but did not intend it, then she is partially blameworthy for its moral flaw.

[20] I am assuming that this is equivalent to there being a correspondence of some kind between I and the author's meaning-intentions. Of course, it is possible for an author to intend that her work be interpreted in a way that deviates from her meaning intentions. For present purposes, however, I am just going to disregard this possibility.

One might worry that this conclusion could be far more simply established by just pointing that an author is not responsible for meanings erroneously ascribed to her work by incorrect interpretations. Now, as above, I eschew talk of correct or incorrect interpretations. But even if one endorses such talk, as well as the suggested link between responsibility and correctness, the argument on offer here might still be in order. After all, rather than explaining responsibility in terms of correctness, one might instead explain correctness in terms of responsibility. That is to say, one could accept the account of responsibility on offer and, using the putative link between responsibility and correctness, conclude that an interpretation is correct when it is both supported and either intended or foreseen by the author. Whether or not this would count as a tenable interest-independent account of correctness is of course another matter.

Let us turn now to cases in which the interpretation, *I*, is unsupported or poorly supported by *W*. My inclination is to treat this as an unsuccessful action. Although one might reasonably praise or blame the author's attempt to produce a work that would be interpreted in this way – if that were her intention – this praise or blame would not stem from her responsibility for *I*-*W*. After all, she did not create a work that supported any such interpretation. Now, of course, someone might go ahead and interpret *W* in this way despite the fact that *I* is unsupported by *W*. And, in such circumstances, the author's act of producing *W* would play a causal role in the production of *I*-*W*. But although this might render the author partly causally responsible for *I*-*W*, causal responsibility is not sufficient for moral responsibility; and it is the latter that is at issue when questions of praise and blame are at stake.

II.4. INTENTIONS AND INTERPRETIVE ROLES

Our central question here is not under what conditions authors are responsible for interpretations of the fictional works they compose; instead the question is what interpretive projects certain appreciators – in particular, those concerned with the attribution of responsibility – ought to engage in. The first thing to note is that, given how the matter has been formulated here, such appreciators should engage only in interpretive projects whose outputs are supported interpretations. In effect, whatever other criteria they bring to bear, appreciators concerned to attribute responsibility are forced to choose interpretations from among the class of those that are supported by the text. In particular, these are interpretive projects which prioritize textual evidence over extra-textual evidence: insofar as permissible extra-textual resources point towards interpretations that conflict with interpretations supported by the text, the latter considerations trump the former. As a result, radical intentionalist interpretive projects – in which

authors' meaning-intentions trump textual evidence – fail to meet the interpretive interests of the appreciators at issue here.[21]

The remaining question is what positive roles authors' intentions and other psychological states should play in an interpretive project designed to meet the interests of such appreciators. Two roles are of interest here. First, minimally they should serve as a criterion of success for whatever interpretive project such appreciators engage in. For appreciators with an interest in attributing responsibility to authors, an interpretive project is successful only insofar as it yields interpretations for which the author in question is in fact responsible. And, as above, an author is responsible for only those interpretations of her work that she intended or foresaw. As a result, an interpretive project that meets the interests of such appreciators is successful only if it generates interpretations authors intend or foresee. As it stands, however, whether or not an interpretation is successful in this sense is pure happenstance. Any project is successful as long as it generates an intended or foreseen interpretation, but nothing need have been built into the project that makes this result likely at all. Consequently, a second role for authors' psychological states is as extra-textual resources that serve as inputs into the relevant interpretive algorithm. As above, in order to yield supported interpretations, the contribution of any extra-textual resources has to be commensurate with the primacy of the text. But a project that includes evidence of authors' meaning intentions and what interpretations of their work they foresaw (or should have foreseen) is much more likely to yield interpretations for which they are responsible. As a result, interpretive projects of this kind are correspondingly likely to satisfy the interpretive interests of appreciators concerned to attribute responsibility to authors.

In effect, what we have here is a limited defence of a kind of moderate intentionalism.[22] The interpretive projects defended here are intentionalist because of the roles played in them by authors' intentions (and other psychological states) – both as extra-textual resources and as criteria of success. But these projects count as versions of moderate intentionalism – rather than strong or radical intentionalism – because of their commitment to the primacy of text. The defence of moderate intentionalism is a limited one because it only applies to appreciators with an interest in attributing responsibility to authors. Insofar as appreciators lack interpretive interests of this kind, no grounds have been offered for thinking that they should engage in some such interpretive project.

---

21 See Hirsch, *Validity in Interpretation*.

22 See Robert Stecker, 'Moderate Actual Intentionalism Defended', *Journal of Aesthetics and Art Criticism* 64 (2006): 429–38.

