Imagination and Emotion in Action

By Jason Robert D’Cruz

B.A. Philosophy (with distinction) Yale University, 2000

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Program in Philosophy at Brown University.

May, 2009
This dissertation by Jason Robert D’Cruz is accepted in its present form by the
Department of Philosophy as satisfying the dissertation requirement of Doctor of
Philosophy.

Date ____________________________

Professor Nomy Arpaly, Advisor

Recommended to the Graduate Council

Date ____________________________

Professor Jaegwon Kim, Reader

Date ____________________________

Professor James Dreier, Reader

Approved by the Graduate Council

Date ____________________________

Sheila Bonde, Dean of the Graduate School
Curriculum Vitae

Jason Robert D’Cruz was born on September 10, 1977 in Ottawa, Canada. After graduating from Saint Michael’s College School in Toronto, he attended Yale University and completed his studies in May 2000 with a B.A. in Philosophy (with Distinction). From 2000-2001 he taught English language and American culture at the Zhejiang Institute of Science and Technology in Hangzhou, China. He commenced graduate studies at Brown University in September 2002. While a graduate student at Brown, he worked as a Teaching Assistant for Professor Bernard Reginster’s course “Existentialism” and Professor Christopher Hill’s course “Epistemology”. He designed and taught “Freedom and Responsibility” as well as “Contemporary Moral Problems”. He currently teaches the course “Deception and Self-Deception” at the Harvard Writing Program. In September 2009 he will commence work as an Assistant Professor in the Philosophy Department at the State University of New York at Albany.
Preface and Acknowledgements

This work could not have been accomplished without the generous help of numerous collaborators and supporters. The original idea for the dissertation was spawned in a particularly stimulating graduate seminar on action theory led by Jaegwon Kim. My fellow graduate students, including the participants in that seminar, provided a wonderful sounding board for my nascent ideas on action and imagination. I had many thought-provoking discussions into the wee hours of the morning with Ali Eslami, Derek Ettinger, Simon Feldman, Allen Hazlett, Lillian O’Brien, Max Pines, Vicky Prickett, Nick Treanor, and with my brother, Andrew D’Cruz.

I received valuable feedback on the material from Chapters One and Two from the job search committee at the University of Miami, and on the material from Chapter Four from the job search committee at the State University of New York at Albany. My readers, Jaegwon Kim and James Dreier, were always available to provide guidance and constructive criticism. My advisor, Nomy Arpaly, showed me by example that ethics could be a creative, subtle, and rigorous enterprise. I credit her with my reinvention as a moral psychologist.

My mother, Filomena D’Cruz, and my brother, Andrew D’Cruz helped me with the copy-editing of the text. My brother, Geoffrey D’Cruz, helped to format the manuscript. My father, Joseph D’Cruz, who defended his dissertation thirty years ago in December of 1978 at Harvard, has been both an example and a mentor. I could not have even
embarked upon this enterprise without the nurturing and selfless support of my family members.

Last but certainly not least, my partner, Shadi Khadivi was unstinting in her optimism and encouragement, even when I was at my most discouraged and despondent.
# Table of Contents

Chapter One:  
The persistent recalcitrance of “arational” actions to explanation and rationalization  
Page 1

Chapter 2:  
Heterogeneous Action Explanation  
Page 25

Chapter 3:  
Cognitive Theories and Desire-like Imaginings  
Page 47

Chapter Four:  
Reality-Directed and Fiction-Directed Emotions  
Page 76

Chapter 5:  
Imaginative Resistance and Capitulation  
Page 111

Works Cited  
Page 143
CHAPTER ONE:

THE PERSISTENT RECALCITRANCE OF “ARATIONAL” ACTIONS TO EXPLANATION AND RATIONALIZATION
In “Arational Actions” (1991) Rosalind Hursthouse describes a subset of intentional actions that elude Davidsonian action explanation since, she argues, they cannot be explained in terms of the agent who performs the action possessing a particular belief/desire pair. Instead, she maintains, these actions can only be explained by the adverting to the fact that the agent is in the grip of some particular emotion. Below is a list of Hursthouse’s examples:

(a) explained by a wave of love, affection, or tenderness – kissing or lightly touching in passing, seizing and tossing up in the air, rumpling the hair of, or generally messing up the person or animal one loves; talking to her photograph as one passes, kissing it;
(b) explained by anger, hatred, and sometimes jealously – violently destroying or damaging anything remotely connected with the person (or animal, or institution) one’s emotion is directed towards, e.g., her picture, letters or presents from her, awards from her, books or poems about her; the chair she was wont to sit in, locks of her hair, recordings of “our” song, etc.;
(c) explained by anger with inanimate objects – doing things that might make sense if the things were animate, e.g., shouting at them, throwing an “uncooperative” tin opener on the ground or out the window, kicking doors that refuse to shut and cars that refuse to start, tying towels that keep falling off a slippery towel rail on to it very tightly and then consolidating
the knots with water; muttering vindictively ‘I’ll show you’, or ‘You
would would you’;

(d) explained by excitement – jumping up and down, running, shouting,
pounding the table or one’s knees, hugging oneself or other people,
throwing things.

(e) explained by joy – jumping up and down, running, shouting, pounding
the table or one’s knees, hugging oneself or other people, throwing things;

(f) explained by grief – tearing one’s hair or clothes, caressing, clutching,
even rolling in, anything suitable associated with the person or thing that
is the object of grief, e.g., pictures, clothes, presents from her (cf. anger
above). (The example of rolling in comes from a novel in which a man
takes his dead wife’s clothes out of the wardrobe, puts them on the bed
and rolls in them, burying his face in them and rubbing them against his
cheeks);

(g) explained by shame – covering one’s face in the dark, or when one is
alone; washing with violent attention to scrubbing and scouring;

(h) explained by horror – covering one’s face in the dark, or when one is
alone; washing with violent attention to scrubbing and scouring;

(i) explained by fear – hiding one’s face, burring under the bed clothes

(j) explained by feeling proud, or self-satisfied, or pleased with oneself –
talking to or posturing to oneself in the mirror. (58)

Hursthouse maintains that all of the above actions share the following common marks:
1. The action is intentional.
2. The agent did not perform the action for a reason (in the sense that there can be no ascription of a suitable instrumental belief).
3. The fact that the agent was in the grip of a particular emotion explains the action (59).

Hursthouse’s project of emotion-based action explanation is a direct and radical challenge to the Davidsonian, who aims to explain actions purely in terms of beliefs and desires. In his landmark 1963 paper, “Actions, Reasons, and Causes” Davidson introduced the idea of a “primary reason”. A primary reason consists a “pro attitude” (a generic term invented to describe desires, fancies, yens, etc.) toward actions with a particular property P (e.g. bringing about a certain result or state of affairs), and an instrumental belief that a particular action, A, has that property. According to Davidson, actions have these mental antecedents as causal antecedents. Using his exact phrase, “a primary reason for an action is its cause” (12).

In this chapter I argue that:
1. Emotion-based explanations are insufficient to explain the particularity of arational actions;
2. Humean explanations of arational actions of the sort the Michael Smith offers are trivially true and therefore not explanatory;
3. The simple combination of a Humean explanation and an emotion based explanation is also not adequately explanatory.

Davidson’s broadly Humean theory of action is appealing for a number of reasons. First, the existence of primary reasons both distinguishes an action from a mere movement and also provides the causal explanation for the action. In this way, the relationship between the justificatory and the explanatory role of reasons is made manifest. The belief-desire complex is not just the causal antecedent of the action; it also provides the material for a reason-giving explanation of the action, rendering the agent’s behavior intelligible. The structure of such an explanation can be made apparent by presenting it as a form of practical reasoning: the constituents of the primary reason, belief and desire, provide the premises, and the action is the conclusion. In this way, the gap is narrowed between the normative realm of reason explanations (he flipped the switch in order to illuminate the

---

While I argue that both of these causal theories are insufficient to give a proper explanation of arational action, I do not deny the causal theory more generally. In fact, I affirm it. I take any causal theory of action to include the following three elements:

1. What distinguishes actions from mere movements of the body is that actions are part of a specific kind of causal sequence.
2. Intentional states place a key role in this causal sequence.
3. Action explanation is causal explanation.

In Chapters 2 and 3 I present my own causal theory of arational actions.
room) and the natural realm of causal explanation (his belief and desire caused him to flip the switch).²

In “Actions, Reasons, and Causes” Davidson argues that any action that is described in a way such that we cannot give it an explanation in terms of primary reasons can be redescribed in a way such that we can. In his famous example, Fred inadvertently alerts a prowler by turning on a light switch, yet, by stipulation, he has no desire to alert the prowler. Davidson argues that this action can be re-described as follows: the man desires to illuminate the room, and believes he can do so by flipping the switch. In giving this explanation, we have already given an explanation of Fred alerting the prowler, since Fred’s alerting the prowler is his moving his finger, and he moves his finger because he wants to illuminate the room.

According to Hursthouse, the intentional actions she describes admit of no plausible re-description such that they are explicable in terms of Davidsonian belief/desire pairs. If she is correct, the explanatory power of Davidson’s theory would be considerably diminished.

² Because of his thesis of the “anomalism of the mental” Davidson himself was committed to the claim there could be no empirical causal laws employing intentional vocabulary. Contemporary philosophers who have aligned themselves with “causalism” have most often jettisoned anomalist monism to espouse a stronger thesis whereby aspects of the content of intentional states are relevant to giving a causal explanation of an action (see Dretske 1991).
I. Emotion-based Explanations.

According to Hursthouse, arational actions elude Davidsonian explanation. Consider some of the paradigm arational actions Hursthouse describes: Jane scratches out the eyes of photo of Joan (explained by anger); a widower rolls around in his dead wife’s clothes (explained by grief); one shouts at an uncooperative tin opener (explained by rage), covers one’s face in the dark out of shame (explained by shame), burrows under the bedclothes (explained by fear). The actions that Hursthouse describes do not seem to be means by which the agent realizes her goals. The commonality in this curious class of (putatively) intentional actions is their seeming pointlessness. Accordingly, it is difficult to identify any appropriate desire and instrumental belief to explain the action. If the action has no obvious end, how can the agent have an instrumental belief about how to achieve that end? Nonetheless, these actions still appear to be intentional: they are certainly not “mere movements” of the body, like the beating of the heart or the functioning of the digestive system. The are things we do, not mere events in the histories of our bodies.

One strategy that a Davidsonian might advert to in order to bring Hursthouse’s recalcitrant cases under the umbrella of belief-desire explanation is to explain them in terms of a desire for self-expression. Candidate explanations might go something like this:
The widower has a deep and inarticulate desire to express his grief. He believes that by rolling around in his dead wife’s clothes he will achieve this self-expression.

Or consider:

The man who screams at and flings the tin-opener desires to express his frustration. He believes that by screaming at and flinging the tin opener he will be able to express this frustration.

Finally:

Jane, who scratches out the eyes of the photo of Joan, desires to express her rage. She believes that by scratching the photo she will express her emotion.

By imputing to these agents a desire for self-expression, we can understand their action as a means to bringing something about (i.e. – self-expression), and we can thereby also impute to them an instrumental belief. Moreover, the desire for self-expression is pervasive and intelligible. However, this explanatory strategy runs into problems. As Hursthouse points out, while the desire for self-expression is prevalent, it is simply not plausible that this desire is present in all instances of arational action:

[W]hat is wrong with this suggestion once again is that it involves ascribing a belief to the agent which should not be ascribed. If I ø in order to express or relieve my emotion, I do so in the belief that my ø-ing will indeed have (or is likely to have) this upshot. And in such
cases, there is the possibility that I am not setting about fulfilling my intention in the right way; that I am open to correction. But arational actions would not usually admit of any possibility of mistake in this way; they are not the sort of action an agent would usually do in the (possibly erroneous) belief that they would achieve this effect. Nor should we accept it as obvious that in every case the agent has the desire to express this emotion, a desire whose content is distinct from that of the desire to, say, throw the tin opener violently on the floor. The ascription of this extra desire requires extra justification. (61)

For Hursthouse’s objection to stand, all we have to grant her is that in some cases where an agent acts out of intense emotion there is no concomitant desire to express that emotion. This does not seem too much to grant, since a desire for self expression is most often absent from the phenomenal feel of impassioned arational actions.

A second strategy that a Davidsonian could deploy to account for Hursthouse’s agents’ seemingly pointless action is by imputing to them a false belief. In this way, the Humean can accept that their action is (objectively) futile, and still explain their action by reference to a desire and an instrumental belief within the agent’s psychological economy. The benefit of this strategy is that it allows us to see clearly how arational action is intentional (it is done (misguidedly) for an end), and it provides us with an illuminating explanation.
However, it seems prima facie implausible that the agents that Hursthouse describes have within their psychological economies the necessary kind of false beliefs. Does the man who rolls around in his wife’s dead clothes believe that he rolling around with his wife? Does the person who screams at the tin opener and savagely flings it to the ground believe that he is injuring the tin opener? Does the person who is afraid and hides his face under the pillow believe that this is a reasonable strategy to keep himself safe? It is sometimes true that when people find themselves in the grip of manic joy, acute anxiety, trauma, infatuation, stress, or strong emotion more generally, the reliability of their belief-forming mechanisms is diminished. In a fit of mania, a person might form the belief that he can max out his credit cards and still avoid bankruptcy; in a state of extreme fear, a hostage might form the belief that his captor is his savior; in a state of romantic infatuation, a person might form the belief that his lover is morally without fault. In each of these cases, however, the irrational belief in question has a stability and a longevity that is absent from the irrational beliefs we would have to attribute to the agents in Hursthouse’s cases to “rationalize” their actions. These beliefs do not show up before or after the target action. Hursthouse (rightly, I think) vigorously rejects this move: “If […] lunatic beliefs are ascribed, these will show up nowhere else in behavior, be sincerely and vigorously repudiated by the agent, and that agent’s momentary acquisition of them will, in turn, be utterly mysterious.” (64) The agents that Hursthouse describes, although they may be described as impassioned, are not lunatics and do not suffer from “temporary insanity”.

3 Of course, this is not always true. Sometimes, for example, the fear and anxiety brought on by a “high-stakes” decision serve as an antidote to self-deception.
A third strategy we may resort to in order to rescue the belief-desire account of intentional action is to deny that Hursthouse’s agents are acting intentionally. We may want to assimilate the violent outbursts and impassioned behaviour she describes to nonintentional physiological behaviour, such as trembling or blushing. Many expressive bodily movements are not actions. Consider, for example, a tremor of fear in one’s voice, or uncontrollable laughter. Many facial movements can be expressive non-actions: smiling, grimacing, widening one’s eye or raising one’s eyebrows in awe or surprise, or contorting one’s face in fear are good examples. This is not to say that smiling or grimacing are always non-actions. They can sometimes be produced intentionally, in order to elicit some reaction from another person or even from oneself. People who adhere to the “fake till you make it” school of self-improvement have been known to laugh deliberately in order to make themselves happy by means of auto-suggestion. A father might raise his eyebrow in order to communicate his disapproval to his naughty son. “Fake” laughter and smiles are often used in order to flatter, seduce, or put others at ease. However, the fact that it is possible for us to convey our emotions via these expressions does not automatically imply that individual is doing so intentionally. Sometimes we intend to communicate via our bodily expressions, and sometimes our bodily expressions convey information utterly independent of (and sometimes contrary to) our intentions. We cannot always tell the difference, even in ourselves, but that does not mean that there is no difference⁴.

---

⁴ In some cases this difference is marked physiologically. For example, a genuine “Duchenne” smile involves distinct muscles which we cannot directly try to move (Goldie 34). This accounts for our ability to tell “fake” smiles from “real” ones.
The behaviour of the agents Hursthouse describes does not fit well into either the category of “self-expressive non-actions” or the category of “actions performed with the aim of self-expression”. The deliberateness with which they act precludes the former interpretation, and the spontaneity with which they act precludes the latter. This is not to say that the sorts of actions that Hursthouse describes are never done with the aim of expressing or “venting” emotion. Sometimes, they may be, as when one deliberately tears at one’s “anger pillow”. But to attribute this aim to all of the agents in her cases does not seem justifiable.

On the other hand, the Davidsonian might want to characterized Hursthouse’s agents as utterly aimless, and thereby deny that they perform intentional actions at all. The pressure to consider arational behavior as non-intentional might come from the intuition that all intentional action is done for an end whereas in the examples Hursthouse describes, the behavior seems pointless. However, a problem with denying that Hursthouse’s agents act intentionally is that their action appears so deliberate. The man speaks to a tin for a sustained period, threatens it, and the flings it. As Joan scratches out the eyes of photo, she explains to Jane why she deserves this.

The way in which these actions are deliberate is brought into focus if we consider what would happen if the agent in question realized that her action was somehow going wrong. What would happen if Joan realized that the photo she was scratching was in fact a photo of her loyal friend Janice, rather than her nemesis, Joan? Surely she would stop her

---

5 This point was put to me by Lilian O'Brien.
scratching immediately. Perhaps she would try to repair the photo, or at least cradle it in her hands. Perhaps she would rifle through her albums to find a proper picture of Jane so that she could “punish” her. These compensatory actions are some indication that the initial action was not altogether pointless. Even if we are not sure what they aiming at, it seems that they aim at *something*.

Moreover, there is a certain appropriateness to these actions as well. Had Joan deliberately picked out a photo Janice in order to punish Jane, we would be rather surprised. It makes sense that she would have picked out a photo of Jane. If the man who flung the can opener had instead gently kissed the recalcitrant can opener, we would be similarly puzzled. It makes sense that he would have flung it.

The “appropriateness” and “normativity” of arational actions together with their deliberateness render dubious the claim that arational agents do not act intentionally. As Hursthouse points out, they are certainly not acting unintentionally. However, I think that this very “appropriateness” which makes it undeniable that arational actions are intentional also makes it clear that Hursthouse’s emotion-based explanation is inadequate.

What makes this mode of explanation inadequate is that emotion-based explanations do not explain why an agent performs *this peculiar action*. The one-word explanation – “anger” – does not contain enough content to explain Joan’s action when she scratches out Jane’s eyes in a photograph. It is simply too generic. While I agree that emotions
should play some role in the explanation of the action, merely identifying the relevant emotion does not seem to me sufficient.

One way to bring the “too generic” objection into relief is to notice that, unlike belief/desire explanations, Hursthouse’s explanation fails to rule out a large contrast class of similar actions that the agent could have performed but rationally ought not to. The generic “grief” explanation does not explain why it would be inappropriate for the grieving widower to roll around in his young lover’s clothes. There is a clear in sense in which the agents in question rationally ought not to perform certain variations on their action, and Hursthouse’s emotion-based action explanations fail to capture this.

This is not to say that I want to deny that emotions play some role in the explanation of action. For example, my fear of a bear in the woods might cause me to run for safety. But notice that in this example, we can still ascribe a belief-desire pair that rationalizes, and ultimately causes, the action. We might explain the action by ascribing to me a desire to be safe from the bear, and the instrumental belief that by running away, I will secure my safety. The emotion of “fear” plays a role in explaining why I form the desire that I do. Moreover, emotions can play a role in explaining both “hot” and “cold” actions. For example, in explaining why I pour cyanide into Vikram’s coffee, we might advert to something like my simmering hatred for and jealousy of Vikram. But again, the

---

6 I do not mean to commit myself to the claim that Hursthouse’s agents are acting on good reasons. I merely mean to point out that there are rationalizing emaplanation in the offing.

7 This is not to say that their actions are all-things-considered rational; it is just to say that there is some room for normative appraisal.
emotion is explanatory insofar as it illuminates why I have the desire to harm Vikram in the first place.

Indeed, emotions are not the only entities in our mental economy that play a role in explaining why we act in the ways that we do. Mood, depression, mania, attention, mental associations, caffeine, and moral principles are only a few items in a very long list. None of these, however, is sufficient by itself to explain intentional action. But each of them can shed light on why an agent has a particular belief/desire pair that ultimately causes her action.

In *The Concept of Mind* (1949), Gilbert Ryle provided an analysis of the sentence “he boasted from vanity” which, like Hursthouse’s analyses, made no specific reference to beliefs and desires. Ryle analyzed this sentence into “he boasted on meeting the stranger and his doing so satisfies the lawlike proposition that whenever he finds a chance of securing the admiration and envy of others, he does whatever he thinks will produce this admiration and envy.” (89) The most obvious problem with Ryle’s analysis seems to be that just because someone boasts from vanity on some occasion does not mean that he will boast from vanity with any kind of regularity, and it certainly doesn’t mean that he will boast with a law-like regularity. Davidson argues further that Ryle’s account derives any explanatory force that it has from the primary reason “the boaster wanted to secure the admiration and envy of others, and he believed that his action would produce this

---

8 For a detailed discussion of the ways in which action can be caused and explained by a complex of mental states that are reason-responsive, merely content efficacious, and brute causal, see Arpaly 2006 pp. 69-72.
admiration and envy” (1980, 7). If the Ryle’s boaster no longer believed that boasting would elicit envy but rather mere contempt, presumably he would stop boasting. As we shall see in the next section, Davidson’s response to Ryle is closely analogous to Michael Smith’s Humean response to Hursthouse. Smith argues that while Hursthouse’s emotion-based explanations are in a sense illuminating, they always presuppose a standard belief-desire explanation. In the following section, I will argue that because of the complex nature of Hursthouse’s examples, Smith’s strategy will not work.
II. Humean Explanations

For Michael Smith, actions are explained by desires (which represent how the world is to be) and beliefs (which represent how the world is, and how it must be changed for relevant desire to be fulfilled). Actions are means by which the agent realizes her goals. Smith’s account requires that all intentional actions be explainable by a complex of cognitive and conative states, specifically, by two beliefs and one desire. The first belief ($b_1$) is an “orienting belief” that represents the initial state of world (e.g. – there is no gin in my glass), the desire (d) represents the final state (there is gin in my glass), and an instrumental belief ($b_2$) represents how one is to get from the initial state to the final state (I must pour myself some gin). Temporally, $b_1$, $b_2$, and d must all occur in the agent’s psychology prior to the action (a).

Smith describes desires as mental states that “represent how the world is to be” (19). Interestingly, if we accept Smith’s analysis of desire, it seems that agents who perform arational actions have none, since their actions do not aim at any goal. But this cannot be right! The agents that Hursthouse describes are positively brimming with desire. In fact, Hursthouse describes a category of desire states that are all too familiar. Moreover, the actions these agents perform seem intimately tied up with their desire. So the important question becomes, what is the best interpretation of what Hursthouse’s agents desire, how is their desire related to their action?

---

9 This, I hope, is a faithful reconstruction of Smith’s theory. See Smith 1987 and 1994.
Smith tries to provide an explanation for the arational actions Hursthouse describes by appealing to a “knowing how” instrumental belief on the part of the agent. He suggests the following analysis of the widower in Hursthouse’s example who rolls around in his dead wife’s clothes (which Hursthouse claims is best explained by the emotion of grief\(^\text{10}\)):

The man is doing what he is doing because he desires to roll around in his dead wife’s clothes and believes that he can do so by doing just what he is doing; that is, by rolling around in those particular clothes that he is rolling around in. To be sure, the explanation doesn’t mention grief. But it is most certainly an explanation we can give of what the man is doing because it correctly identifies the man’s reasons for action, in the sense of the belief/desire pair that produced his action, relatively bizarre though they may be. (22)

While somewhat trivial, Smith’s Humean explanation does do some explanatory work. We get a sense of this explanatory work when we consider the case where the agent’s belief is false\(^\text{11}\). To borrow a famous example from Bernard Williams\(^\text{12}\) (1981, 102), we have an explanation for why a man drinks a glass of petrol if we know that he (falsely)

\(^{10}\) Although I bracket the worry in this discussion, I suspect that the example of the grieving widower is one of Hursthouse’s least compelling, since the action seems plausibly explainable in terms of a non-trivial instrumental belief (the man desires to be reminded of his wife’s presence by touching and smelling her belongings, and he believes that he will be so reminded by rolling around in her clothes). For the purposes of this discussion, we must assume that this is not the explanation for the man’s action. A clearer example of an arational action would be if the man talked to his dead wife, fully aware that she could not hear him.

\(^{11}\) Peter Goldie (2000) makes this point in “Explaining Expressions of Emotion”.

\(^{12}\) Here I use Williams’ example for my own purposes, and not to illustrate the internal/external reason distinction he had in mind.
believes the glass contains gin and he desires to drink gin. Similarly, if the grieving
widower falsely believes that he is rolling around in his dead wife’s clothes, when in fact
he is rolling around in his new lover’s clothes, we still have a perfectly good explanation
of his action. The explanation of this action will involve the belief that the widower can
do what he wants to do – rolling around in his dead wife’s clothes – by doing what he is
doing – that is, by rolling around in the clothes that he thinks belong to his dead wife.
We should not overlook that Smith’s Humean explanation correctly predicts that if the
man falsely believed that his new lover’s clothes were his wife’s clothes, he would roll
around in them. And if he falsely believed that his wife’s clothes were his new lover’s
clothes, he would not roll around in them. Moreover, Smith’s explanation has the virtue
of avoiding the problem of being too generic in the way that Hursthouse’s account is.
Unlike the “grief” explanation, Smith’s explanation specifies the exact action that the
man performs. “He is grieving” can be the explanandum of a great variety of actions; not
so for Smith’s explanation.

However, it is natural to find Smith’s explanation for the action of the widower
unsatisfying. Smith is right that in order to act, it is necessary that the agent have
knowledge of how to perform the particular action. If one is persuaded that knowing how
is a species of knowing that, then we can impute to the agent a belief that he can roll
around in his wife’s clothes “by rolling around in those particular clothes that he is
rolling around in.” (22) More generally, he believes that he can ø by ø-ing in this way.
However, the obvious problem with this sort of explanation is its glaring triviality. While
it may be true that agents have such a belief, what disqualifies it as a good explanation is
that it is not informative (especially when we compare it to the grief explanation).

Indeed, the sort of explanation that Smith gives for the widower seems applicable to every bodily action.

Smith concedes that his explanation of the widower’s action has an air or triviality to it, but he does not find this worrisome. He maintains that his explanation “simply prompts the question ‘And why would anyone want to roll around in his dead wife’s clothes?’” He concludes, “here is where a supplement to the Humean’s explanation is both required and possible” (1998, 22). Smith maintains that this explanatory “supplement” puts the belief/desire pair that ultimately explains his action in a context which allows us to make sense of them. And the reason why reference to the emotion contextualizes the belief/desire pair is that “grief at the loss of a loved one is, by definition, a state in which we are disposed to think, and to desire, and to do, all sorts of things: cry, dwell on memories of the loved one, seek out things that remind us of the loved one and hold them close, and so on and so forth. Given that grief is such a state, it should therefore come as no surprise that we can explain the man’s action by citing the fact that he is grieving for his dead wife.” (22) In Smith’s response to Hursthouse, we hear clear echoes of Davidson’s response to Ryle. The Davidsonian explanation for why the man boasts is that he desires to elicit envy, and believes that he can do so by boasting. Given this most basic explanation, the further question arises naturally: why did the man have the desire to elicit envy? The answer to this question is that the man was vain.
Smith argues further that “while an explanation in terms of an emotion presupposes the availability of a Humean explanation, the reverse is not true. Agents sometimes act on their desires but not on the basis of any emotion.” (23) But even if we grant that the Humean explanation is in some way “presupposed,” it is not at all clear what contribution it makes to the overall explanation. If it provides little explanatory insight to the agent’s action, it must be jettisoned as the primary explanation for the agent’s action. And even if its explanatory value is derived from the more informative grief explanation which presupposes it, Smith never makes explicit what connection there is between the two.

Even after Smith’s Humean explanation and its supplement has been given, the agent’s action is still bewildering, baffling, flummoxing. Smith’s explanation leaves it utterly opaque to the interpreter how the agent’s action might be related to facts about his other desires, beliefs, emotions, principles, values, and so on. As a result, there is little sense in which her action is rationally (or even irrationally!) intelligible.

Consider the following two contrast cases:

(A) Action: Emily switches on the fan.
Humean Explanation: Emily was feeling hot and desired to feel cooler. She believed that by turning on the fan, she would cool herself.

(B) Action: Sam turns on the fan.
Humean Explanation: Sam was feeling hungry, and desired to rid himself of his hunger. He believed that by switching on the fan, he would alleviate his hunger pangs.
There is a clear sense in which the action explanation in (A) is successful, and that in (B) is unsuccessful. This is true even if we stipulate that the explanation in (B) picks out the exact belief/desire pair that ultimately causes the action. And the reason why the action explanation in (B) is unsuccessful is that, even after learning it, the interpreter of Sam’s action is still baffled. Davidson writes, “A reason rationalizes an action only if it leads us to see something the agent saw, or thought he saw, in his action [...]” (1980, 4). Now, if Sam wanted to alleviate his hunger, what on earth could he have seen that would allow him to realize his goal in switching on the fan? Why does he have such a strange belief? Could it be that he thinks that by switching on the fan, he is sending a secret message to god, who will in turn satisfy his hunger? There must be something about Sam that we do not know; and only when we have piece of the puzzle in our possession will we be in possession of an adequate explanation of his action. Davidson remarks,

> When we ask why someone acted as he did, we want to be provided with an interpretation. His behaviour seems strange, alien, outré, pointless, out of character, disconnected; or perhaps we cannot even recognize an action in it. When we learn his reason, we have an interpretation, a new description of what he did, which fits into a familiar picture. The picture includes some of the agent’s beliefs and attitudes; perhaps also goals, ends, principles, general character traits. (1980, 10)

With Hursthouse’s “arational” agents, the desire and the belief that Smith attributes to them clamor for further explanation. Smith realizes this, and offers an emotion-based
“supplement” to his belief/desire pair. The problem with his strategy is that the simple combination of the Humean explanation with the emotion explanation does not make it at all transparent why the grieving widower desires to roll around in his dead wife’s clothes, or why Jane desire to scratch out the eyes of a photo Joan. These actions are far too baffling to be explained by the combination of these two explanations, one of which is trivial, and the other of which is generic.

Peter Goldie (2000) describes a sense in which some desires are “primitively intelligible”, that is, they cannot be explained by anything other than the emotion which gives rise to them. For example, the desire to flee some object is primitively intelligible given that the agent is afraid of this object. The desire to harm or punish some person is primitively intelligible given that the agent is angry at this person. In cases like this, a Humean belief-desire explanation, supplemented by an emotion explanation would often be perfectly adequate. These impulsive actions have a characteristic shortsightedness to them; as Élisabeth Pacherie puts it, “They are done without regard for their consequences other than the immediate result aimed at.” (76) They manifest a lack of cognitive integration insofar as their appropriateness with respect to furthering goals and plans is not considered. The examples that Pacherie gives of these sort of actions include punching someone in a bar brawl, running away in fright, or hugging one’s fellow supporters in joy at the victory of one’s football team (76). The psychologist Nico Frijda (1986, 2002) describes action tendencies that are associated with strong emotions as clamoring for attention and for execution. An agent who acts from passion acts in a way
that is intelligible even if the desire we must impute to him seems to contradict other standing desires that we know him to have.

What is puzzling and unique about many of Hursthouse’s cases of “arational” action is that they are not “primitively intelligible” in this way, even if they are in some sense familiar. Recoiling in fear and lashing out in anger are both universal and almost autonomic. This is not true of an agent in the grip of rage who scratches a photo, or agent in the grip of grief who roles around in clothes, or an agent in the grip of shame who covers his face when no one is watching, or an agent in the grip of anger who talks to a can opener. These actions have a complexity to them that makes it seem as if they are done for an end. Smith emphasizes the strong intuition that intentional action must be done for an end. And if this is right, intentional action must have some cognitive component. The challenge is to provide a cognitive explanation that is informative, and to demonstrate the relation it bears to the emotion explanation.

A striking feature of many of the agents Hursthouse describes is that their behavior is somehow theatrical. I think that noticing this feature might provide us with an important insight for explaining their actions. Actors in plays behave as if propositions that they know to be false are in fact true. Because they are acting in the mode of as if, the type of propositional attitude that guides them in their actions is not belief. What’s more, they are not merely buffeted about by random passions: it is not uncommon to describe good acting as precise. In the next chapter I will outline a strategy for finding a better explanation for arational action on the model of pretending and imagining.
CHAPTER TWO:
HETEROGENEOUS ACTION EXPLANATION

All the best stories in the world are but one story in reality – the story of escape. It is the only thing which interests us all and at all times, how to escape.

- Arthur Christopher Benson
Recall: Michael Smith’s plausible analysis of desire requires the agent to “represent how the world is to be” (19). But since Hursthouse’s agents seem not be aiming at any intelligible goal, it is not clear in what sense they are representing “how the world is to be”. On the other hand, the agents that Hursthouse describes are positively brimming with desire. In fact, Hursthouse describes a category of desire states that are all too familiar. Moreover, the actions these agents perform are intimately tied up with their desire. So the question becomes, what is it that they desire, how is their desire related to their action?

A commonality among many of the agents Hursthouse describes is a background desire that represents something that is unattainable, or that is immoral, or that is impractical. It would not be surprising if the person who flings the can opener desires to open a can, but is unable to. Jane, who gouges out the eyes of the photo of Joan, may very well feel like gouging out Joan’s eyes, but of course this would be immoral. (Or perhaps it is undesirable because she might get caught and punished. Or perhaps it is undesirable because all things considered Joan would never want to do something as awful as that.) The person who violently destroys someone else’s pictures, letter, awards, books, poems, and music may desire to rid himself of painful memories that he finds difficult to expunge. The man who rolls around in his dead wife’s clothes desperately yearns that his wife were still alive, but nothing he can do will bring her back to life.
So in a sense, all of these agents do represent the way the world is to be. The problem is that, for reasons of morality, pragmatism, or powerlessness, they cannot make the world the way they want it. As is often the case, the world is recalcitrant to the will. But this does not stop them (or us) from desiring. So what role can desire play in action when the object agent desires is for some reason unattainable?

To be sure, the roles that frustrated desires play in our mental economy are diverse. Often when one’s desire is frustrated, one experiences short-lived dysphoria, after which the desire and the unhappiness both disappear completely with a little distraction. For example, if I desire to eat a piece of cheese, but I realize that my brother has eaten all the cheese in the fridge, I will typically feel mildly upset for the short period up until my attention has fixed itself on something else. Even long-held desires can rapidly cease to endure. For a period of months, I had a desire to be a geneticist. But when I discovered I was useless in the lab, this desire just seemed to “evaporate”.

However, in some cases, frustrated desires lead to strong emotions and brooding moods that endure for longer periods of time. For example, had I procured a pricey piece of Catalonian goat cheese, and my brother, who cannot tell cheddar from camembert, absconded with it, my emotional response (anger, resentment) might very well last a while. This is especially likely if there is nothing I can do about it, such as punishing my brother, or extracting an apology. Situations like this are particularly vexing because not only is my desire frustrated, but also I am utterly inefficacious. In circumstances like this, an agent’s imagination can come to play a pivotal role.
Using Michael Smith’s schema, the agent has the orienting belief, \( b_1 \); she represents how the world is (the can is not open, the painful memories remain, the man’s wife is dead, Joan has not received the punishment that she is due). The agent also has \( d \); she represents how the world is to be (the can is open, the painful memories have vanished, the wife is alive and well, Joan is blind). What the agent lacks is \( b_2 \), the instrumental belief. This should not be surprising, since the agent simply has no idea how her desire could possibly be realized within the limiting confines of moral and practical imperatives. When an agent is unable to conceive of a course of action that might bring about what she desires, why does she act at all? If actions are means by which agents (try to) realize their goals, there seem to be no candidate actions. Yet Hursthouse’s agents act.

I think that the answer to this puzzle becomes apparent when we realize that in arational action, emotion, desire, and action are related to each other in a more complex way than is the case in typical action. With arational action, the connection between emotion and desire is not primitively intelligible in the way that it is in Davidson’s cases from “Actions, Reasons, and Causes”. In the grip of intense fear, for example, it is primitively intelligible that an agent forms a desire to flee. We can easily fit this desire into an integrated and coherent picture of the agent’s overall psychology. The “rationalizing” explanation could go as follows. We might say that his fearful emotion is partially constitutive of a judgment or appraisal that he is in danger. We might also impute to him the background belief that in the face of a threat, rapid flight is a good way to secure his
safety. If we add to this picture the desire for self-preservation, it should not be at all puzzling why the desire to flee will often issue from the emotion of fear.

A desire to shout at and fling a can opener is not primitively intelligible in this way (which is not to say that it is altogether exotic or uncommon). Consider Hursthouse’s character (call him “Jim”) who vituperates the can opener and flings it to the ground. The can refuses to open, and Jim gets angrier and angrier. When one is angry or frustrated, it is always more satisfying to have someone to be angry or frustrated with. But in this case, there is no one else in the kitchen; there is just Jim, the tin opener, and the tin. He shouts at the tin opener (“You stupid piece of garbage!”) and flings it savagely to the ground.

Jim’s action is undoubtedly expressive. But still, Hursthouse is right: we should not make the unfounded assumption that Jim has the desire to express his emotion at the moment he flings the can opener. We often act in ways that are expressive of an emotion, but have no particular desire to express that emotion. As Hursthouse puts it, “the ascription of this extra desire requires extra justification.” (61) It is better to say that Jim desires to curse at and fling the can opener. Hursthouse is also right that Jim does not suddenly lose his marbles and form the false belief that the can opener is sentient. The attribution of temporary insanity is not apt. Recall that Hursthouse’s description of Jim’s action is that he does something that “might make sense if [the can opener] were sentient”. Also recall that Hursthouse describes the can opener as “uncooperative” (58).

---

13 To see the distinction here, think back to instances in which you acted in a way that was expressive of an emotion, but you would have preferred to keep that emotion hidden.
Although Hursthouse herself does not make this connection, I will argue that the best explanation of Jim’s behavior requires the supposition that Jim imagines or pretends that the can opener is sentient. This is not to say that Jim does this deliberately.

A candidate explanation for Jim’s action might go like this. Jim’s initial desire is to open the tin (call this D1). The world, however, is recalcitrant to Jim’s will: the can simply will not cooperate. As a result, Jim no longer has an instrumental belief (*B1) about how to open the tin: he does not see it as possible. Consequently, Jim feels the typical dysphoria that comes with impotence: he feels frustrated (call this emotion E). As a coping mechanism for mitigating emotion E, he imagines that the tin-opener is sentient and can understand his curses. Instead of feeling merely frustrated, he begins to feel angry (E2). Now, within the pretence we can explain Jim’s action. Jim has the desire-like imagining (D-LI) to mete out vengeance on the tin opener. He also has the belief-like imagining (B-LI) that he can injure it by verbally abusing it, and by flinging it to the ground. This explains why he shouts at it and flings it to the ground.

In arational actions, including actions borne of frustration, the connection between initial desire to action is more circuitous that in ordinary action. Whereas ordinary action proceeds directly from belief and desire, Jim’s action is explained by an initial desire (D1) that is frustrated, the lack of a relevant instrumental belief concerning how to realize his desire (*B1), the resulting emotion and coping mechanism that leads to the imagining, and the desire-like imagining (D-LI) and belief-like imagining (B-LI) that are the proximate causes for his action.
A pattern emerges for arational actions that issue from frustration of the will. When an agent desires something that, for some reason or the other, she does not have the power to realize, she finds herself impotent and powerless. As a mechanism for coping with and mitigating this dysphoria, the agent imagines or pretends that she is in some way efficacious: she can punish and injure inanimate objects; she can escape her shame by hiding from the world under her pillow; she can mete out vengeance by scratching a photograph; she can erase the past by destroying keepsakes and trinkets. In the imagined world that the agent creates, her actions have a meaning for which a standard belief-desire explanation is available, substituting mental states of imagining for belief and desire. Even though the imagining serves a purpose, it is best not to consider the imagining itself as an intentional action. It is more like a sub-personal purpose, a pattern of behavior that has an aim, even though it is never deliberately aimed at by an agent.

Imaginings play a role in explaining action in a wide variety of contexts. Consider, for example, the dyed-in-the-wool atheist physicalist whose older brother has recently died after being hit by a truck. He feels a mixture of grief at the profound loss, abandonment, and fear of the responsibilities that he must now face alone. In a state of acute grief, he finds himself talking to his brother’s grave. He stares at the grave, then glances up to the sky, and exclaims, “You bastard! I can’t believe you left me here to take care of mom!” Shall we say of this man that he abruptly and instantly disabuses himself of his atheist-physicalist world-view? Shall we impute to him the temporary belief that there really is an afterlife, and that his brother’s spirit has survived the death of his body? I submit that
we should not say either of these things. Rather, we should acknowledge the obvious: the man knows all too well that he will never talk to his brother again no matter how hard he tries. The world will not allow for this. As a way of coping with his anguish, the man imagines that he is able to communicate with his dead brother. His grief, his frustrated desires, and the resultant belief-like imaginings and desire-like imaginings are what explain his action. None of the elements in the action explanation is inessential.