III. THE ACCIDENTAL AUTHOR

The account of authorial responsibility on offer here makes room for the heretofore underdeveloped notion of accidental authorship. Just as an action can be accidental in the sense of being neither intended nor foreseen, so too can the products of authorial compositional activity. But, as above, in order for something that an agent has done to count as an accidental action, it needs to be performed by means of an action that is performed intentionally, despite not being performed intentionally itself. So, in the example considered above, Mary accidentally performed the action of dropping Fred into the fire by means of intentionally pushing the relevant lever forward, when she did so without intending or foreseeing that she would thereby drop Fred into the fire. Analogously, with regard to composition, what the author creates intentionally is the (uninterpreted) work itself.[23] And by means of so doing she might accidentally create an interpreted-work – a work-interpretation pair – if her work ends up being interpreted in a way she neither intended nor foresaw it would be. Three different sorts of accidental authorship are of interest here: cases in which the interpretation is supported by the work; cases in which the interpretation is supported by the text, but not by the work itself; and cases in which the interpretation is supported neither by the work nor by its constituent text. I will consider each in turn. Before doing so, however, a terminological note may be in order. One might worry that cases in which an interpretation is not supported by a work do not involve any kind of authorship and, hence, talk of 'accidental authorship' in such cases is out of order. The model here, however, is accidental action. And to deny that what I am calling 'accidental authorship' is any kind of authorship is akin to denying that what we all call 'accidental action' is any kind of action. But if I cause you injury without being at all responsible for so doing, we would say that I injured you accidentally rather than saying that I did not injure you at all. Similarly, if a work I author is interpreted in a way for which I am not responsible, it is reasonable to describe this as a case of my having authored the interpreted-work accidentally rather than not having authored it at all.

Consider, first, the film *Withnail and I*, which tells the story of a pair of out-of-work actors – Withnail and Marwood – living in Camden Town in the late 1960s.[24] The film was written and directed by Bruce Robinson and was intended to be

[23] One might, of course, reasonably or unreasonably insist that a work is inevitably interpreted by the author herself. My own inclination is to suppose that while an author may inevitably intend her work to have a certain meaning, having (first-person) meaning-intentions for a work is distinct from having a (third-person) interpretation of it. Nevertheless, nothing of import hangs on this point.

[24] Robinson, *Withnail and I*.

a semi-autobiographical account of his and his friend Vivian MacKerrell's own experiences in Camden Town at that time. According to one interpretation of this film, the character Withnail is secretly in love with the narrator Marwood. But this is an interpretation that Robinson presumably neither intended nor foresaw: after all, he intended the Marwood character to be based on himself and the Withnail character on Vivian MacKerrell, and there is no evidence that he believed MacKerrell to be secretly in love with him. But this interpretation is supported by the work: it makes sense of a number of episodes in the film in which the topic of homosexuality plays a central role.[25] As a result, this counts as a case of accidental authorship in which the interpretation in question is supported by the work. Moreover, the film is arguably better on this accidental interpretation: it yields a more unified plot and avoids the charge of homophobic portrayals of gay characters.[26] As a result, insofar as it is interpreted in this way, we have an instance of what might be called 'aesthetic good luck': Robinson accidentally created an interpreted-work that was aesthetically better than what he had intended to create and, to that extent, he got aesthetically lucky.

Second, consider Ed Wood's notorious film *Plan 9 from Outer Space*.[27] It could be – and, perhaps, sometimes is – interpreted as a postmodern send-up of a science fiction film despite the fact that Wood intended it as a serious genre film. Now according to Stephen Davies, works 'take their identity from the circumstances of their creation'.[28] And if this is right, any interpretation that fails to situate a work in the art-historical and linguistic/symbolic context of its creation is one that can only be supported by the text that constitutes the work and not by the work itself. As a result, insofar as a work's identity-conferring circumstances of creation includes the author's categorical intentions – intentions to produce a work of a certain kind or category – the postmodern interpretation in question is not supported by the film *Plan 9 from Outer Space*. After all, it conflicts with Wood's categorical intention to produce a serious science fiction film. Nevertheless, the postmodern interpretation is supported by the filmic text, even if not by the film itself.

Now, what is important to note here is that we arguably have two objects of evaluation accidentally produced by Wood – the interpreted-work and the interpreted-text – each of which he created by means of creating *Plan 9 from Outer Space* (without intending or foreseeing that it would be given a postmodern interpretation). Moreover, because the postmodern interpretation is not

25 Alward, 'Rogue or Lover'.
26 Ibid.
27 Ed Wood, dir., *Plan 9 from Outer Space* (1959; Los Angeles: Image Entertainment, 2000), DVD.
28 Davies, 'Authors' Intentions', 224.

supported by *Plan 9 from Outer Space*, the interpreted-work consisting of the film and the postmodern interpretation is flawed aesthetically. After all, one dimension of the aesthetic value of a representational artwork consists in how its content is embodied in the work; and insofar as its meaning is captured by an unsupported interpretation of it, it is not embodied in the work at all. But because the postmodern interpretation is supported by the filmic text of *Plan 9 from Outer Space*, the interpreted-text is not flawed in this way. As a result, although accidentally creating the interpreted-work might count as aesthetic bad luck, accidentally creating the corresponding interpreted-text could still count as aesthetic good luck.