In order to see the power of the imagination account, it is useful to contrast it with the “changed belief” or “temporary insanity” account. Consider the action explanation that goes as follows: in a fit of sorrow the grieving atheist temporarily adopts the belief that his brother’s spirit has survived the death of his body. The problem is that if the man thought his brother could actually hear him, and that this was to be their first communication after his untimely death, presumably he would not address him as, “You bastard!” Moreover, given that he knows that his brother was hit by a truck, it is unlikely that he takes him to be blameworthy for “abandoning” him. Granted, we do sometimes feel and express moral emotions such as indignation that are not warranted by our beliefs. However, one would think that if this man really believed himself to be in communion with his brother, he would be more careful about what he was saying to him.

This account of arational action has many advantages over the accounts critiqued earlier. It allows us to see the way in which the agent’s action is intentional by referring to an aim or end the agent has (within a pretense). It does not rely on imputing to the agent lunatic beliefs or temporary insanity. Unlike Smith’s account, it provides an illuminating
explanation for the agent’s action that takes into account both the affective and cognitive states of the agent and makes their relation comprehensible. And unlike Hursthouse’s account, it does not provide generic explanations (like “grief” or “anger”) for specific actions. It is my hypothesis that in order to make sense of arational actions, we must appeal to what I will call “heterogenous explanation”, that is, explanation that appeals both to real-life beliefs and desires, as well as belief-like imaginings and desire-like imaginings.

To see why the action explanations for certain types of action require appeal to both imagination and belief, consider the case of Sheila, who regularly performs a paradigmatic “arational” action:

Whenever Sheila’s computer “refuses to cooperate,” she curses at it using the most vulgar, offensive, and cruel language that she can summon.

Interestingly, when Sheila speaks to people who refuse to cooperate, she is demure, subtle, and diplomatic.

It is obvious that if Sheila really believed that the computer was sentient and capable of understanding, she would never treat it as she does. Clearly, Sheila’s belief that the computer is not sentient and does not understand English plays a role in explaining her action. Given her personality, if Sheila had even the slightest doubt, she would not behave in this way. On the other hand, the belief-like imagining that the computer can understand English and does feel must also play a causal role in Sheila’s cursing at her
computer, since her action only makes sense in this light. Any explanation of Sheila’s action that does not mention both the real-life belief and the belief-like imagining would not be sufficient to explain her action. Without the latter, her action would be opaque and inscrutable: why on earth does she scream at something that cannot hear, feel, or understand? Without the former, it would appear as if Sheila’s behavior were wildly inconsistent with her character: how could kind and demure Sheila use such violent and vulgar language? To reiterate: For an adequate explanation of Sheila’s action we must appeal to both the belief-like imagining and the real-life belief.

It has traditionally been held by philosophers that one of the key differences between beliefs and imaginings is that the former are intimately connected to action, while the latter are not\(^\text{14}\). When an agent believes that \(p\), and she is considering a course of action, she may take \(p\) into consideration when deciding how to act. On the other hand, if an agent merely imagines that \(p\), then generally speaking she will not take \(p\) into consideration when deciding how to act (unless of course, she is suffering from some psychological pathology). If my hunch is correct, this way of conceiving of the difference between belief and imagining has been a serious impediment to our understanding of arational actions. In cases like Sheila, both belief and pretense are intimately connected to action and to each other in rather complex and subtle ways.

But should we say that the invocation of a pretense is itself an action? I do not think we should. I can see two reasons against this position. The first is given by David Velleman

and applies to children: “the desire-belief explanation of pretending makes the child out
to be depressingly unchildlike.” (256) It just seems contrary to the phenomena that
children who pretend purposefully and intentionally keep “a firm grip on reality while
mounting an appearance conceived as such.” (256) When a child truly enters into a
fiction, he does not act out of the desire to represent the fiction. I think the same thing
goes for an adult who shouts at a can opener. Hursthouse has argued that it is simply
contrary to the phenomenology that the man “desires to express himself”. Similarly, I
think it is contrary to the phenomenology that the man “desires to invoke a pretense”.
The only desire he has at the moment is the desire to revenge himself on the can opener!
In cases of frustration, adults tend to “click into” a certain pretense, as described by the
locution “coping mechanism”.

The phenomenon of “clicking into” a pretense is not exclusive to impassioned actions, or
actions borne of frustration. We “click into” different pretenses depending on, among
other things, our surroundings and our state of mind. For example, the following
phenomenon has often been noted: a twenty-something returns to his parents’ home for a
visit, and acts as if he were eighteen years old again\textsuperscript{15}. The sorts of things that motivate
him, the states of mind that move him to action, are certainly not beliefs and desires that
he would ever avow. Presumably, they are all beliefs and desires that he has abandoned
long ago. But when he is home, they are somehow activated once more and his actions
are inscrutable (perhaps even to himself) without appeal to them.

\textsuperscript{15} This example was suggested to me by Nomy Arpaly.
Velleman’s “Inhibition” Account

A good reason to think that imagination can motivate action is that the best explanation for the class of actions that Hursthouse terms “arational” make essential appeal to the motivating force of belief-like imaginings and desire-like imaginings. There are some actions that would be utterly inscrutable otherwise. Take Hursthouse’s case of “muttering imprecations under one’s breath”. To explain such an action without appeal to the imagination, we would have to resort to reasons that are simply recherché. David Velleman (2001) puts this point well:

There is nothing that we both want to do and believe ourselves to be doing by talking to ourselves in this way. If someone stopped us on the street and asked “Why were you just muttering and shaking your head like that?” we could not offer an answer that began with the words “I wanted…” What could we have wanted? To walk along muttering and shaking our heads? Hardly. (264)

The inclusion of states of imagination as elements in action explanation raises the question, when and how does an imagining have priority over a countervailing belief? Velleman argues that the relationship between real-life beliefs and belief-like imaginings is one of “inhibition”. He holds that pretend actions in adults are often “inhibited” by countervailing real-life beliefs:
Thus, for example, most deliberate imagining is accompanied by countervailing beliefs, embodying the subject’s knowledge of the facts that he is imagining to be otherwise, such as his knowledge that an imagined pail of water is really a chair. These beliefs exert their worn motivational force, which can be expected to compete with that of the subject’s imagination. Ordinary beliefs are not regularly accompanied by countervailing beliefs, and so their motivational force encounters less competition. I have also hypothesized that the motivational force of imagining comes under an inhibition, whose effects can be detected for example, in the way that we lower our voices when talking to ourselves. (272)

While Velleman’s general point (that both pretense and belief make a causal contribution to action) is well taken, I don’t think that the metaphor of motivational forces that “outweigh” or “inhibit” each other is altogether apt. In a related footnote, Velleman describes the agent as acting on the “vector sum” of all motives combined (footnote 50). This locution, too, I find infelicitous. It is not as if the motivational force of the belief-like imagining and the desire-like imagining is “added up” to the countervailing belief and desire, and when the “vector sum” is positive one is moved to act, although in a qualified or mitigated way. In the case of Sheila, her action is intensified as a result of the countervailing belief that the computer cannot feel or understand language. Were it not for this countervailing belief, she would not speak with such force and venom. The pretend action can only be explained by a complex interaction between beliefs and
desires within the pretense and real-life beliefs and desires. The real and imaginary
cognitive and conative states interact with each other as reasons rather than as forces.

Consider some of the relevant attitudes in Sheila’s psychology:

Within the imagining:
Orienting Belief-like imagining: The computer can feel and understand language.
Desire-like imagining: I want to punish the computer.
Instrumental belief-like imagining: I can punish the computer by shouting violent insults
at it.

Outside of the imagining:
Orienting Belief: The computer cannot feel or understand language. If I shout at the
computer it will not be hurt, since it cannot understand language or experience emotions.
Desire: I do not want to hurt or insult anyone using coarse and violent language.

The relationship between the motives in real-life and the motives within the pretense is
clearly not one of “outweighing” or of “inhibition”. I think that Velleman somewhat mis-describes the case of the woman who lowers her voice when talking to herself. Surely
the woman lowers her voice out of self-consciousness, and not as a result of the
“opposing force” of a countervailing belief that she has no interlocutor. When in her
house, the woman might feel freer to talk to herself out-loud. This is not due to an
impoverishment of the motivational force of her belief that she has no interlocutor, but
rather an alleviation of self-consciousness.
The position that I am advancing is not uncontroversial: Shawn Nichols and Steven Stitch have argued forcefully against the view that imaginings can motivate\textsuperscript{16}. They object that if imaginary desires have the same causal power as real desires, then a child who has an imaginary desire to eat a pie along with an imaginary belief that a glob of mud is a pie, would actually eat the mud pie (which he doesn’t). In Chapter 3 I will address this objection.

Heterogeneous Action, Pretend Games, and Virtual Agency

A recent *Wall Street Journal* article entitled “Is this man cheating on his wife?” chronicled the life of Ric Hoogestraat, 53-year-old man with a long ponytail, graying sideburns, a salt-and-pepper handlebar mustache, and a large paunch. His look is described as that of a “cross between a techie and the Grateful Dead fan that he is.” (Alter, 2007) During different periods of his life, Hoogestraat worked as an elementary school teacher, a call center operator, a computer graphics instructor, and a vendor for herbs and essential oils at Renaissance fairs.

But Mr. Hoogestraat also has another persona. For an average of six hours a day, and fourteen hours at a stretch on weekends, he inhabits an “avatar” named “Dutch Hoorenbeek” on the multi-player, immersive game called Second Life. His life as the avatar “Dutch” is rather different from his life as “Ric”:

[Dutch] looks like a younger, physically enhanced version of [Ric]: a biker with a long black ponytail, strong jaw and thick handlebar mustache. In the virtual world, he's a successful entrepreneur with a net worth of about $1.5 million in the site's currency, the linden, which can be earned or purchased through Second Life's Web site at a rate of about 250 lindens per U.S. dollar. He owns a mall, a private beach club, a dance club and a strip club. He has 25 employees, online persons known as avatars who are operated by
other players, including a security guard, a mall concierge, a manager and assistant manager, and the "exotic dancers" at his club. He designs bikinis and lingerie, and sells them through his chain store, Red Headed Lovers.

All of this has proven rather distressing for his long suffering (real) wife, Sue Hoogestraat, who pays household bills, cooks, does laundry, takes care of their three dogs and empties ashtrays around the house. Even worse from her perspective is the fact that her husband Ric, in his cyber life, is married to a tall, wiry redhead named Tenaj Jackalope, an avatar played by a 38-year-old Canadian woman whom he has never met. “Dutch” and “Tenaj” share a great deal of romance, intimacy, and trust; Ric and Sue, not so much. The case raises a number of fascinating psychological, sociological, and ethical questions. There is also a very practical question to be answered: Is Mr. Hoogestraat cheating on his wife? According to an account recently articulated by David Velleman, the answer would seem to be “yes”.

In a paper entitled “Bodies, Selves” (forthcoming), David Velleman describes the manner in which “virtual agency” is exercised in multi-user role-playing games such as Second Life. Velleman notes that a player of Second Life will typically attribute to himself the various actions that his Second Life avatar performs in the virtual world. He will say things such as I wielded the knife or I was late for class or I got married to Cindy last year. Velleman makes the surprising claim that we have good reason to take these avowals literally. He maintains that “although the player’s actions in the virtual world are merely fictional, he is really performing them.” (4) This interpretation is in sharp
contrast to the way that we generally interpret the action of actors in theater and film, or of children in pretend games. When an actor reports *After I got off the subway I stabbed Tony Soprano* or *I commit suicide in Act V* it would be very odd to interpret what he says as literally true. Similarly, in pretend games, it would be very odd to understand a child’s utterance of *I killed the pirate* literally. So why should matters be different for virtual play? Velleman articulates two central disanalogies between make-believe in pretend games and acting, and the phenomenon of virtual play. Unlike the imaginary world of children’s fantasies, the virtual world of Second Life is both *recalcitrant* and *determinate*. In what follows, I will contrast these characteristics with the characteristics of the episodic imaginary worlds of Hursthouse’s characters. 

According to Velleman, a distinctive feature of virtual play is that players cannot make stipulative additions to the virtual truths in the game. For this reason, the virtual world manifests a certain *recalcitrance* to the will that is characteristic of real life. This recalcitrance is part of what explains the players’ intense psychological engagement for extended durations. They must struggle to achieve their goals, just like in real life. Moreover, since their beliefs about the virtual world are constantly being reality-tested, there is always the chance that they might be getting things wrong. Unlike a pretend

---

17 In “Bodies, Selves” Velleman argues that virtual play has a similar structure to that of Freudian analytic transference: “In the analytic transference, the patient may attempt to seduce the analyst, but not literally. Patient and analyst are rather embodying fictional versions of child and parent, and the attempt by the one to seduce the other is a fictional attempt at a fictional seduction. But the patient is really making that attempt, is really the agent of that unreal action. Actions carried out within the transference are not make-believe; they are fictional actions literally performed.” (3)

18 Velleman describes a third feature of virtual play that he calls “opacity” (roughly, the players’ inability to see who the real people behind the other avatars are).
game of pirates where a child can simply stipulate, “I found the treasure! It was in the ship’s hull!” and thereby make it true in the fiction, in virtual games like *Second Life* the player must actually search for and discover the treasure. The treasure may very well be difficult to find, and there is a chance that the player, despite his best efforts, will not find it. A player’s hypothesis that the treasure is located in, say, the ship’s hull, may turn out to be false. Although a child playing pretend pirates may imagine that his hypothesis turns out to be false, it will only turn out false if he so decides. As a result, participants in games of make believe cannot be frustrated or disappointed in a way that it is possible for inhabitants of *Second Life*.

Although Velleman is right to point out that virtual worlds can be challenging and recalcitrant, there is an obvious sense in which they are *less* recalcitrant than the real world. For example, in Second Life there is no recalcitrant beer belly or pimple. And so there is a real sense in which playing Second Life is a kind of escapism. It’s not that the inhabitants of Second Life do not have problems: a world without conflict would also be without interest. But the salient difference with the real world is that the inhabitants of Second Life are able to *choose* their problems. Many choose struggles that are noble: combating racism, reversing environmental degradation, or fighting economic injustice. Others pursue less altruistic passions: conquest, seduction, and sexual gratification. The sorts of challenges that Hoogestraat-as-Hoorenbeek chooses to face range from the challenges of expanding a successful business, to the minutiae of lingerie design. He does not choose to struggle against paperwork, bureaucracy, ungrateful children, boredom with his spouse, or chronic disease.
Although Hursthouse’s agents may also be indulging in a kind of escapism, it is not of the same order. First, they do not choose the sorts of problems they struggle with. Treacherous friends, deceased family members, and shoddily designed can openers are not the sorts of challenges that one would normally elect to face. Second, the “escape” from reality they derive from imagining that they can punish a can opener, mete out justice at a distance, or commune with the dead, is fleeting and ephemeral. This sort of magical thinking is unlikely to persist.

The motivation of Hursthouse’s agent who rolls around in his dead wife’s clothes is structurally similar to that of Ric Hoogestraat. Each of them is motivated to escape into the realm of the imagination because the real world is in some way recalcitrant to his desire (the grieving widower wants his wife back, while Hoogestraat wants a different kind of wife, at least for a while). However, the degree of recalcitrance within each of the imagined worlds is quite different. Hoogestraat must work to win the heart of his cyber beloved, and there is a chance that he will fail. This is not analogous to the case of the widower: it is true in the imagining that he is with his wife, just because he imagines it so. Hoogestraat’s escape from reality is ultimately more enduring. His suspension of disbelief is more robust and stable, perhaps in part because the imagined world shares some of the recalcitrance of real life.

Besides their distinctive recalcitrance, Velleman describes the way in which virtual worlds have a very high degree of determinateness relative to their players’ knowledge as
compared to the worlds of pretend play. What is fictional in a world of make-believe includes only what has been stipulated or enacted by the players, plus some of what follows logically from their overt contributions. In a virtual world, what is fictional far outruns what any of the individual players knows. The players need explore to the world in order to find out what is true in the world; they can only see from the perspective of their avatars. Moreover, their actions have determinate consequences. When two inhabitants of Second Life get married, they cannot later stipulate that the marriage never happened: they must get a divorce.

The episodic fantasies of Hursthouse’s characters do not exhibit anything like this kind of determinateness. When Joan imagines that she scratches out the eyes of Jane, she need not worry about the consequences for Jane’s family or the long arm of the of the law. Indeed, there is simply no fact of the matter to many questions we might have about people and events in these kinds of fantasy worlds.

The fantasies of Hursthouse’s characters exhibit a great deal of what Tamar Gendler terms “disparity” (2003). She describes two ways in which imaginative content can differ from non-defective belief content, and thereby exhibit disparity:

[W]hat I successfully imagine may be incomplete, in the sense that some of its features may remain permanently – even explicitly – unspecified and unspecifiable; and what I successfully imagine may be incoherent,
the sense that some of its features may be conceptually or logically incompatible. (137)

When the grieving atheist calls his brother a bastard at his grave, there need not be a fact of the matter about whether his brother really deserves this vituperation, whether this is the first time they have talked since his death, or whether there will be an opportunity to make amends. In a virtual world, many or all of these sorts of conditions would have to be specified.

Perhaps part of what accounts for the radical disparity in many of Hursthouse’s cases is that none of them involve interactions with other people. When imagining and pretending is a joint activity, certain principles of generation (conventions that govern which features of the actual world are mapped onto the real one) must be respected in order for each agent to understand the significance of other agents’ pretend action. The way in which fictional truths are generated in joint activities will be explored further in Chapter Three.
CHAPTER THREE: COGNITIVE THEORIES AND DESIRE-LIKE IMAGININGS
While the heterogeneous account of arational action explanation has many virtues, it also introduces some vexing questions. Chief among them is, “Can the imagination motivate action, and if so, how?” Recently, philosophers have been paying increasing attention to this important question. We often take the paradigm of imagination to be a state that is not connected with real-life action. Indeed, often this is the case: when I lie in my hammock and daydream about flying, or about 19th century France, or about peeling a pomegranate, I am not generally moved to do anything. But in some cases of imagining, those which contribute to an act of pretence, the imagination does seem to be implicated in my action.

In Chapter Two I argued that the best explanation for the class of actions that Hursthouse describes makes essential appeal to belief-like imaginings and desire-like imaginings. Indeed, I think that these actions would be utterly inscrutable otherwise. Take Hursthouse’s case of muttering imprecations under one’s breath. To explain such an action without the imagination, we would have to resort to reasons that are simply recherché. David Velleman (2001) puts this point well:

This behavior eludes desire-belief explanation. There is nothing that we both want to do and believe ourselves to be doing by talking to ourselves in this way. If someone stopped us on the street and asked, “Why were you just muttering and shaking your head like that?” we could not offer an answer that began with the words “I wanted…”
What could we have wanted? To walk along muttering and shaking our heads? Hardly. (264)

I argue that these sorts of actions become comprehensible and rationalizable only if we give an explanation in terms of belief-like imaginings and desire-like imaginings. The position that I am advancing is not uncontroversial: Nichols and Stitch have argued forcefully against the thesis that imaginings can motivate. If their analysis is perspicuous, then an important part of my project of explaining arational actions by adverting belief-like imaginings and desire-like imaginings will have been undermined.

In the first part of this chapter, I will outline in broad strokes the Nichols and Stitch theory. In section II, I argue that the theory lacks the necessary resources to explain the complex mapping relation between belief and imagining that is seen in pretend action. In section III, I make the case for desire-like imaginings.

---

I. The Nichols and Stitch Cognitive Theory of Pretense

Nichols and Stitch maintain that in order to explain the phenomenon of imaginative pretense we must make certain assumptions about the basic architecture of mind. Their account is built from observations of common empirical phenomena relating to belief and imagination. They describe five common features of pretense that their theory is designed to explain (2000, 118-122):

1. *Starting from an initial premise.* Typical episodes of pretense begin with basic assumptions about what is to be pretended: for example, “Let’s have a tea party.” The pretender can either generate the initial premise herself, or the initial premise can be supplied by someone else.

2. *Inferential elaboration.* From the initial premise, together with background knowledge, the pretender can draw inferences about what is going on in the pretense. For example, in the now famous “tea party” experiment by Alan Leslie (1987), an experimenter pretends to fill up two cups in front of a child, and then turns one cup upside down. When the young child in the experiment is asked which cup is empty, he points to the one that has been turned upside down.

3. *Non-inferential elaboration (embellishment).* Pretenders often elaborate scenarios in a way that is non-inferential in order to “fill out the story”. For instance, in playing the game, “Grocery Shopping”, a child may decide to buy six plums, three bananas, and a box of cereal.
4. Production of appropriate pretend behavior. Pretenders engage in action that is suitable to the particular pretense at hand. A child pretending that a banana is a telephone puts one end of the banana to his ear, and the other to his mouth.

5. Cognitive quarantine. Events that occur in a pretense typically have limited effect on the post-pretense cognitive state of the pretender. The child who uses the banana as a telephone to talk to daddy at the office does not end up believing that he has really talked to daddy.

Nichols and Stitch take their task to be to describe what the basic cognitive architecture of the mind must be like in order to account for the five types of phenomena described above. They assume that imagining, like believing, is a propositional attitude, a contentful mental state. According to their theory, beliefs and imaginings differ in their functional roles, that is, they differ in their connections to various mental mechanisms. They represent this by positing a “Belief Box”, containing representations of the world, a “Desire Box”, containing representations of the world as we would like it to be, and a “Pretense Box”, containing representations that can have the same sort of contents as beliefs and desires, but do not “represent the world as it is or as we’d like it to be, but rather […] what the world would be like given some set of assumptions that we may neither believe to be true nor want to be true.” (122)20 They hypothesize that the Pretense Box is used in a variety of tasks including mind-reading, strategy testing, and

---

20 Of course, a “box” is not meant to denote a spatial location in the brain; rather, it is meant to denote a class of mental states that share “an important cluster of causal properties that are not shared by other types of states in the system” (Nichols and Stich 2000, 121). In their early writings, they refer to the “Pretense Box” as the “Possible Worlds Box”, or “PWB”. In later writings (post-2000), they abandon this expression because it carries too much philosophical baggage. To avoid confusion, I will always use the term “Pretense Box”, even when referring to their pre-2000 writings.
empathy; they speculate that the evolutionary function of the Pretense Box was to facilitate reasoning about hypothetical situations.

A key component of the Nichols and Stitch theory is that beliefs and imaginings are “in the same code”. Nichols elucidates the single code hypothesis as follows: “The core claim of the single-code hypothesis is that a mechanism that takes pretense representations as input will process those representations much as it would process isomorphic belief representations.” (Nichols 2004, 131) As a consequence of beliefs and imaginings being a single code, the very same inference mechanism that operates on updating beliefs in the light of new evidence also operates on imaginings. They call this mechanism “The Updater”.

We are now in a position to get a preliminary glimpse of how Nichols & Stitch model pretend action. At the initiation of an episode of pretense, a pretense initiating premise is added to the Pretense Box. Along with the pretense initiating premise, all of the contents of the Belief Box are also added to the Pretense Box. The problem, however, is that some of the contents of the Belief Box will likely be inconsistent with some of the contents of the Pretense Box. For example, there will probably be an item in the Belief Box with the content “the banana is not a telephone”. This is where the Updater comes in. The Updater goes through the representations in the Pretense Box “eliminating or changing those that are incompatible with the pretense premises” (2000, 124). In this way, inferential chaos is avoided: “Everything in the pretender’s store of beliefs gets
thrown into the [Pretense Box] except if it has been filtered out (i.e. altered or eliminated) by the Updater”. (2000, 125)

Also contained within the Belief Box are a number of “Scripts” that detail the way in which certain types of situations typically unfold. For example, Nichols and Stitch asked subjects to participate in a “Fast Food” game in which they were to pretend that they were ordering a meal at a fast food restaurant. They found that most of the subjects in the experiment followed a standard pattern: order the food, pay for the food, eat the food. They seemed to following a “script”. These scripts, however, do not fully specify the way in which a pretense gets filled out. There are still many choices to be made (for example, in the “Fast Food” game, the pretender might choose to order a fish burger, chicken nuggets, onion rings, etc.). To account for this phenomenon, Nichols and Stitch add yet another component to their theory: the “Script Elaborator”. The Script Elaborator fills in the heretofore unspecified details, the elements of the pretence “that can’t be inferred from the pretense premise, the (filtered) contents of the Belief Box and the pretender’s knowledge of what has happened earlier on in the pretense.” (2000, 127)

We now have all the key ingredients of the Nichols and Stitch theory. The Nichols and Stitch “boxology” diagram is presented below. The boxes (Belief, Desire, Pretense) represent functionally defined groups of mental representations. The diamonds (Inference, Perceptual process, Body monitoring system, Script elaborator, Decision

---

making system, and Action control system) represent mental mechanisms that take the contents of the boxes as inputs or outputs.

Each element of the “mental architecture” that Nichols & Stitch propose plays a role in explaining the five phenomena outlined earlier. The final element needed is an understanding of how imagination and motivation are connected in pretend action.

Conspicuous by its absence is a box containing conative states analogous to the belief-like states in the Pretense Box. Nichols and Stitch see no need for such a box. Instead, they posit a real desire that pretenders have: “Pretenders behave the way they do because they want to behave in the way that is similar to the way some character or object behaves in the possible world whose description is contained in the [Pretense] Box” (2000, 128).
II. The complexity of belief-imagining mappings

In this section, I will first present a series of examples that illustrate the sophisticated manner in which representations in both the Belief Box and the Pretense Box play a role in explaining pretend action. These examples call into question the Nichols and Stitch characterization of The Updater as a mechanism designed to maximize consistency. The examples will also serve to demonstrate that certain representations contained in the Belief Box, which are absent from the Pretense Box, are necessary in explaining actions within pretense episodes.

Version 1

Sinan is teaching his young son, Ali, what to do if there’s an emergency in the house and his parents have not come home from work yet. He gets Ali to go through a pretend rehearsal of what he would do were he to notice a fire in the kitchen. Just to make sure that Ali doesn’t actually call the fire department, Sinan had previously disconnected the phone when Ali was not in the room. Ali runs to the phone and pretends to dial 911 (while not fully pressing the buttons, since his dad has taught him never to call 911 unless there is a real emergency), and says in a calm voice, “Fire department! I am at 173 Williams street in Providence. There is a fire in the kitchen, and my parents are not home.”

Version 2

911\textsuperscript{22}

This example is based on an example given in Nichols and Stitch’s 2000 paper in *Cognition.*
Sinan is teaching his young son, Ali, what to do if there’s an emergency in the house and his parents have not come home from work yet. He gets Ali to go through a pretend rehearsal of what he would do were he to notice a fire in the kitchen. *Before they begin their rehearsal, Sinan disconnects the phone from the wall socket in a conspicuous way so that Ali can see what he has done.* Ali runs to the phone and dials 911, fully pressing down the buttons. He says in a calm voice, “Fire department! I am at 173 Williams street in Providence. There is a fire in the kitchen, and my parents are not home.”

Version 1 and Version 2 of 911 are very similar: they both begin with the same pretense-initiating premise (there’s a fire in the kitchen and mom and dad are at work). Further, both versions utilize the same props (a telephone) and both versions activate the same belief-like imaginings (there’s fire in the kitchen, no one is home) and the same desire-like imaginings (to warn the fire department, to be safe). The salient different between the two scenarios is that in Version 2 Ali has a belief that he does not have in Version 1, that is, in Version 2 Ali believes that his father has disconnected the phone, while in Version 1 he does not have this belief since Sinan disconnected the phone surreptitiously. This difference in belief is what explains why in Version 1 Ali merely taps the buttons on the telephone, whereas in Version 2 he presses them.

While simple and schematic, 911 is telling. In interpreting Ali’s action and trying to come to something like what a Davidsonian would call a rationalizing explanation, we would have to take Ali’s (real-life) belief regarding whether the telephone is plugged in into consideration. Indeed, if Ali’s had pressed rather than tapped the button in Version
1, his father might very well have been baffled, and sought further explanation. Why did Ali 
do that when I taught him never to dial 911 unless there was a real emergency? Was he 
being deliberately obdurate? Was he trying to rile me? Or did he get so caught up in the 
game, that he lost track of what was real and what was pretend? Let’s see… did he seem 
surprised when someone picked up the phone on the other end?

Careful analysis of this case leads to a seemingly paradoxical result for the Nichols and 
Stitch picture. Consider Ali’s belief in Version 2 that “the phone is unplugged”. In 
giving an illuminating rationalizing explanation for the pretend action within this game, 
we must advert to a real-life belief that is both “activated” and “inert”. On the one hand, 
it is “activated” insofar as it makes it the case that Ali presses rather than taps the buttons; 
the presence of this belief is the “difference maker”. On the other hand, within the 
pretense it is “inert” insofar as, were it true in the pretense that the phone is unplugged, 
Ali would plug it in before calling 911. Since Ali does not behave in this way, there can 
be no representation with the content [the phone is unplugged] present in the Pretense 
Box. So, we have it that a representation that plays a causal role in regulating a pretend 
action is absent from the Pretense Box.

Ali is able to draw the inference “the phone is unplugged, so it is safe to press the 
buttons” but he is able to quarantine23 the belief “the phone is unplugged” so that, within

23 Here, my use of the term “quarantine” departs from that of Tamar Gendler (2003), who uses the term to 
describe the extent to which “events within the pretense-episode are taken to have effects only within that 
pretense-episode (e.g. the child does not expect that “spilling” (pretend) “tea” will result in the table really 
being wet), or more generally, to the extent that proto-beliefs and proto-attitudes concerning the pretended state of 
affairs are not treated as beliefs and attitudes relevant to guiding action in the actual world.” If what I am 
arguing is correct, then there is something akin to Gendler’s “quarantining” that goes on in the belief-to-
pretense direction, as well as the pretense-to-belief direction that Gendler describes.
the pretense, it still makes sense to dial 911 immediately, rather than first plugging in the phone. Indeed, one of the requirements of playing a game of pretense is that we keep track of pretense and reality in this manner, that we manage to simultaneously attend to, and blind ourselves to, certain beliefs.

Recall that according to Nichols and Stitch, the contents of the Pretense Box are composed of the pretense initiating premise, together with the representations in the Belief Box. Since some of the items in the Belief Box will often (though not necessarily) contradict the pretense initiating premise (e.g. Ali pretends that there is a fire in the kitchen, but believes that there is no fire in the kitchen) inferential chaos is avoided with the help of the the Updater, which “goes through the representations in the [Pretense Box] eliminating or changing those that are incompatible with the pretense premises. Thus, these representations are unavailable as premises when the inference mechanism engages in inferential elaboration on the pretense premises.” (2000, 124)

What is notable about 911 is that it demonstrates that the work of the Updater is more complex than might have initially been conceived. In Version 2, the representation “the phone is unplugged” has clearly been purged from the Pretense Box, since Ali does not pretend to plug in the phone again. However, the belief “the phone is unplugged” is not incompatible with the pretense initiating premise “there’s a fire in the kitchen”. The latter does not rule out the former; indeed, it does not even make it less probable. Regardless, Ali does not plug the phone back in; he does not even pretend to plug the phone back in. Clearly, it is not fictional that the phone is unplugged (since if it were, then the pretense
would lose all coherence – why is Ali dialing 911 when the phone is unplugged? Does he believe that the phone will magically work even though it is unplugged? Is he pretending this? Unlikely.) Even though the Updater is doing its work, it is not targeting inconsistency or “inferential chaos”.

So: why does the Updater update here, if not for consistency? The obvious answer is that Ali knows very well why his father unplugged the phone in the first place. He did it to better facilitate the pretense (by doing this, he allows Ali to press the buttons on the phone, which is a better simulacrum of what he would actually do were there to be a real emergency). It is this understanding that explains why Ali’s belief, “the phone is unplugged” is updated to “the phone is not unplugged” in the pretense. The explanation for this “update” is Ali’s further belief, “My dad unplugged the phone on purpose, for our game, and he does not intend for me to pretend that the phone is unplugged.”

*Mud Pies: Variations on a Theme by Walton*

The way in which pretend action is circumscribed by real beliefs is a complicated issue. It is not a simple matter of which of the items in the Pretense Box survives the editing of the Updater. The following examples serve to illustrate some of this complexity.

In *Mimesis as Make-Believe* Kendall Walton describes the (now canonical) case of children who shape, bake, and “eat” mud pies. Consider these three related cases based on Walton’s original example:

*Version 1 (the original): The children are pretending to make and eat pies. Instead of*
flour, water, baking soda, etc, they use *mud*. After they have finished “baking” (putting the gobs of mud into the “oven” (really a drawer), they “eat the pies”, that is, they bring the pies up to their lips, and then, giggling, put them back down. After they are “finished eating”, they exclaim, “Wow! Those were scrumptious!”

*Version 2*: The children are pretending to make and eat pies. Instead of flour, water, baking soda, etc, they use *bagel dough*. After they have finished “baking” (putting the bagel dough into the drawer), they “eat the pies”. They bring pies up to their lips, take a little nimble, and then, giggling, put them back down. After they are “finished eating” they exclaim, “Wow! Those were scrumptious!”

*Version 3*: The children are pretending to make and eat a pie. Instead of flour, water, baking soda, etc, they use *crumbled Oreo cookies*. After they have finished “baking” (putting the crumbled Oreos into the drawer), they “eat the pies”. In doing so, they bring the “pies” to their mouths, and proceed to eat copious amounts of the crumbled Oreos.

These variations on Walton’s original example give some indication of the variety of ways in which beliefs and imaginings contribute to causing and explaining behavior. In all three examples, the pretense initiating premise as well as the “script” are the same. However, in *Version 1*, the real-life belief that the props being used (gobs of mud) are inedible causes and explains that the children act in a way that is only *symbolic* of the way they would act, were they to be really eating pie. In *Version 2*, the children’s belief that the prop being used (bagel dough) is edible in small amounts but not very tasty, causes and explains that the children act in a way that is a *truncated* variation on the way
that they would act were they to be eating a real pie; that is, the children only take little small bites (even though, outside of the game, they would not nimble on bagel dough at all). Finally, in Version 3, the children’s belief that the prop being used is edible, sweet, and delicious causes and explains the fact that they eat copious amounts of “pie”, in much the same full-blooded way that they would eat copious amounts of real pie.

In interpreting the children’s action and in constructing a rationalizing explanation, we must advert to their beliefs outside of the pretense in much the same way that we did in 911. The appropriateness of the chosen mode of pretense – symbolic, truncated, or full-blooded – can only be explained if we make reference to certain representations which are not contained within the Pretense Box (for example, in Version 1, it is certainly not contained within the Pretense Box that the pies are made out of an ingredient that is inedible). We can only give a rationalizing explanation if we advert to the children’s real life beliefs about the props that they are using in the pretense. If the item “the dough is inedible” were added to the Pretense Box, it would make no sense for the children to continue baking. If the children in Version 1 were actually to eat the mud and make themselves sick, we would have to admit that something had gone very awry. Were the children purposely making themselves sick to win their parents’ attention? Or, did they get so caught up in the game of pretend that they failed to keep in mind salient beliefs about the props they were using? As in 911, we see that, barring exceptional circumstances, a pretender’s belief, which is not included in the representations in the Pretense Box, plays a regulating role in pretend action. We also see that the “Updater” does not just target inconsistency.
My final example, *Tea for Two*, illustrates how in complex, coordinated episodes of pretense, we can only explain the actors’ actions by adverting to: (1) the contents of their respective Pretense Boxes; (2) their real beliefs about each of the contents of other’s Pretense Boxes, and (3) their real beliefs about the contents of each other’s Belief Boxes.

Unlike children, adults rarely engage in deliberate and elaborate episodes of pretense.

This does not mean that they are unable to. Consider for example, amateur and professional actors. What actors do is different from what children do, insofar as actors are usually given highly detailed instructions about what to say by scriptwriters, and what to do by directors. But this is not always the case. For example, the dramaturgical technique of “retroscripting” eschews traditional scripts in favor of scene outlines of varying levels of detail from which actors improvise. Actors are given what Nichols and Stitch would call a “pretense-initiating premise” as well as some general rules and guidelines, and are allowed considerable leeway in developing the scene. The following example of a pretense episode between two adults can be interpreted as an instance of retroscripting.

---

24 This technique has recently been used to critical acclaim in Larry David’s hit series, *Curb Your Enthusiasm*. David (both the lead writer and main character) gives different actors varying levels of information about the plot, which makes possible very real expressions of surprise when an actor is made aware of a plot twist mid-episode.
**Tea for Two**

Imagine that two people, Andy and Bertrand, are playing the game “Tea for Two”.

A: [pretending to switch on the “electric kettle” (really a flower vase) to boil the “water” (really just air)] What a lovely day for a tea party!

B: [pretending to set out the “teacups” (really glasses), the “sugar bowl” (really a coffee mug), and a “small jug of milk” (really a salt-shaker)] Yes, and it’s four o’clock: precisely time t!

A: [making the “click” sound of an electric kettle switching off] What sort of tea do you prefer?

B: [setting out the “crumpets” (really Oreos)] You wouldn’t happen to have Lady Grey?

A: [putting a “tea bag” (really a grape) into each “teacup”] Of course I have Lady Grey!

B: [grinning with anticipation] Ah, splendid!

A: [pretending to pour the “water” from the “kettle” into the “teacups”, and then immediately adding “milk”] Lady Grey is also my favorite.

B: [picking up the “sugar bowl”] Do you take sugar?

A: [proffering his “teacup”] Yes please!

B: [bringing the “teacup” to his lips] There’s nothing like a refreshing cuppa!

A: [sighing in approval] Indubitably!
While Andy and Bertrand seem to be following all the rules for generating fictional truths by using props, Bertrand has some salient beliefs about tea drinking that Andy lacks. Being British and a regular tea-drinker, Bertrand thinks that tea is best made in a teapot, preferable one that is covered by a tea cozy. An American and a coffee-drinker, Andy does not own a teapot, and has never heard of a tea-cozy. In fact, prior to playing the game “Tea for Two”, he had never heard of Lady Grey tea. Bertrand has the (justified, true) belief that pouring milk into tea immediately after pouring hot water over the tea bag produces a weak, grey, milky liquid that barely deserves to be called tea. Andy has yet to add this notion to his conceptual repertoire.

What is interesting about this game is the pragmatic admixture of analogy and disanology with reality. Clearly, some of the inferences that Andy and Bertrand make within the pretense are isomorphic to the inferences that they would make in reality. For example: the water is boiling when the electric kettle makes a “click”. On the other hand, because of Andy’s ignorance, there are some important disanalogsies: in reality, when milk is added right after hot water, the result is bad-tasting tea; in Andy and Bertrand’s pretense, this is not the case. To the contrary, both players pretend to be greatly enjoying their tea.

Were Andy and Bertrand really to have afternoon tea, Bertrand would have to work hard to hide his disgust after tasting Andy’s terrible tea. In the real world, unlike in pretend games, the world manifests a recalcitrance that can be very inconvenient. Were Bertrand to say something like, “There’s nothing like a refreshing cuppa!” after taking his first sip,

---

25 See Walton Mimesis as Make-Believe pp. 38-42.
he would be dissimulating. Moreover, if Andy were to go along and say, “Lady Grey is also my favorite!” he’d be flat out lying. But, within the game “Tea for Two”, it need not enter into either pretender’s Pretence Box that Bertrand is revolted by the tea he’s drinking, or that Andy is lying about his knowledge of tea to save face.

Of course, if Bertrand and Andy had wanted to steer the “Tea for Two” game in a slightly different direction, they could have. Andy might have admitted his ignorance (“Bertrand – what the hell is ‘Lady Grey’?”); Bertrand might have shunned Andy’s offering (“Andy, I’m sorry to tell you, but your tea tastes like dishwater”). But what’s interesting is that it’s largely up to them whether to take the pretense in this direction or not. If Andy and Bertrand decide that they want the tenor of “Tea for Two” to be “a jolly good time” rather than “a peevish afternoon” then it need not be included in the Pretense box that Andy is lying to his friend about his favorite tea, or that Bertrand is disgusted by what is in his cup. Of course, there needs to be some level of cooperation in order for both players to want to keep playing, and for the game to maintain its coherence. For example, if Bertrand were to say something like, “Andy, how is it that you are putting a tea-bag into my cup when I know very well that you never have tea in your house?” this might signal the beginning of the end of the pretense episode.

Part of what makes it fictional that the tea that Andy has made is not disgusting is Andy’s and Bertrand’s evolving and mutual decision that the game Tea for Two be such that the tea is a delight. And part of what motivates this decision on the part of Bertrand is the way in which Bertrand interprets Andy’s intentions. Interestingly, it is only given
Bertrand’s knowledge that Andy lacks certain knowledge about how to make tea that allows him to interpret Andy’s intention that the pretense be one where they make delicious tea.