Finally, consider Bruce Springsteen's song 'Born in the U.S.A.' The song contains lyrics which express a very negative view of the United States in general and its treatment of veterans in particular, such as the following:

> Born down in a dead man's town
> The first kick I took was when I hit the ground
> End up like a dog that's been beat too much
> Till you spend half your life just covering up.[29]

Despite this fact, it has been widely interpreted as a patriotic, pro-American anthem. But as should be clear, this interpretation is simply not supported by the text of the song, let alone the song itself. As before, we can take both the interpreted-work and the interpreted-text to have been accidentally created by Springsteen by means of writing the song. It is, after all, unlikely that he either intended or foresaw that it would be interpreted in a way that explicitly contradicts the lyrics he wrote. But in this case, because the interpretation is unsupported by either the work or the text, both the interpreted-work and the interpreted-text are aesthetically flawed. The upshot is that while an author can accidentally create any number of interpreted-works or interpreted-texts, unless the interpretations in question are minimally supported by the text, this can only count as bad aesthetic luck.

Peter Alward
Department of Philosophy, University of Saskatchewan,
9 Campus Dr, Saskatoon, SK, S7N 5A5 Canada
peter.wb.alward@gmail.com

[29] Bruce Springsteen, 'Born in the U.S.A.', *Born in the U.S.A.* (1984; New York: Columbia Records, 2015), CD.

BIBLIOGRAPHY

Alward, Peter. 'Butter Knives and Screwdrivers: An Intentionalist Defense of Radical Constructivism.' *Journal of Aesthetics and Art Criticism* 72 (2014): 247–60.

–––––––. 'Rogue or Lover: Value Maximizing Interpretations of *Withnail and I*.' *Projections* 12 (2018): 39–54.

Beardsley, Monroe C. 'Intentions and Interpretations: A Fallacy Revived.' In *Aesthetics and the Philosophy of Art: The Analytic Tradition*, edited by Peter Lamarque and Stein Haugom Olsen, 189–99. Oxford: Blackwell, 2004.

Carroll, Noël. 'Art, Intention, and Conversation.' In *Intention and Interpretation*, ed. Gary Iseminger, 97–131. Philadelphia: Temple University Press, 1992.

–––––––. 'Interpretation and Intention: The Debate between Hypothetical and Actual Intentionalism.' *Metaphilosophy* 31 (2000): 75–95.

Currie, Gregory. *The Nature of Fiction*. Cambridge: Cambridge University Press, 1990.

Davidson, Donald. 'Actions, Reasons, and Causes.' *Journal of Philosophy* 60 (1963): 685–700.

Davies, Stephen. 'Authors' Intentions, Literary Interpretation, and Literary Value.' *British Journal of Aesthetics* 46 (2006): 223–47.

Hirsch, E. D. *Validity in Interpretation*. New Haven, CT: Yale University Press, 1967.

Krausz, Michael. 'Interpretation and Its "Metaphysical" Entanglements.' *Metaphilosophy* 31 (2000): 125–47.

Levinson, Jerrold. 'Intention and Interpretation in Literature.' In *The Pleasures of Aesthetics: Philosophical Essays*, 175–213. Ithaca, NY: Cornell University Press, 1996.

Nannicelli, Ted. 'Ethical Criticism and the Interpretation of Art.' *Journal of Aesthetics and Art Criticism* 75 (2017): 401–13.

Parker, Richard. 'Blame, Punishment, and the Role of Result.' *American Philosophical Quarterly* 21 (1984): 269–76.

Robinson, Bruce, dir. *Withnail and I*. 1987. Los Angeles: Anchor Bay, 2001. DVD.

Springsteen, Bruce. 'Born in the U.S.A.' *Born in the U.S.A*. 1984. New York: Columbia Records, 2015. CD.

Stecker, Robert. 'The Constructivist's Dilemma.' *Journal of Aesthetics and Art Criticism* 55 (1997): 43–51.

–––––––. *Interpretation and Construction: Art, Speech, and the Law*. Oxford: Blackwell, 2003.

–––––––. 'Moderate Actual Intentionalism Defended.' *Journal of Aesthetics and Art Criticism* 64 (2006): 429–38.

Walton, Kendall. *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Cambridge, MA: Harvard University Press, 1990.

Wood, Ed, dir. *Plan 9 from Outer Space*. 1959. Los Angeles: Image Entertainment, 2000. DVD.