To see that this is true, consider the case where Bertrand plays *Tea for Two* with his English friend, Martin. Here, Bertrand assumes that Martin knows very well how to make good tea; in fact, he knows Martin to be a particularly fastidious tea drinker, and something of a tea connoisseur. Now consider what would happen in the game were Martin to pour the “milk” in directly after the “hot water”. Surely, Bertrand would sense something amiss. Given that Martin cares so passionately about tea, surely he would not prepare his tea in this careless way? In this case, it would be utterly appropriate for Bertrand to say something like, “Martin – this tea is revolting!” to which Martin might reply, “Oh Bertrand, I was just testing to see whether you still remembered what a good cuppa tastes like after all those years across the pond!” In complex, coordinated games of joint pretense, players must constantly keep track of what the other pretender takes to be true within the pretense, and what the other pretender takes to be true in real life. To a large extent, there must be mutual understanding and agreement about what belongs in the Belief Box and what belongs in the Pretense Box.

Usually, the boundary between the real and pretend realms is permeable only in the real-pretend direction: that is, pretenders will assume that many real beliefs that they have also hold true in the pretense (e.g. – the water is boiled when the electric kettle goes
“click”). In exceptional cases, however, items that originate in the Pretense Box find their way into the Belief Box.

Consider what would happen if Andy and Bertrand were to play *Tea for Two* a day or two after they had had tea for real at Andy’s place. Suppose that on that day, Andy had prepared tea in his usual inept way (by adding milk before the tea had properly steeped), but that Bertrand was too polite to say anything. Now, suppose that the next day when they play *Tea for Two*, Bertrand, bearing in mind that this is just a game, is less demure than he ordinarily is, and says something like, “Andy, this tea is a little weak. Maybe next time you should allow it to steep for a while.” In this variation of the story, Andy might very well add the item “One should always wait for tea to steep before adding milk” into his Belief Box. The etiology of this belief is rather interesting. It started in Bertrand’s Belief Box, was added to Bertrand’s Pretense Box in the manner prescribed by the Nichols and Stitch theory; from there it made its way into Andy’s Pretense Box, in accordance with the conventions of coordinated joint pretense, and then, (as a result, perhaps, of some Gricean implicature) it crossed the (normally non-permeable) boundary into Andy’s Belief Box.

The Nichols and Stitch theory takes part of the explanandum of a good theory of pretense to be the way in which, after an episode of pretense is over, items in the Pretense Box do not find their way into the Belief Box. I take it that part of the explanandum of a fully developed theory of pretence is the interesting and principled way in which items from
the Pretense Box *do* sometimes cross the Pretend-Belief boundary, and find their way into the subject’s belief box.

To summarize our findings thus far:

1. In the explanation of a pretend action, we must often advert to a belief that is *not* one of the beliefs that are added to the pretender’s Pretense Box.

2. Items in the Belief box often serve to regulate pretend action, determining whether the action will be symbolic, truncated, or full-blooded.

3. Barring unusual circumstances, successful pretenders are able to quarantine these regulating representations to the Belief box, preventing them from entering the Pretense box.

4. To account for the phenomena in *911* and *Mud Pies*, the functioning of Updater must be much more sophisticated than how Nichols and Stitch characterize it. The Updater must expunge from the Pretense box much more than representations that are logically contradictory to the pretense-initiating premise.

5. In complex, coordinated games of pretense, representations in one pretender’s Pretense Box can find their way into another pretender’s Belief Box.
III. The Case for Desire-like Imaginings.

Gregory Currie argues forcefully that imagination must have a systematic connection with desire if we are to make sense of fact that we are often emotionally moved by what we imagine (204). Currie points out that when one is emotionally moved by a real-life event, both a belief and a desire must come into play. Suppose that I believe that Sonia Gandhi will become the next Prime Minister of India. I will only be emotionally moved by this belief if I bear a related desire. Perhaps I desperately want it to be the case that Gandhi become prime minister; in that case I will feel elated. Or perhaps her taking power is the last thing I want to happen; in that case, I might feel dread. The emotion of elation or of dread is only explainable if the appropriate desire is present. So, too, with the imagination. When I am saddened by the fate of a fictional character, it is because I desire (within the scope of the imagining) that he not have such a fate. In the absence of that desire-like mental state, it would be very difficult to explain my sadness.

Recall that in the mental architecture that Nichols and Stitch posit, there is no conative analogue to the belief-like representations that exist in the Pretense Box. There is no box analogous to the Desire Box for desire-like imaginings that might motivate agents to act in pretend ways. Nichols and Stitch think that real desires can do the work. They argue that what motivates pretend action is the desire to act as we would if the things we imagine were true. In their lexicon, we desire to act in a way that is “similar to the way in which some character or object behaves in the possible world whose description is contained in the Pretense box” (2003, 37).
Taking up Walton’s case of children who pretend to make and eat mud “pies”, they argue that it cannot be the case that imaginary desires motivate people to act, since “if this imaginary desire has causal powers similar to the causal powers of the real desire with the same content, then the desire would produce the wrong behavior. An (imaginary) desire to eat some pie along with an (imaginary) belief that the gob of mud is a pie would presumably lead the pretender to actually eat the mud pies. But pretense behavior of this sort is rarely seen in children and almost never in adults.” (2000, 134)

I think that it may be the case that Nichols and Stitch misconstrue the nature of desire-like imaginings (or “pretend desires”). Pretend desires are generated from within an episode of imagination, and similarly, pretend desires make their contribution to motivation from within the episode of pretense. Responding to Nichols and Stitch, Gendler argues: “[Nichols and Stitch] mis-identify the relevant parallel: what the (imaginary) belief and (imaginary) desire lead to is an (imaginary) action – that is, an action that represents the eating of a (mud) pie, such as making certain gestures and sounds typically associated with eating. […] A circumscribed pretend-belief and a circumscribed pretend-desire conjoin to produce a correspondingly circumscribed pretend-action.” (2006, 5) I agree with Gendler that Nichols and Stitch have misidentified the relevant parallel. However, I do not think that it is necessary to characterize the pretend-belief itself, and the pretend-desire itself as “circumscribed”. The “circumscribed” nature of the pretend-action is best explained by the agent’s real beliefs. (Recall: In the variations on the mud pie example above, the explanation for why
the children’s eating was merely symbolic, truncated, or full-blooded was the children’s belief about what the props in the game were made of).

The “Inner Tension” Argument for Desire-like Imaginings

In *Recreative Minds*, Gregory Currie and Ian Ravenscroft make the case for desire-like imaginings by focusing on the spectator’s experience of watching a tragic play. Their chosen example is Shakespeare’s *Othello*. They contend that the spectator experiences an “inner tension” in viewing this work. On the one hand, he wants Desdemona to avoid her cruel fate; on the other hand, he does not want the play to end in a way other than Shakespeare intended:

> When I am sorry and upset about the fate of Desdemona, I am not sorry that this fiction has it that an innocent and good-hearted girl suffers a cruel fate. One might be sorry about that, deploring that there are fictions with such unhappy outcomes. This is not what at least many of us are sorry about; we are glad that Shakespeare’s fiction has it this way. (21)

Currie and Ravenscroft think that this “inner tension” in the spectator is evidence of a competition between real-life desires and desire-like imaginings: “Part of the inner tension one experiences in watching the play derives from the fact that we experience a desire-like imagining that Desdemona flourish, combined with a (genuine) desire that the play be one which will ensure that that desire-like imagining is unsatisfied.” (21-22) So,
according to Currie and Ravenscroft the “inner tension” is explained by two competing conations in the spectator’s psyche:

**Real Desire:** A tragic ending to a tragic play; that Desdemona be killed.

**Desire-like imagining:** A virtuous woman avoids a cruel fate; that Desdemona be saved.

In a review of *Recreative Minds*, Shaun Nichols acknowledges that the spectator of *Othello* might feel the sort of tension that Currie and Ravenscroft describe, but he denies that we need appeal to desire-like imaginings to explain this phenomenon. Nichols argues that ordinary, real-life desires can do the necessary work:

The tension we experience in watching *Othello* arises because we have conflicting *real* desires about the play. One desire is that the fiction have it that Desdemona be saved; the other desire is that the play be tragic. I both want it to be the case (fictionally, of course) that Othello not kill Desdemona, and I also want it to be the case that the narrative be tragic.26

(3)

So, according to Nichols, the “inner tension” that one experiences when watching *Othello* is best explained by two competing *real-life* desires:

---

26 I’m not sure that the best way to characterize the spectator’s real-life desire is “that the play be tragic”. I doubt that most spectators would be introspectively aware of such a desire (although admittedly introspection about desire is not incorrigible). However, I doubt that if Desdemona were to flourish in the play, most spectators would come away disappointed that the play failed to be tragic. Rather, I think they would be disappointed that the play failed to end the way Shakespeare wrote it (they bought tickets to see a play by Shakespeare!). But this is a minor point and nothing substantial rests on it as far as I can see.
Desire 1: That the play be tragic (and Desdemona die).

Desire 2: That Desdemona be saved in the fiction.

So which is it? Do we desire that Desdemona be saved in the fiction (Nichols), or do we desire-like imagine that Desdemona be saved (Currie and Ravenscroft)? It is very difficult to maintain a firm grasp on what this question means; it is even more difficult to adjudicate it. Part of the controversy seems to hinge on the issue of how “deeply” a reader or spectator “enters into” a fiction. Does the spectator merely have cognitions and conations about the fiction (Nichols), or does she generate cognitions and conations within the fiction (Currie and Ravenscroft).

One point in favor of the Nichols and Stitch view is that desire-like imaginings cannot be as easily willed into existence as make-beliefs can. To make-believe that I am Napolean, I just need to close my eyes and begin to imagine. But although I can imagine that I desire to win the battle of Waterloo, I cannot so easily desire-like-imagine to win the battle. One way to see that we do not have voluntary control over our desire-like imaginings is to notice that we cannot emote at will, even by imagining (I cannot get myself to feel the disappointment of having lost).

Still, my intuitions in this matter ultimately lie with those of Currie and Ravenscroft. However, I think that the case for desire-like imaginings is made stronger when we take up the case of pretend action (rather than that of the spectator in the theater). In “The
Aim of Belief” David Velleman makes the case for desire-like imaginings by examining the nature of pretend play in children:

An especially imaginative child may come up with his own way of pretending to be an elephant, but not by considering which behaviors would be most suitable to an elephant act, as if he were an impressionist honing some zoological schtick. Rather, the child’s method is to imagine being an elephant and then wait and see how he is disposed to act.” (2000 257)

Velleman argues that to suppose that a child’s pretense is always motivated by a belief and a desire denies that the child ever enters into the fiction of being something other than he is: “In order to enter into the fiction, the child would have to act it out; and in order to act it out, I think, he would have to act out of imagining it, not out of a desire to represent it in action. A child who was motivated by such a desire would remain securely outside the fiction, thinking about it as such – that it is as a fiction to be enacted” (2000, 257). While I find the picture that Velleman draws to be compelling, his language is rather abstract. If we accept Nichols’ account, must we presume that the children who pretend to be elephants stand “securely outside of the fiction”?

One thing to notice is that there is an asymmetry between the sorts of things that we desire about the fictions we engage in, and the sorts of things we desire within the fictions we engage in. We often desire that unfortunate things happen to us in pretend
games, or to characters in fictions. I might desire that an upstanding yet boring character in a film face some sort of personal crisis, just so that the film becomes more interesting. Or, in a game of pretense, a child might want their “safari” game to include a snake falling from the trees onto their heads. These desires about the goings on in fictional worlds are markedly different than the sorts of desires we normally have. (Normally, we would not desire that an upstanding person face a personal crisis, or that we have a dangerous encounter with a snake.) Our desires within fictions, on the other hand, more closely track our ordinary desires about the real world. My desires regarding Desdemona’s fate are relevantly similar to my desires about real people like her in analogous situations.

---

27 This example is from the contribution of Tyler Doggett and Andy Egan to the Online Philosophy Conference 2007: http://experimentalphilosophy.typepad.com/online_philosophy_conference/
CHAPTER FOUR
REALITY-DIRECTED AND FICTION-DIRECTED EMOTIONS

My endeavor should be directed to persons and characters supernatural, or at least romantic; yet so as to transfer from our inward nature a human interest and semblance of truth sufficient to procure for these shadows of imagination that willing suspension of disbelief for the moment, which constitutes poetic faith.

- Samuel Taylor Coleridge *Biographia Literaria*
Imagine that you are sitting in a darkened movie theater watching a scary movie about small, creepy, wily, and cruel gremlins. The infernal creatures appear unstoppable. Spreading out across the country, they occupy the Empire State Building, the Capitol, and White House. They even peek their evil little heads out from behind the gigantic Hollywood sign. They are thirsty for power and brimming with ill will. One of the gremlins turns toward our brave and handsome protagonist with a maniacal and crazed look.

Your face betrays a paroxysm of terror. You twitch in your seat, bite your nails, claw at your armrest, grab whoever’s next to you….

The case described above illustrates well what philosophers have dubbed “the paradox of fictional emotions”. Often, a decisive moment in the development of a philosophical theory occurs when a paradox is generated out of some phenomenon that looks utterly banal or at least very familiar. Philosophers have long puzzled over how it is even possible for a person to get worked up about people and states of affairs that she knows to be unreal. To put the problem slightly differently, how is it that the inhabitants of fictional worlds are able to produce sighs, moans, and tears in our world? How can you pity someone that you know very well does not exist? How can you dread something that will never happen? The paradox of fictional emotions can be expressed as the conjunction of three distinct elements; while each element has some plausibility to it, it is not possible that all three are true.
1. In order for us to respond emotionally to what we learn about various people and situations, we must believe that those people and situations exist. (For example, I will not pity your sister unless I believe that you have a sister. I will not fear the imminent recession unless I believe that a recession is imminent.)

2. These “existence beliefs” are absent when we knowingly engage with fictions. (When we watch a horror movie, we do not believe that the “monsters and ghouls” really exist. When we read a tragic novel, we do not believe that the suffering heroine really exists.)

3. We are in fact emotionally moved by fictional characters and situations. (Often at horror movies we are afraid. Often when we read tragedy we pity the heroine.)

The second and third elements of the paradox are deliverances of common sense. Many “ordinary” subjects (that is, non-philosophers) who are acquainted with narrative fictions will be inclined to endorse them, especially if they do not already have prior commitments to a theory of fiction. The first element of the paradox is a substantive philosophical thesis about what it means to respond emotionally to something. In this chapter I propose to side with the volk and to abandon the substantive philosophical thesis. Many philosophers have seen the solution to the problem quite differently.

Irrealists about fiction-directed emotions maintain that the apparent emotions that we

---

28 The paradox of fictional emotions is at its most paradoxical for emotions elicited by tragedy, viz. grief and fear. There is less of an air of paradox for emotions such as mirth and disgust. For example, it seems much less puzzling that we giggle at Seinfeld’s shenanigans than that we weep at the fate of Anna Karenina. Also, there is a more robust isomorphism between the things we find funny or disgusting in the real world and the things we find funny or disgusting in fictional worlds than there is between the things we find fearful or sorrowful in the real world and the things we find sorrowful or fearful in fictional worlds. For example, I may be surprised to find how much pity I am made to feel for a character, but I am never surprised at what jokes I find funny in a fiction. That said, non-mimetic elements such as lighting and laugh tracks can sometimes mitigate or intensify the mirth or disgust we experience with fictions.
experience in response to fictional worlds are not real emotions. Irrationalists about fiction-directed emotions (following Socrates of Plato’s Ion\textsuperscript{29}) hold that although the emotions that we feel when we get “caught up” in a fiction are real, we are irrational for having such emotions.

The view that appreciators\textsuperscript{30} of fiction are irrational for emotionally engaging with fictional narratives seems to smack of a kind of philistinism. Surely it cannot be true that those of us who are most moved by Tolstoy and Dostoyevsky, or by Bergman and Antonioni, are peculiarly irrational for being so moved? Moreover, surely there are a number of fine discriminations to be made in assessing the rationality of our emotional response to art? It is a very different thing to experience a sense of foreboding when watching Hitchcock’s The Birds than to experience that same emotion response when watching Sesame Street. Any view with wholesale irrationalism as a consequence would seem to be predicated on some deep misunderstanding of the nature and purpose of

\textsuperscript{29}Soc. Well, Ion, and what are we to say of a man who at a sacrifice or festival, when he is dressed in holiday attire and has golden crowns upon his head, of which nobody has robbed him, appears weeping or panic-stricken in the presence of more than twenty thousand friendly faces, when there is no one despoiling or wronging him;—is he in his right mind or is he not?

Ion. No indeed, Socrates, I must say that, strictly speaking, he is not in his right mind.

Soc. And are you aware that you produce similar effects on most spectators?

Ion. Only too well; for I look down upon them from the stage, and behold the various emotions of pity, wonder, sternness, stamped upon their countenances when I am speaking; and I am obliged to give my very best attention to them; for if I make them cry I myself shall laugh, and if I make them laugh I myself shall cry when the time of payment arrives.

\textsuperscript{30}Throughout the chapter I will use the term “appreciator” as compact way of saying “viewer, reader, or listener”. Despite the word’s ugliness, it serves as convenient shorthand, and has gained wide use in the relevant literature.
mimetic art\textsuperscript{31}. In this chapter I will not take up topic of the rationality of emotion in detail, although I will briefly discuss, and try to defuse, some of the motivation for holding irrationalism about fictional-directed emotions\textsuperscript{32}.

The roots of irrationalism can be traced to the deeply held (yet mistaken, I will claim) intuition that a warranted emotion requires a belief in the existence of its object, the first element in the paradox described above. To understand the source of this intuition, consider the following example. Suppose that I were to tell you a story about Shaun, a friend of mine who had a particularly hard-scrabble childhood, with a mother who was a drug-addict and a father who was physically abusive. Shaun went to a school on the wrong side of town where only a small proportion of the students finish with high school diplomas and even fewer go on to college. Despite his awful home life and the lack of decent schooling available to him, Shaun’s curiosity, passion for learning, and hard work resulted in his graduating with excellent grades. He was the darling of his teachers, and the pride of the school district. What’s more, against all odds, Shaun managed to gain admission to Brown University, the only admit from a Providence public school in his graduating year.

\textsuperscript{31}This is not to say that the nature and purpose of mimetic art is altogether clear. Martha Nussbaum (1986) and Susan Feagin (1983) have described the way in which moral character can be refined by fiction-directed emotions. Gendler and Kovakovich (2005) have applied empirical insights from Damasio, Harris, and Bechara to argue “simulated emotions” play a central role in practical reasoning.

Supposing that you were convinced by the story, that you held certain attitudes regarding justice and merit, and that the narrative was told in a way that was detailed and vivid rather than schematic and condensed, it is likely that you would experience first pity for Shaun followed closely by admiration. Now suppose that I were to tell you that my friend Shaun never existed, and that I had made the whole thing up. There is a strong case to be made that you (rationally) ought to be disabused of your feelings of pity and admiration. (And you will likely feel something like anger or resentment toward me for leading you on, depending on how caught up in the story you became). Were you to continue feeling pity and admiration for Shaun, it would seem that something had gone seriously amiss; you would be less than perfectly rational.

This, I think, is the basic intuition underwriting irrationalism. While I think the intuition points to something that is real, the hard work comes in interpreting it. The example above, and other examples of the sort presented by Colin Radford, only serve to demonstrate that if an emotion is predicated on a belief, and if one discovers that the relevant belief is false, then it is irrational to go on experiencing the emotion. What examples of this sort do not demonstrate is that an emotion is necessarily irrational if it is not based on a belief. You can be open to the (commonsense, I think) view that one’s emotional life should track changes in one’s stock of beliefs, without being at all committed to the (tendentious, I think) view that emotions that have as their cognitive

---

33 For other examples of this stripe, see Radford (1975).
constituent imaginings, considerings, hypotheses, construals, or seeings-as, are thereby irrational.\textsuperscript{34}

Irrealists such as Kendall Walton (1990, 1997), Gregory Currie (1990), Jerrold Levinson (1997), and Stacie Friend (2006) reject the third element of the paradox. Rather than maintaining that we are irrational for experiencing fiction-directed emotions, they deny that we experience fiction-directed emotions at all. That is, they deny that we are really afraid at the movies, or that we really pity a novel’s heroine. Rather, they posit that we imagine that we are experiencing these emotions. By adopting this strategy, irrealists are able to maintain that emotions require beliefs in their objects, and that we do not believe (temporarily or otherwise) that fictional objects really exist.\textsuperscript{35}

The canonical and most detailed irrealist account is due to Kendall Walton. Walton’s irrealism about fiction-directed emotions is best understood in context. In *Mimesis as Make-Believe* he develops general theory of representational art whereby works of fiction

\textsuperscript{34} Radford’s second major argument for irrationalism is the similarity of fiction-directed emotions to certain types of phobias (which are not based on beliefs). For a compelling response to this line of argument, see Richard Joyce (2000). Joyce objects that the absence of control, counter-productiveness and maladaptiveness found in phobias is generally absent in the case of fiction-directed emotions, and so the cases are not analogous.

\textsuperscript{35} Here a number of semantic and metaphysical puzzles arise. For a proposition of the form, “I fear ø” or “I pity ø” or “I hate ø” or “I love ø” it is conventionally thought that we quantify into the name position, committing us to the existence of whatever is referred to by the name. It is not hard to see why this might be problematic when ø is, say, a gremlin. Meinong’s famous solution to this problem was to maintain that there are some things that do not exist (see Parsons 1980, Crittendon 1991). Another tact is to hold that fictional characters exist as real entities in other “worlds” that are causally and spatiotemporally isolated from our own (see Lewis 1978). Needless to say, neither of these solutions is for ontologically faint of heart. Van Inwaagen (1977) and Salmon (1983) treat fictional characters as abstract entities; fictional characters are also sometimes treated as abstract kinds or sets of properties (Wolterstorff (1980), Larmarque (1983)). It is far beyond the scope of the current project to undertake a cost-benefit analysis of each of these strategies.
function as \textit{props} in games of make-believe. Just as dolls and toy trucks prescribe imaginings of real people and real trucks in children’s games of make-believe, works of fiction are designed to prescribe imaginings about their content. Imagining what is prescribed is what it means to participate in the game authorized by the work.

In engaging with a fiction, we imagine that we are learning about actual facts, and we respond to the fictional events from within this pretense. When I read Edith Wharton’s \textit{The House of Mirth} I imagine myself to be reading a nonfiction report of actual events. I imagine discovering that there is really a person named Lily Bart who is gradually, painfully, and tragically loses her social position in hypocritical New York society. I imagine \textit{of} my reading the novel that this very experience counts as learning about actual people and situations. Thus, for Walton, all imagining involves a kind of self-imagining. As he puts it, “the minimal self-imagining that seems to accompany all imaginings is that of being aware of whatever else it is that one imagines.” (1990, 29). It is essential that part of what is being imagined has a content referring to oneself. When I imagine some fictional truth, I imagine myself believing it.\textsuperscript{36}

On Walton’s account, my psychological/emotional/physical response all constitute aspects of my participation in the relevant game of make-believe. Suppose that in reading \textit{The House of Mirth} my imagining learning of Lily Bart’s unjust fate is sufficiently vivid that a discernible lump forms in my throat and tears well up in my eyes.

\textsuperscript{36} The supposition that one’s psychological state of imagining is part of the content of one’s imagining is itself controversial. Insofar as I can imagine being deceived, it seems a I can imagine something that does not include as part of its content my believing it. For a critique along these lines, see Moran (1994, 87-92)
I know, of course, that in reality Lily never existed, and that the story is altogether made up. Nonetheless, it is in part facts about my feelings (a psycho-physical state that Walton terms a “quasi-emotion”) that make it fictionally the case that I experience pity. Just as it is only fictionally the case that I am reading a true report of Lily’s suffering, it is only fictionally the case that I pity her.

We are now in a better position to understand the motivation behind the surprising claim that in engaging with fictions appreciators do not experience genuine emotions. Walton concedes that this view might seem at odds with the most pervasive and striking feature of our experience of mimetic arts, the very feature that leads Aristotle to remark in his Poetics that, "the plot [of a tragedy] must be structured […] so that the one who is hearing the events unroll shudders with fear and feels pity at what happens." (1970, 53 b 5) However, Walton avers that Aristotle’s words are best not taken literally. To illustrate his position, Walton proffers the by now canonical example of Charles:\footnote{37While compelling, Walton’s example is not representative of the typical way in which fictions frighten us. Generally, we are not afraid for ourselves, but for some character or characters in the fiction.}

Here is an example of the most tempting kind: Charles is watching a horror movie about a terrible green slime. He cringes in his seat as the slime oozes slowly but relentlessly over the earth, destroying everything in its path. Soon a greasy head emerges from the undulating mass, and two beady eyes fix on the camera. The slime, picking up speed, oozes on a new course straight toward the viewers. Charles emits a shriek and clutches
desperately at his chair. Afterwards, still shaken, he confesses that he was “terrified” of the slime.” (196)

The reasons why one might attribute an emotion (specifically, fear) to Charles should be obvious. First, just try telling Charles that he was not “really” afraid. Charles explicitly avows that he was “terrified”, and if Charles is like most people, he will stick to his guns in face of skepticism. Second, Charles evinces many of the involuntary psycho-physiological symptoms of fear: the cringing, the shrieking, the clutching, the sweating, the trembling. If Charles’ first-person reports as well as behavioral manifestations unambiguously point to fear, why should we doubt that he is in fact experiencing fear?

Despite these indications, Walton resists this move. Rather, he maintains that the evidence before us is only sufficient to impute to Charles something that he dubs “quasi-fear”:

Was he terrified of [the slime]? I think not. Granted, Charles’s condition is similar in certain obvious respects to that of a person frightened of pending real-world disaster. His muscles are tensed, he clutches his chair, his pulse quickens, his adrenaline flows. Let us call this physiological-psychological state quasi-fear. But it alone does not constitute genuine fear. (196)
So what is the extra ingredient that differentiates real fear from quasi-fear? It is certainly not *intensity*: Charles is shaking in his boots! To cleave us from the seemingly commonsense view that Charles fears the slime, Walton brings our attention to some features of Charles’s experience that he thinks speak against characterizing Charles’s experience as one of real fear. First, he notes that Charles is not naïve: he knows that the slime is not really threatening him and that he is in no danger. (If he did believe the slime were real, he would probably alert the authorities!) Charles cannot be said to be afraid, Walton argues, because Charles does not have the *belief* that the slime endangers him. Furthermore, Charles shows no inclination to leave the theater or to call the police. In fact, Charles does not display any of the action-tendencies that we usually associate with being in a state of fear. All of his reactions to the fictional slime are non-deliberative non-actions (palms sweating, muscles tensing, etc). The nature of Charles’s mental state, Walton maintains, is non-motivating, and he concludes that “[f]ear emasculated by subtracting its distinctive motivational force is not fear at all.” (202) So, Charles is not really afraid; rather, to use Walton’s terminology, *it is fictional* in Charles’s game of make-believe that he is afraid, and part of what makes it fictional are the psycho-physical reactions that constitute quasi-fear.38

To be sure, Walton does not disallow that when you watch a horror movie, you *feel* something that is very similar to real fear. What Walton simply denies is that these

---

38 Throughout this chapter I presuppose a broadly cognitive-evaluative theory of the emotions, whereby emotions are individuated by the subject’s evaluation of the emotion’s intentional object (Lyons 1980, de Sousa 1987, Greenspan 1988, Roberts 1988). I will not argue for the correctness of this account; however, the account is presupposed by most of the interlocutors in the realism/irrealism and rationalism/irrationalism debates in aesthetics.
psycho-physical responses are sufficient to constitute real fear. Real fear has three necessary components:

1. The affective: The feeling of fear, and its physical manifestations
2. The cognitive: The belief that you are endangered
3. The motivational: The desire to act to as to secure your safety.

According to Walton, Charles’ experience of watching the green slime lacks the cognitive and motivational components necessary to attribute to him real fear. At no point does Charles believe that he is endangered (on the contrary, he knows that he is safe and sound in a movie theater). Similarly, Charles is not motivated to flee the green slime coming towards him. Only in very rare cases do people actually flee the cinema. In critiquing Walton’s view, I will first assess the belief condition and then the motivational condition on genuine emotion.

I will use term judgementalism to refer to the view that in order to experience an emotion, we must have a belief about the object of that emotion39. On this view, (genuine) fear requires the belief that someone or something endangers someone or

39 I will not attempt here to resolve some thorny questions about the nature of emotions (e.g. whether all emotions are object-directed in the way that fear and pity seem to be, and whether for that matter emotions form a natural class at all (see Amelie Rorty (1978) and Paul Griffiths (1997) for an answer in the negative and Robert C. Roberts (2003), chapter 1 for an answer in the affirmative). For the purposes of this chapter, I will make a few assumptions about emotions that I hope my readers will not find unpalatable. First, I will assume that emotions are mental states that have a characteristic phenomenology. Second, I will assume that a rough and ready distinction between emotions and moods is that an emotion has an intentional object whereas a mood need not.
something that you care about (which may include yourself). (Genuine) pity requires the belief that someone or something you care about is suffering.\(^{40}\)

Following Noël Carroll (1997), I will use the term **thought theory** to refer to the view that the cognitive component of emotion can consist of a variety of different kinds of *thoughts*, which may include beliefs, as well as unasserted thoughts, construals, seeings-as, entertainings, imaginings, etc. Both judgementalism and thought theory are types of cognitivism insofar they both assert that emotions must have a constitutive cognitive component, but thought-theory is more ecumenical than judgementalism is. The candidate emotions that satisfy judgementalism form a proper subset of the candidate emotions that satisfy thought theory.

I think that judgementalism is false. By now there is a host of compelling examples in the literature that indicate that emotions do not require *beliefs* about their objects. One of the earliest and most persuasive examples is due to Patricia Greenspan’s (1981, 162; 1988, 18). Greenspan describes a subject who, having been bitten by a rabid dog in the past, is now deathly afraid of harmless Fido, a harmless old hound that is well known to him. Although he feels fear in the presence of Fido, he knows very well that Fido will not hurt him. This belief is in evidence when he doesn’t shield his children from Fido, or run away screaming. What is interesting about the example is that fear described *does* seem to have a cognitive content; it not simply the physiological and phenomenological aspect

---

\(^{40}\) Walton himself remains agnostic about judgmentalism.
of fear. Yet, it seems wrong to impute to the subject a belief that he is endangered by Fido:

Instead of supposing that his beliefs come into momentary conflict whenever Fido comes near, it seems simpler, and preferable from the standpoint of rational explanation, to take this as a case where emotions parts from judgment. It exhibits the tendency of emotions, in contrast to a rational agent’s beliefs, to spill over to and to fix on objects resembling their appropriate objects in incidental ways. (Greenspan 1988, 18).

A variety of phobic responses seem to tell against judgmentalism. Consider: The observation deck of the CN Tower in Toronto features a glass floor which visitors are encouraged to walk across. Often you will find young children gleefully jumping up and down on the glass, while their parents huddle toward the concrete frame, terrified to step across. It is highly unlikely that the adults in question have conflicting beliefs about whether the glass is safe to walk on. If they did, they would never allow their children to walk across (on the minimally charitable assumption that adult visitors to the CN Tower do not tend to be negligent parents). But it seems equally undeniable that these parents experience fear: their hearts are beating quickly, their palms are sweating, they show a marked reluctance to advance any further onto the floor41.

41 Tamar Gendler (forthcoming) discusses a variety of examples of this sort, including reluctance to drink from a glass in which a cockroach was placed, even though the glass has since been completely sterilized, or the reluctance to eat fudge in the shape of feces. To
Admittedly, the way in which we should understand phobias is a matter of some philosophical and psychological controversy, so we would do well with some other types of examples. Noël Carroll (2001) describes well the way in which merely entertaining a certain kind of thought is often sufficient to generate the experience of fear or squeamishness in a subject. Unlike a belief, a thought has a propositional content that is unasserted.

While cutting vegetables, imagine putting the very sharp knife in your hand into your eye. One suddenly feels a shudder. You need not believe that you are going to put the knife into your eye. Indeed, you are not going to do this. Yet merely entertaining the thought, or the propositional content of the thought (that I am putting this knife into my eye), can be sufficient for playing a role in causing a tremor of terror. (234)

There are many cases drawn from everyday life where the mere thought of something is sufficient for generating fear. When I sit on the small turbo-prop airplane that goes from T.F. Green airport to Toronto, merely focusing my attention on the thought of violent turbulence is enough to make me close my eyes and firmly grasp my tray-table.

There are many types of counterfactual imaginings outside of the context of deliberate “fictions” seem to suffice as the cognitive constituents of emotions.

account for the “belief-behavior mismatch” Gendler posits the existence of a mental state alief: “so-called because alief is associative, action generating, affect-laden, irrational, agnostic with respect to content, shared with animals, and developmentally and conceptually antecedent to other cognitive attitudes” (8).
Consider the Holocaust survivor who experiences a feeling of dread when he imagines what the world might have been like had Hitler won the war. In order to experience this emotion, he need not have any beliefs about Hitler coming back from the dead, or about the likelihood of a Nazi resurgence. He just needs to think of the unthinkable.

The phenomenon of engaging in entire imagined conversations with real people is familiar to many. Merely imagined exchanges between friends and lovers can alter the contours of very real relationships. Upon having a solitary experience that is for some reason remarkable, one might imagine trying to recount it to a close friend or spouse. How one imagines this conversation might succeed or fail can be revelatory about the relationship itself, and can provoke feelings of intimacy (“he’s the only one who might understand”) or estrangement (“he would never get it”).

Not all examples of emotions based on unasserted thoughts require thoughts about possible future states of affairs. Often when we think back to close calls and near misses, we experience fear, or something akin to it. Consider the case of a driver who on her sleep-inducing morning commute to work narrowly avoids veering off into the path of an oncoming tractor-trailer. Safe in her office cubicle, she knows with certainty that the danger is no longer present. However, she cannot extinguish the memory of it, which she experiences as a painful jab. Throughout the day she finds herself cringing involuntarily, even though she is well aware that the danger is safely in the past.
Even in cases in which a person learns only in abeyance that she was in danger or in proximity to danger, fear is often a typical response. Consider a case where, upon disembarking from a flight, a woman learns that the flight that left two hours after hers had a serious mechanical malfunctioning and came to a bad end. Even though the woman herself was never actually in danger, the danger faced by her “possible world counterpart” who boarded the ill-fated plane is itself sufficient to generate fear in her. Even if we change the case so that despite mechanical malfunctioning the other plane ultimately lands safely, it would not be surprising if the woman still experienced fear.

Fear is not the only emotion that we experience we when think about past, future, possible, or impossible states of affairs. Often we experience embarrassment in the absence of any belief that we have done something humiliating. Just as the phenomenon of the “near miss” can occasion fear, it can also spur mortification. It is a common experience to think back on something that you came close to saying or doing, and to feel acute embarrassment. Consider the following case. Upon meeting for the first time a young faculty member with an exceptionally boyish face, you labor under the misapprehension that he is an advanced undergraduate student. Thankfully, nothing that you say or do betrays your misunderstanding. Still, when days later you learn from another colleague that the person you met was in fact an assistant professor, it would not be surprising if you now felt mortified about all things that you almost said to him42.

42 Here it is interesting to note that one can also be made to feel vicarious mortification on the part of a fictional character (the BBC series *The Office* is particularly adept at eliciting such emotional responses). It would not be surprising if we found examples where audiences were made to feel vicarious embarrassment at the near embarrassing predicaments of
Indeed, thoughts directed toward mere *possibilia* form a significant component of our emotional lives. Once we recognize this fact, it should seem less mysterious that we experience emotions that are directed towards *fictional* characters and states of affairs. If we are willing to acknowledge the reality of this wide range of emotions that are not constituted by belief, I think we should also acknowledge the reality of fiction-directed emotions. As to the question of *just why it is* that we respond emotionally to the non-actual, I think we can safely leave this line of inquiry to the cognitive psychologists and neuroscientists. For our purposes, the important thing to notice is the pervasiveness of the phenomenon, and the fact that fiction-directed emotions are in a sense unexceptional.
Motivation

In a widely quoted passage, Walton famously contends that, “fear emasculated by subtracting its distinctive motivational force is not fear at all” (1990: 201-202) I cannot really pity Anna Karenina since, the thinking goes, I am not in the least moved to help her. Nor can I really fear the slime, since I do not find myself fleeing the theater and notifying the authorities.

While Walton is no doubt correct in observing that there is often a motivational asymmetry in our response to fiction and reality, it is important to notice there are some (uncontroversial) genuine emotions that may not motivate us to do anything in particular. Some emotions may simply be experienced as too faint or too fleeting to issue in any sort of action-tendency. I may experience mild embarrassment when I come back to my car and notice that my parallel parking job is less than a command performance. But as I drive off, the emotion quickly dissipates and is soon forgotten. Of course, we could imagine scenarios in which my embarrassment leads me to set off with a compensatory swiftness, or in which my mortification leads me to treat my significant other with uncharacteristic brusqueness. But if the emotion is dim and transitory, there is a good chance that it will not result in any action at all. And this is perfectly consistent with my feeling embarrassed.

Even strong emotions do not always move one to act. Consider, for example, admiration. I may admire Lance Armstrong for his indomitably will and for his stamina. I might be motivated to buy a yellow “livestrong” bracelet. I might be motivated to be more
disciplined myself. But then again I might not. My admiration for him might not motivate me to do anything at all. And this is perfectly consistent with my admiring him.

Similar cases can be cooked up for pity. When I watch the nightly news and learn about a multi-car pile up on Route 95, I may be moved to feel pity for the victims. But just as I can do nothing for the inhabitants of fictional worlds, so also I am powerless to help inhabitants of this world after they have died. (Perhaps if I were of a different turn of mind, I might pray for their souls, but I am not).43

Sometimes emotions fail to motivate because of a belief that the emotion is not warranted, and the person experiencing the emotion is distinctly aware that the emotion is unwarranted. For example, Stewart may feel a sharp pang of jealousy in seeing his wife engrossed in deep conversation with his intelligent and handsome colleague. But his knowledge of his wife’s abiding love and faithfulness prevents him from acting on this feeling. Of course, we could imagine scenarios in which Stewart’s sublimated jealousy leads him to act with cool cruelty or passive aggression, but this need not be the case.

And a kind of self-control that is borne of self-awareness is consistent with experiencing jealousy, even strong jealousy. Indeed, even intense emotions that a subject believes to be warranted and which ordinarily issue in some sort of action tendency might fail to motivate because the agent is depressed or anxious. Severe depression can render

---

43 More examples of genuine emotions directed at states of affairs that we can do nothing about are found in Matravers (1998) and Gaut (2003). Matravers proffers this compelling case: “In writing Robinson Crusoe, Defoe drew heavily on the journal of the real-life castaway, Alexander Selkirk. A reader of Selkirk’s journal could surely be moved to emotion at his suffering. However, the fact that Selkirk was marooned in 1704 (and died in 1721) makes it impossible – really impossible – for the reader to help him […]. Hence, if the absence of such a connection [to action] is no reason to deny that we feel emotions towards Alexander Selkirk, it is not reason to deny that we feel emotions towards Robinson Crusoe either.” (69)
someone catatonic and inert even in the face of intense fear or sympathy. It can sometimes be the case that an emotion that characteristically issues in an action is drained of its motivation force.

Finally, an emotion can fail to motivate simply because the agent cannot conceive of any action that would count as responding to the emotion. If I read a detailed and vivid account of the trial and execution of Joan of Arc, I may be moved to profound pity for her. But I am not motivated to try to help her. Barring backward causation or time travel, I have no instrumental beliefs about how I could help her. Similarly with fear, if I read a vivid historical narrative of Spanish Civil War, a historical period so barbaric that learning of its details induces real trepidation, I am not motivated to do anything in particular.

It might be argued that any emotion that I experience causes me at least to have certain dispositions, even if these dispositions are not activated. For example, if someone were to ask, “What do you think of Lance Armstrong” I would be disposed to tell them that I found his steeliness and his stamina impressive. Or if a therapist were to ask Stewart to recount any instances of jealous feelings in his relationship, he may be disposed to tell her about the episode with his handsome colleague. But this is surely not the kind of motivation that Walton is concerned with, since fiction-directed emotions will also lead me to have these sorts of dispositional action-tendencies. Reading a memoir of the Holocaust may dispose me to be concerned about the plight of Darfurian refugees. And I
would be very surprised if becoming engrossed with Ayn Rand’s *Atlas Shrugged* has not disposed some to vote Republican.

Perhaps Walton’s point is that when a fiction-directed emotion and a reality-directed emotion are of a similar kind, the reality-directed emotion tends to be more potent in its tendency to motivate. So, while pity for my brother motivates me to help him, pity for Anna Karenina does not motivate me to help her. And while my fear of the undertow motivates me to swim back to the beach, my fear of the fictional green slime does not motivate me to flee the theater. While I agree that such asymmetries exist, I do not think that the best explanation of them lies in fact that one kind of emotion is fiction-directed and one kind of emotion is not. Rather, the reason why I am not motivated to help Anna or flee the slime is that I am simultaneously aware that neither Anna nor the slime really exist, and so nothing that I can do could possible count as helping Anna or fleeing the slime. This is not to say that there is no conation present in my psychological economy with the content “flee the slime”. In fact, many things I that I do in my seat evince this mental state. I cringe, bite my lip, cover my eyes with my hands (an action that is at least symbolic of flight, much like that arational actions discussed in Chapter One and Two). But the characteristic fleeing response to fear is obviated by a concomitant awareness that the object of fear is not real. This is not so different from the characteristic helping response to pity being obviated by the awareness that it is impossible to help (as in the Joan-of-Arc case), or the characteristic lashing out response to jealousy being obviated by the awareness that jealousy is not warranted. Indeed, the characteristic action-tendencies associated with emotions maybe intensified, diminished, or completely extinguished by a
variety of attitudes and perspectives the subject may have on the object of the emotion. The subset of emotions that are fiction-directed are not exceptions to this pattern.

Still, one may, like Radford and Plato, be possessed of the conviction that there is nonetheless something suspicious and out of the ordinary about strong emotions that are directed toward (mere) fictions. One way to bring out this intuition is to make appeal to the intrinsic unpleasanthness of certain kinds of emotions like fear. If spectators feel real fear when they watch a horror film, why do they keep coming back for more? This can seem paradoxical. Surely they can’t all be masochists? [cite Radford or Levinson here]

It is not altogether clear what it means to say that fear is intrinsically unpleasant. But fear certainly isn’t exclusively unpleasant, and this is all I require for my purposes. As Richard Joyce (2000) has pointed out, sometimes fear may appeal as a kind of “arousal jag”. If one feels uninterested, or deadened, or bored, fear can function as an antidote. Bungee jumpers and mountain climbers seek out and delectate fear in carefully controlled conditions. Indeed, they might only attempt to climb a mountain or dive from a dam if it is really daunting. Lonely Planet would probably sell a lot fewer “adventure travel” guides if it were true that fear is exclusively unpleasant. Fear in the right amounts and under the right conditions can be like sourness or bitterness in the right amounts. To experience uncontrollable and unmitigated fear for prolonged periods at high intensity
would make life a living hell. But for some personalities, just enough fear, at the right time, under the right conditions, can render life more interesting.

My strategy thus far has been one of demystification. My aim has been to demonstrate, by way of example, that fiction-directed emotions and reality-directed emotions are in fact not so different from each other in their cognitive and motivational aspects. It is not really so strange or exceptional that that we experience emotion in the absence of a belief in its object, or that we can be profoundly moved without being motivated to act, or that people regularly seek out frightening or vicariously embarrassing experiences. But even if you think that these kinds of phenomena are still somehow strange or exceptional, the important thing to notice is that they are not restricted to our experience of fiction.

In the final section of this chapter, I want to describe some of the ways in which our experience of fictions often involves a complex and sometimes messy inter-mixing of beliefs and imaginings. I will argue that the emotions that we experience as a result can only be properly accounted for by appealing to both belief and imagining. Fiction-directed emotions of fear, pity, suspense, disgust, and admiration can be intensified or mitigated by thoughts and beliefs that are directed toward the real world. Conversely, reality directed emotions can be influenced and colored by our experience of fiction.

---

44 Vicarious embarrassment is another emotion that is most often unpleasant, but which can be appealing in the right amounts and under the right circumstances. (For instance, you might feel vengeful satisfaction in witnessing your rival squirm with humiliation). In the realm of fiction, the television series “Curb Your Enthusiasm” and “The Office” capitalize on eliciting cringe-inducing responses. They allow the viewer to imaginatively indulge in scenarios that are so mortifying that they would be unbearable, if not for the awareness that none of it is real.
A useful point of departure for this line of argument is an analysis of fiction-directed suspense. Kendall Walton finds confirmation of his theory of fiction in our experience of suspense and surprise in fictional works. He notes an interesting phenomenon whereby works “manage so often to survive multiple readings or viewings without losing their effectiveness.” (259). It is initially puzzling that suspense and surprise in fictions can survive our knowledge of how things turn out. When a child re-reads Jack in the Beanstalk for the umpteenth time, she knows with some certainty that Jack will safely escape with Ogre’s golden harp. Yet again and again she feels the same gripping suspense when the Ogre discovers Jack and chases after him. The key to solving this puzzle, Walton thinks, is to mark a distinction between what we know to be fictional and what fictionally we know. The child knows that in the fiction Jack will escape; however, fictionally she does not know whether or not Jack will get caught. “Fictionally she is genuinely worried about his fate and attentively follows the events as they unfold. It is fictional in her game during a given reading or telling of the story that she learns for the first time about Jack and the Giant. […] It is the fact that fictionally she is uncertain about the outcome, not actual uncertainty, which is responsible for the excitement and suspense of her experience.” (261) Although I will not draw the all the same conclusions that Walton does⁴⁵, I think that in his treatment of suspense and surprise in fiction, Walton has

⁴⁵ One thing to notice is that this phenomenon is not exclusive to our experience of fictions. Consider a case when a grandfather recounts to his grandchildren the story of his and his wife’s escape from a war, and their subsequent settlement in America. One can well imagine that these children experience repeated cycles of fear, suspense, and relief on multiple tellings of the story, just as they would in the context of a fiction. The pattern that Walton notices seems to me to issue more from the nature of narrative than from fictionality.
cottoned on to a phenomenon that is key to our understanding of how the imagination works in engaging with fictional works.

For it to be possible that we experience repeated cycles of suspense and relief even when we know with a high degree of certainty how things will turn out, we must have a considerable capacity for quarantining some of our beliefs so that they do not interfere with our emotional response to imaginings. In the case of the child listening to the story of Jack and the Beanstalk, the child is able to quarantine her knowledge that Jack will escape unscathed, which is what allows her to experience repeated cycles of suspense and relief. This quarantine, however, is clearly imperfect, for it is difficult to deny that most of the time, the suspense and surprise we feel on a second or third reading is discernibly less intense than the suspense and surprise we feel on the first. Were this not the case, “spoilers” would not be so-named and we would not get annoyed at film and book critics who give away endings in their reviews. To put the point in Walton’s terms, it is not always possible to keep what we know about a fiction from infecting what fictionally we know. In the case of “spoilers”, the appreciator of the fiction may feel cheated of suspense because advance knowledge about the plot is foisted upon him. To the extent that he is unable to quarantine the knowledge he has acquired about the work’s dénouement, he will be robbed of the uncertainty that gives the work its formal structure. To put it in Walton’s terms, he will not be able to keep separate what he knows to be

---

Walton’s choice of the “Jack and the Beanstalk” story is well chosen. It seems me that children have a far greater capacity to experience suspense on repeated readings or tellings than do adults. Perhaps this indicates a greater capacity on the part of children to quarantine meta-beliefs about a fiction from beliefs that they have from “within” the fiction. Or it might just be that children have a preference for repetition.
fiction from what fictionally he knows. Of course, the extent to which spoilers spoil
varies greatly from subject to subject. There exist people who routinely read the end of a
novel before reading the beginning and the middle. Conversely, there are people who
clap their hands to their ears whenever anyone so much as mentions a film they have not
yet seen. There is probably no single explanation to why any particular individual falls to
one side of this spectrum or the other. Sometimes the explanation will reside with the
person’s capacity for quarantining her knowledge about the fiction while she “inhabits”
the fiction. For other people, the explanation will reside with the extent to which they
value and appreciate the cultivation of suspense in fictional works.

Just as there are instances in which an appreciator aims to preserve a sense of suspense
when experiencing a fictional work, so also there are instances in which appreciators of
books and films aim to mitigate the feelings of suspense and fear generated by
imaginatively engaging with fictions47. Consider the sort of child who has little
experience with thrillers and horror movies, but is dragged along to the latest blockbuster
by some insistent older friends who help sneak him into the theater. As suspense and fear
mounts, he looks askance at the screen, trembling. His friend notices how frightened he
is, and says to him, quietly, “Remember, it’s only a movie!” In directing his mind to this
thought, the child feels the some of the paralyzing fear and uncomfortable uneasiness
dissipate.

47 In “Emotional Strategies and Rationality” (2000) Patricia Greenspan describes the self-
management of emotions as part of an instrumental strategy conducive to practical
rationality.
Directing one’s mind’s to the thought, “It’s only a story” is perhaps the earliest strategy we learn as children to modulate our emotional response to fiction. As we become more sophisticated appreciators of books, films, television, and theater, we learn to marshal and employ our knowledge of narrative structure and genre as well. Imagine you are roped into watching a television drama that you don’t think is all that great. It is filled with suspense, and although your fear response is being deftly manipulated, you don’t judge the program to be of much quality: the acting is bad, the dialogue is stilted, the characters are cookie-cutter, and the plot is hackneyed. Despite all this, the show is not so inept so as to prevent it from generating a good deal of suspense.

The protagonist of the series, an attractive blond female spy, finds herself trapped behind enemy lines. The sadistic Syrian police interrogator has a gun to her head with his finger on trigger. Sweat is dripping down her face as she tries to wriggle free of the ropes binding her to a rusty metal chair… Your fingers are clammy, you are biting your lip, and your heart is beating quickly. Although you sometimes enjoy suspense in movies, this is getting too much, and, given that you don’t have much faith that this program has any redeeming value, you see no point in this. Still, you cannot get yourself to turn off the set, and you cannot stop feeling fear and suspense. What does one do in a predicament such as this one?

One way to calm your frayed nerves without averting your eyes from the screen is to direct your mind to a thought like this: “There is still forty minutes until this show is over – they’ll never kill off the main character now”. Or you could think to yourself, “they
can’t allow this character to be killed or maimed – the actress who plays that role is the only reason people watch this wretched show”. These sorts of thoughts, which are directed toward considerations that are external to the fictional “world” itself, can serve to subdue and mitigate our fiction-directed emotions.

It should be clear by now that beliefs and desires that we have about the real world play a real role in determining our emotional response to fictional events. In some ways, this seems obvious. Someone who works with reptiles as part of her job is unlikely to be gripped by fear when watching Indiana Jones descend into a pit full of what she can clearly tell are non-venomous snakes. Our real-world knowledge, reality-directed emotions, and history of emotional response, often regulate our emotional response to fiction. Someone who has recently escaped a war may fall to pieces when watching *Apocalypse Now*. Someone who has felt the sting of racial prejudice may feel a heightened sense of indignation as he watches members of the Ku Klux Klan burn down a black family’s house in Spike Lee’s *Malcolm X*. In these sorts of cases, the extent to which the viewer feels reality-directed fear, grief, or indignation rather than fiction-directed grief, fear, or indignation is very difficult to determine. Should we say that the viewer’s previous experience primes him for a more or less forceful episode of fictional quasi-emotion? Or should we say that the viewers’ real emotions are being elicited and intensified by his experiences of fictional emotion? Cases like this can make the real-emotion/quasi-emotion distinction appear rather forced and artificial.
In cases of what I will term “inappropriate affect”, a fictional work elicits a strong emotional response in an appreciator, but we would hesitate to attribute to her a “fictional emotion”. This phenomenon often occurs when an appreciator experiences an emotion that the author of the work clearly does not intend to elicit. For instance, imagine someone reading a Harlequin novel with a sex scene that is supposed to evoke thoughts, feelings, and associations of romance, arousal, and wonder, but instead the reader experiences episodes of hilarity and amusement. Or consider someone who watches what is supposed to be a “heart-warming” romantic comedy and comes out feeling depressed and despondent about the cheapness and superficiality of people’s emotional lives and attachments. In these cases, the appreciator enters into the “world” of the fiction, but she makes judgments that the author does not intend for her to make, and these judgments lead her to have emotions that the author does intend for her to have. In these cases of “inappropriate affect”, the appreciator’s emotions are explained in part by his knowledge of the facts of the fictional world, and in part by the appreciator’s own real normative commitments (which differ from those of the author, or at least from those which the author expresses in the work).

In some cases, however, features of the appreciator’s emotional response, although elicited by the fictional world, are directed towards matters in the real world. Consider the viewer of a science-fiction series who feels indignation, rather than grief or fear, when a black character dies early in the program, because she notices that a preponderance of the characters who die early on the show are black, and she takes this

---

48 (When I label the response “inappropriate” I do not intend to make a comment on warrant of the emotion.)
as an indication of systemic racial prejudice. In such a case, it would be wrong to attribute to the viewer an emotion that is merely fiction-directed.

There are also more complicated cases where an appreciator experiences the kind of emotion that the author intends to elicit, but feels it for the “wrong” reason. Consider a committed environmentalist who reads Michael Crichton’s 2004 novel, *State of Fear*. A major premise of the novel is that modern governments, media, and fundraising organizations use fear to control the opinions of their citizenry, and that the hysteria over global warming is attributable to scientifically unverified claims. It is clear that Crichton is trying to elicit feelings of indignation on the part of the reader. The reader is meant to feel indignant that the government and the scientific establishment in this story would try to scare the unsuspecting electorate into thinking that there is a global warming crisis. Upon reading the novel, the committed environmentalist may very well feel indignant, but she will most likely feel indignant that Crichton would air such politically and ethically corrosive views, even if they are framed within the context of a work fiction. Even though the emotional response is elicited by a fictional work in virtue of that work’s content, what the environmentalist feels is best construed as reality-directed indignation, and the object of that indignation is Michael Crichton and his book.

For a large class of cases it becomes very difficult to determine whether an emotion response is reality-directed or fiction-directed. Take, for instance, our emotional response to historical and philosophical novels and films. When feel sadness, elation,
admiration or disappointment for a character that we know represents a real historical figure, (say, Napoleonic in *War and Peace*), do we experience mere quasi-sadness, quasi-elation, quasi-admiration, or quasi-disappointment? If one feels despair when Ivan of *The Brothers Karamazov* says, “If there is no immortality, there is no virtue" and as a result, one cannot get out of bed the next morning, surely one has felt more than mere quasi-despair?

The explanation for why we feel which emotions, and at what intensity, includes a whole host of factors, some of them reason-responsive, some of them merely content efficacious, and some of them neither. You might cry at a performance of Puccini’s La Boheme because:

1. Because you pity poor Mimi who was abandoned by Rodolpho in her sickness and frailty.
2. Because the actor playing Rodolpho reminds you of your ex-husband.
3. Because the sound of the flutes and the cellos always creates a lump in your throat.
4. Because you drank a little too much red wine at dinner.

To categorize the emotion as either a “real” emotion or a fictional quasi-emotion is a fool’s errand. It effects an unreal discontinuity between our attitude towards fictional characters and events, and our attitudes toward things in the real world.

---

49 For a detailed treatment of how finely woven these three levels can be, see Arpaly 2006 Chapter 2.
Often an emotional response that starts off as straightforwardly fiction-directed seeps over the fiction-reality boundary. In watching Hitchcock’s classic *The Birds*, it is natural to suffer from a palpable sense of eeriness, anxiety, and discomfort. And so it is not that surprising that days or even weeks later, you might experience a sense of foreboding when you notice a group of crows congregating somewhat suspiciously in the corner of a parking lot. This tendency for fictions to “stay with us” and insinuate themselves in our emotional lives is not uncommon.

In the light of all of this, it should not be so baffling that Hursthouse’s characters, or the grieving atheist who speaks to his brother’s grave, or Sheila who rails at her computer, experience emotions that can only be made sense of when we attend both to their beliefs about the real world, and to the content of their flights of fancy. What relief or catharsis could Jane who scratches out the eyes in the photograph procure if she were not imagining that the abrasions she makes are injuring Joan, the person represented in the photo? Similarly, for Sheila, her imagining that her computer is sentient is what makes it possible that she experiences the vengeful emotions related to “punishing” her computer. The grieving atheist who speaks to his brother’s grave experiences consolation only if he imagines that he speaks with his brother. In each of these instances, a cognitive component of the emotion is found in the imagination of the subject. But this is not to say that the emotions in question are not also based on belief. Jane’s anger is partly constituted by the belief that she has been betrayed by Joan; Sheila’s feelings of frustration issue from belief that she has been stymied;
the grieving brother’s feelings of anguish issue from the belief that someone whom he loves in no longer around. In this way, each of these agents experience emotions that simultaneously about the real world, and about their “interior fictions”. Each of these emotions is made possible by a “framing belief”, an awareness that this particular imaginative flight of fancy is just that – a flight of fancy. Just as their actions remained baffling and opaque without appeal to both belief and imagining, so also their emotional strategies remain inscrutable without attending to both their avowed convictions and to their fantastical imaginings.

Ethically speaking, we are interested in both the facets of people’s emotions that are reality directed, and the facets of emotions that are imagination directed. We might be disquieted, for example, by the teacher who owns a collection of voodoo dolls representing his students, or by the movie watcher who roots for Hannibal Lector. We may also be uneasy about the spectator who feels no ambivalence about fully imaginatively engaging with the musical “Showboat”. We often says things like, “I can’t believe that you sympathized with Harvey Pekar– he was such a self-absorbed pig!” or “How could you not feel any pity for Willy Loman?” Our identification with, or repudiation of, characters who are “merely” fictional can sometimes reveal aspects of our real-world commitments and values. Such ethical analysis is very tricky business, and will be left to Chapter 5.

50 We see a similar pattern with what Patricia Greenspan (2000) has termed “strategic” emotions. In order to whip themselves into a frenzy for the 1972 Summit Series, the Canadian national hockey team might have imagined that the players on the Soviet team were heartless henchman of the Evil Empire. But their polite and cordial exchanges after the series was over seems to attest to the suspicion that this was simply a useful fantasy all along, and that the players were aware of this fact.
CHAPTER FIVE: IMAGINATIVE RESISTANCE AND CAPITULATION

Our sense of the horror and dreadful atrocity of such conduct, the delight which we take in hearing that it was properly punished, the indignation which we feel when it escapes this due retaliation, our whole sense and feeling, in short, of its ill desert, of the propriety and fitness of inflicting evil upon the person who is guilty of it, and of making him grieve in his turn, arises from the sympathetic indignation which naturally boils up in the breast of the spectator, whenever he thoroughly brings home to himself the case of the sufferer.

- Adam Smith, *Theory of the Moral Sentiments*.

It may just be one of the main offices of art less to represent the world than to represent it in such a way as to cause us to view it with a certain attitude and with a special vision.

(167)

- Arthur Danto, *The Transfiguration of the Commonplace*. 
I. Keeping Cognitive Order

In a recent paper entitled “How I Really Feel About JFK” (2003), Stacie Friend marshals a sustained effort to explain, defend, and revive Kendall Walton’s theory of quasi-emotions. She maintains that Walton’s opponents (the “thought theorists”) rely on certain idiosyncratic features of examples that are based on purely fictional characters (such as Anna Karenina or Raskolnikov). She contends that we can only appreciate the full force of the paradox of fiction (roughly speaking, the paradox that I set out at the beginning of Chapter Four) when we turn our attention to fictions that treat real persons and events (such as Shakespeare’s Henry V, or Friend’s preferred example, Oliver Stone’s 1991 film JFK). Friend contends that we cannot explain central features of our engagement with reality-based fiction without the resources of something akin to Walton’s emotion/quasi-emotion distinction (36).

One way to appreciate what Friend takes to be the explanatory advantage of Walton’s irrealism is to compare emotional responses to fictions with phobias that we ourselves know to be irrational. In both cases, the emotions in question are not based on beliefs. But only in the case of phobia, does this lack of belief entail irrationality. If I am frightened by Fido even though I know very well that he is harmless, I am irrational. If I feel frightened while watching a film about menacing gremlins, I am not irrational. But just as I do not believe that I am endangered by Fido, so also I do not believe that I am endangered by the gremlins. So, what accounts for the difference in our assessment of rationality? Friend thinks that the Waltonian has a ready and perspicuous answer to this question: “[O]nly full-fledged emotions require beliefs to be rational, while quasi-
emotions do not.” (40). In her detailed analysis of our emotional response to *JFK*, Friend deploys a similar strategy, a strategy that she claims is not available to the thought theorist. According to Friend, the thought theorist is unable to account for the apparent conflict between a spectator’s emotions that are based on her beliefs about the assassination of JFK, and the spectator’s emotions that are based on the “fictionalized” presentation of the event in *JFK*.

The main subject of Stone’s *JFK* is not President John F. Kennedy, but Jim Garrison, the real-life New Orleans District Attorney who prosecuted the trial of Clay Shaw, a man who was accused of the assassination of JFK and later exonerated. Stone’s film is largely based on Garrison’s 1988 book, *On the Trail of the Assassins* and it portrays Garrison in a markedly positive light. As Friend puts it, “It is more than clear that in response to the film, we are supposed to imagine that Jim Garrison is a person of inestimable integrity, whose dedication to the truth is genuinely heroic, and that Clay Shaw is a shady, arrogant individual whose acquittal depends on the continued operation of the conspiracy” (41). In engaging with *JFK*, we are to imagine of the real Garrison that he is incorruptible and heroic, and we are to imagine of the real Shaw that he is corrupt and dishonest. But according to Friend, the facts of the case are rather in opposition to Stone’s propagandist portrayal. In fact, “Garrison and his office bribed witnesses, took statements made under hypnotic suggestion as testimony, and otherwise presented no evidence for the guilt of

---

51 I am not sure that Friend’s diagnosis of the irrationality of phobic responses is entirely perspicuous. Part of what makes phobic responses irrational seems to be the presence of a (justified) belief that the object of fear is not in fact a real danger.

52 In the discussion that follows, I simply defer to Friend’s account of Stone’s film and the actual facts of the case. (I have no expertise in this matter).
Clay Shaw, who was acquitted in under one hour. After years of prosecution, Shaw was left bankrupt and died an early death” (42).

The understanding that Garrison’s conspiracy obsession and “cavalier attitude toward the truth” destroyed the life of Clay Shaw forms part of what Friend refers to as her belief stock. Even while watching JFK, Friend has these beliefs (even if only dispositionally). However, the film prescribes that she imagine something quite different, namely that Garrison is heroic and that Shaw is contemptible. Luckily, Friend observes, this does not prove problematic: “I imagine that Garrison is admirable, while I believe that he is contemptible.” (42) Friend contrasts her response with “Naïve Nellie” who, persuaded by Stone’s propaganda, comes to believe that Garrison was fearless and Shaw was the agent of a real conspiracy. Whereas Friend compartmentalizes her imaginings, Nellie incorporates them (43). While Nellie forms the belief, “Garrison is noble”, Friend forms the belief, “In JFK, Garrison is noble.” This belief is not in contradiction with the belief contained in Friend’s belief stock “Garrison is contemptible”.

So, Friend experiences certain emotions each of which is based on a constitutive belief. She feels disgust for Garrison, and pity for Shaw. These emotions form part of Friend refers to as her emotional stock whose composition varies in predictable ways with changes in her beliefs (insofar as she is rational). However, in response to the film, Friend feels sympathy for Garrison and his noble pursuit of truth. She wants Garrison to win the case. To maintain “cognitive order” Friend must separate these emotions about

---

53 Friend borrows this terminology from Richard Gerrig’s Experiencing Narrative Worlds (1993)
Garrison from her real attitudes toward Garrison, in the same way that she keeps her beliefs about Garrison separate from her imaginings about him. If not, she will simultaneously admire and disdain Garrison on the same count, and this is an irrational state of mind. Nellie, on the other hand, has no need for this separation. Based on the beliefs that she forms in response to the film, she genuinely admires Garrison. It would seem that Friend is no less rational than Nellie. Friend avers Walton’s account best explains the facts: “Walton has a simple explanation of the fact that I remain as rational in my emotions as Nellie, since there need be no conflict between my genuine contempt and my quasi-admiration: it is not literally true that I admire Garrison, only fictionally true; really I detest him” (44). Even while Friend is imagining that Garrison is heroic, she still detests him. Her (belief based) contempt has priority over her (imagination based) admiration (45). This priority of belief over imagination allows her to maintain “cognitive order”

According to Friend, the thought theorist, who will characterize both her contempt and her admiration as real emotions, lacks the conceptual resources to explain how cognitive order is maintained.

So I feel admiration of Garrison because I imagine that he is admirable, and I feel contempt for him because I believe that he is contemptible.

But this way of putting it implies that imagining, of Garrison, that he was admirable is sufficient to explain genuine admiration of him. That just seems false. Compare the following case. Suppose you believe Nelson Mandela to be a great man and you admire him. Now, if I made up a story in which Mandela tortured kittens, and you knew this story to be
pure fantasy, it would be inexplicable for you to change your feelings toward him. Similarly, my imagining that Garrison was a noble pursuer of truth would not explain a change in my feelings toward him. That is, my emotional stock should remain exactly the same, even while I am imagining that Garrison is different from how I believe him to be. (44)

It is difficult to disagree that to alter one’s real attitudes and feelings toward Garrison or toward Nelson Mandela merely because one imagines of Garrison or of Mandela that he behaves in ways contrary to what one knows about him is downright irrational (and, I think, morally problematic to boot!). However, what this constraint on rationality tells us about the nature of the emotions in fictions is not altogether clear.

The first thing to notice is that “quasi-emotions”, just like “real emotions”, are subject to warrant. I am not just as rational in my terror of Big Bird as I am in my terror of the poltergeist. That quasi-emotions have these kinds of “correctness conditions” is evident in our genuine perplexity when we witness someone weep while reading P.G. Wodehouse, or burst into laughter while watching a film directed by Kristof Kieslovski. So, it is not true that “anything goes” when it comes to emotions directed toward fictions. Indeed, fictional representations can themselves elicit irrational phobic responses (for someone deathly afraid of Fido, the barking dogs in Balzac’s “L’Auberge Rouge” will likely be very disquieting; for the snake phobic, Indiana Jones might be

---

54 The assessment of the rationality of emotions directed toward fictions is somewhat more complicated than that of “ordinary” emotions. We may be perplexed by a spectator’s emotional response even when the spectator responds in the way that is prescribed. For example, I may be shocked that my friend weeps at Hugh Grant romantic comedy, even though weeping is what is prescribed.
unwatchable). The case of phobias elicited by fictions problematizes Friend’s claim that “only full-fledged emotions require beliefs to be rational, while quasi-emotions do not.” (40). Beliefs play a role in determining the warrant of our fiction-directed emotions. If you are terrified by Fido-in-the-film, yet you do not believe that Golden Retrievers are dangerous, your quasi-emotion is not warranted and thereby irrational.

The second thing to notice is that most of the time when one alters one’s attitude towards someone as a result of their resemblance to another person (whether real or fictional), one manifesting a kind of irrationality. My feelings of fear directed toward my postman are irrational if they can be explained in terms of the postman’s physical resemblance to a high school bully I knew (especially given my knowledge that physical attributes do not reveal characterological ones). My feelings of resentment toward a teacher are irrational if they can be explained in terms of my teacher’s idiosyncratic diction being similar to that of my absent father (especially given my knowledge that these speech patterns are not indicative of personality traits). Similarly, I am irrational if I start to admire Garrison after watching Stone’s propagandist film, if I know that Stone’s character misrepresents the real Garrison. There seems to be no need to distinguish real emotions and quasi emotions to explain the way in which cognitive order is upended. Walton and

---

55 Empirical research seems to bear out view that people experience the same phobias in their imaginative life as they do in their ordinary life. In one experiment, subjects were presented with an imaginary scenario involving a snake encounter. Although almost all of the subjects showed physiological signs associated with fear, the fear responses were stronger in people who were snake phobic than in people who were phobic about public speaking and not about snakes (Lang et al. 1983).

56 Our ability to keep our emotional stock in line with our belief stock is distressingly weak. Just being introduced as “the man who did not rob the bank” can compromise one’s good name. For more examples of these “priming” phenomena see Bargh, Chen, and Burrows (1996).
Friend’s ontological distinction at the level of emotion does no more explanatory work than a simple cognitive distinction.

However, there are certain occasions when it does not seem irrational to alter one’s emotions toward a person as the result of an affect-laden encounter with another person (fictional or real) who resembles him. Sometimes we are blind to certain virtues or vices in one person’s character until we encounter similar virtues or vices in another person. I may fail to notice Sally’s lack of social conscience because I am bamboozled by her charm and good looks. But when go dinner with Sally and her friend, Diane, who is just as heartless but not as pretty, I may realize that I had been blind to a serious flaw of character in Sally. It is important to note that I need not have learned anything new about Sally after meeting Diane. Rather, I end up noticing something about Sally that had previously failed to register.

Fictional representations of real people can work in similar ways. Upon reading Robert Grave’s novel, *I Claudius*, a sympathetic rendering written from the perspective of the emperor himself, one might feel a diminished sense of opprobrium for Claudius’s corruption and manipulation. Even if one was previously aware of the facts of Claudius’s stammer, limp, and nervous tics, the act of vividly imagining what it might have been like to rule Rome with qualities that make you appear mentally deficient may very well awaken a sympathy in the reader that was previously absent. Similarly, although one is unlikely to learn anything new about the biography of revolutionary France’s queen in

57 For an analysis of how these sorts of “dawnings” can be rational processes, see Arpaly (56).
Sophia Coppola’s film *Marie Antoinette*, one might well feel less disdain for her lack of social conscience upon being made vividly aware of the extent to which she was isolated from anyone outside of the court. So it seems that we are not *always* irrational in altering our emotional stock as a result of our engagement with fictions.

In some cases, it seems indisputable that changes to our emotional stock that occur as a result of fiction-directed emotions are occasions for learning and emotional refinement. After reading *Anna Karenina*, you might realize that the relationship between fidelity and betrayal is much more complicated than you earlier imagined. This realization might well have a bearing on your emotional response to real life cases of infidelity in your own life or that of your friends.

Of course, none of this is to deny that in Friend’s Garrison case, one would be irrational to alter one’s emotional stock as a result of an affect-laden encounter with Oliver Stone’s fictionalized Garrison. The phenomenon of having to “work” to keep our emotions in line with our beliefs is not peculiar to our experience of reality-based fictions, or even of fictional depictions more generally. Suppose that you learn from some gossipy neighbors that your upstairs housemate, Jimmy, is an ex-con. They claim to know this because the postman, who happens to be a friend of theirs, once told them that Jimmy regularly receives mail from the Department of Corrections, and that ex-cons who are still on probation get those sorts of envelopes in the mail. Although you strive to be calm and open-minded, you still feel disquieted in your dealings with Jimmy. You find yourself altering certain habits: when you are inside the house, you make sure the door is dead
bolted; instead of leaving your rent check in the front hallway, you mail it directly to the landlord; you decide to protect wireless internet connection with a password. Months later, you strike up a conversation with Jimmy and, much to your embarrassment, you learn that he is an architect who specializes in the design of prisons and college dorms. Sheepishly, you admit to yourself that you were too quick to draw conclusions, and that you should not have relied on the testimony of your nattering neighbors. You stop using the deadbolt, and resume the practice of leaving your rent check in the front hall. But despite your embarrassing volteface, the emotional residue of your past suspicion remains. When Jimmy approaches you unawares on the front stairway one morning, you freeze in a way that is not fully explicable in terms of your ordinary startle reflex. When you learn that Jimmy will be vacationing in the Berkshires next week, you feel relaxed in a way that is not altogether understandable in terms of your pleasure at having the house to yourself.

Previous to your discovery of Jimmy’s occupation, you may have had some justification for your wariness of him. But now you have none. To put the problem in Friend’s jargon, the contents of your emotional stock have not kept up with certain changes in the contents of your belief stock. You still experience aspects fear and relief from fear, even though you now lack any justification. You still feel disquieted, even though you yourself realize that there is no reason for disquiet. You have a belief-emotion mismatch, and insofar as this is true, you manifest a certain kind of irrationality. For you to be rational in your emotional response to Jimmy, your emotions must “catch up” with your beliefs. And with enough exposure to and interaction with Jimmy, the chances are pretty
good that, eventually, they will. In the meanwhile, your prevailing inappropriate emotions toward Jimmy might occasion a certain measure of guilt. The experience of guilt is not uncommon when one’s gut feelings fail to accord with one’s reflective beliefs (think of the political liberal who cannot help but feel a bit green about the gills at the sight of two men kissing).

This phenomenon has its analogue in the imagination. We are often reluctant to allow ourselves to engage in fantasies that we morally disapprove of. Similarly, we are loath to imagine people that love and respect behaving in ways that we find repulsive. (If you are asked to imagine that your beloved is a molester, it doesn’t help to be told that it’s only a fantasy). Fictions that are based on reality produce a similar resistance, which often results in a failure or refusal of the appreciator to emotionally engage with the work. Appreciators are likely to emotionally abandon fictions which mandate that they imagine that, for instance, the Holocaust was a hoax, or that blacks are intellectually inferior. In the light of this, certain aspects of Friend’s JFK example seem quite unlikely. Someone who truly despises Garrison for having ruined a decent man’s life, and who has taken enough interest in the case to have read up on it, is rather unlikely to feel admiration toward Garrison in the film. More likely, she is repulsed by Stone’s propagandist portrayal, and she fails to engage with the film sufficiently to experience the emotions that are prescribed by it. Friend emphasizes that reality-based fictions mandate that we imagine things of real people. But it would seem that the only way a “Garrison-hater”
would fully engage with Stone’s film is to imagine that Stone’s Garrison is “purely fictional”\textsuperscript{58}. And even this would be a feat too great for many of us.

\textsuperscript{58} Part of what seems to do the work in Friend’s example is that it uses admiration (which is more a settled attitude) rather than fear. The use of admiration helps make plausible Friend’s claim that we do not experience genuine admiration toward Garrison. However, her choice of admiration makes her example less persuasive insofar as it seems rather unlikely that the real Garrison-hater feels \textit{any} kind of admiration toward Garrison.
II. Hume’s Puzzle

In his 1757 essay entitled “On the Standard of Taste,” David Hume observed a curious asymmetry in how readers engage with narrative works: while they are quite willing to imagine all manner of absurdity when reading the works of Homer, Ovid, Aristo, or Cervantes, they are unwilling or unable to engage imaginatively with morally counterfactual worlds, that is, fictional worlds where deviant moral principles obtain. Just under two-hundred-and-fifty years later, Hume’s puzzle has gripped a group of philosophers working at the intersection of ethics, philosophy of mind, and aesthetics. In a recent paper Tamar Gendler dubbed the problem “The Puzzle of Imaginative Resistance” (2000).

The main object of Hume’s essay is to characterize the nature of the “true aesthetic judge”. According to Hume, the central qualification of such a judge is that he “preserve his mind free from all prejudice” (239). Hume is not here endorsing the Kantian view that a true aesthetic judge must take the view from nowhere, exercising a kind of sensus communis. Rather, Hume takes a more contextualist position. Early on in the essay he adopts the stance that the ideal critic must take the point of view of the work’s intended audience. In order for work of art to be properly appreciated, it “must be surveyed from a certain point of view, and cannot be fully relished by persons, whose situation, real or imaginary, is not conformable to that which is required by the performance.” (239). To properly appreciate a work “a critic of a different age or nation, who should peruse this

59 All references to Hume’s essays are to David Hume, Essays Moral, Political, and Literary. Ed. Eugene F. Miller (Indianapolis: Liberty Fund, 1985)
discourse, must have all these circumstances in his eye, and must place himself in the same position as the audience [envisioned by the creator of the work.” (239). To the extent that a judge departs from the required point of view, his taste, “evidently departs from the true standard; and of consequence loses all credit and authority.” (240) Hume scholar Michelle Mason describes Hume’s judge as “less an impartial observer than a cultural chameleon” (60).

However, for all the open-mindedness, cosmopolitanism, and cultural sensitivity of the true aesthetic judge, Hume maintains that there are still important limits on what can and should be appreciated. In the final pages of the essay, Hume claims that a true judge should not “relish” works or parts of works that conflict with a moral standard of correctness of which the judge is confident.

Where speculative errors may be found in the polite writings of any age or country, they detract but little from the value of those compositions. There needs to be but a certain turn of thought or imagination to make us enter into all the opinions which then prevailed and relish the sentiments or conclusions derived from them. But a very violent effort is required to change our judgment of manners, and excite sentiments of approbation or blame, love or hatred, different from those to which the mind from long custom has been familiarized […] I cannot, nor is it proper that I should, enter into such [vicious] sentiments. (253)
Here Hume makes both a descriptive claim and a normative one. He holds that it is in fact “very difficult” to enter into moral sentiments that are markedly different than our own. Further, he holds that we ought not enter into such sentiments. Contemporary philosophers working on the “problem of imaginative resistance” have focused on the former claim. This will also be my focus. However, in assessing the purported difficulty in entering into vicious sentiments, I will need to make appeal to the “moral hazard” of doing such a thing.

Walton’s “Supervenience” Solution

Hume is not primarily concerned with the depiction of immoral thought and action itself. Rather, he is worried about fictions in which "vicious manners are described, without being marked with the proper characters of blame and disapprobation" (246). In “Morals in Fiction and Fictional Morality” (1994) Kendall Walton considers a pair of sentences that exhibit exactly this quality (37):

*In killing baby, Giselda did the right thing; after all, it was girl.*

And

*The village elders did their duty before God by forcing the widow onto her husband’s funeral pyre.*

In each of these fiction fragments, there is a description of an action followed by an evaluation of that action. After presenting the puzzling sentences, Walton asks the rhetorical questions: “Why shouldn’t storytellers be allowed to experiment explicitly with

---

60 Hume makes the further claim that such works are aesthetically flawed. I will not address this rather thorny question. For an excellent survey on recent approaches, see Noël Carroll’s “Art and Ethical Criticism: An Overview of Recent Directions of Research” (2000).
worlds of morally different kinds, including ones even they regard as morally obnoxious? There is science fiction; why not morality fiction?” Walton is not without a theory (although tentative) to explain this paradoxical piece of phenomena. He begins with a claim about the supervenience of moral properties on natural properties:

Moral properties depend on ‘natural’ ones and, I believe, in the relevant manner (whatever that is); being evil rests on, for instance, the actions constituting the practices of slavery and genocide. This, I suggest, is what accounts (somehow) for the resistance to allowing it to be fictional that slavery and genocide are not evil.

But even if I believe that A supervenes on B, it does not follow that I cannot fictionally assent to the existence of A in the absence of B. As philosophers have known for a long time, it is perfectly possible to entertain the Cartesian presuppositions even if I believe that mental states supervene on physical states. So, Walton needs a further argument to address the explanandum. He provides one:

We still need an explanation of why we should resist allowing fictional worlds to differ from the real world with respect to the relevant kind of supervenience relations. My best suspicion, at this moment, is that it has something to do with an inability to imagine these relations being

---

61 In my reading of Walton I am indebted to Gendler (2000) and Matravers (2003)
different from how we think they are, perhaps an inability to understand
fully what it would be like for them to be different. (46)

To summarize, Walton’s explanation is that it is simply unimaginable (perhaps because it
is conceptually impossible) that the supervenience relations between moral properties and
natural properties are other than they actually are. We could not fully understand what it
would be like for them to be different. And so it is not surprising that in just these cases
we are incapable of going along with an author’s mandate to imagine. Tamar Gendler
(2000) has labeled this the “impossibility hypothesis”.

In “The Puzzle of Imaginative Resistance” (2000) Gendler critiques the impossibility
thesis by arguing that it is possible to imagine conceptual impossibilities. As evidence of
this claim, she presents a story – “The Tower of Goldbach” – that contains the conceptual
impossibility “twelve both is and is not the sum of five and seven” (66). Gendler’s
strategy is to distract the reader from closely attending to the conceptual impossibility,
which is deftly hidden in the story. She gets the reader to focus on certain elements of
the story, and thereby ignore others.

There is some controversy regarding whether Gendler really gets the reader to imagine a
conceptual impossibility. The important thing to notice, though, is that literary authors
who create immoral yet sympathetic characters pursue a strategy that is structurally
similar to Gendler’s. In Milton’s Paradise Lost, Satan’s pride and ambition lead to his
tragic end, but he is also presented as dreadfully alluring, and many readers succumb to

---

62 See Kathleen Stock 2003.
his allure. Indeed, it has traditionally been considered a mark of Milton’s genius as an artist that he is able to elicit profound sympathy for evil incarnate. It would seem that although strong, the mechanisms that underpin imaginative resistance are not insuperable. If Walton was right and it really is impossible to imagine morally counterfactual worlds, we would likely be much less concerned with the capacity of art to corrupt.
Matravers “Report Model” Solution

Derek Matravers (2003) proposes a novel solution to the puzzle of imaginative resistance that does not presuppose the “impossibility hypothesis”. He argues that the explanation lies in our apprehension of the limits of the authority of the fictional narrator (96). He bases his explanation on a theory of how we engage with fictions called the “report model”\(^{63}\). On the report model, we interpret fictional narratives to be “reports” of events told to us by someone with knowledge about them. The narrator is like a trustworthy foreign correspondent who relays information to us. Instead of sending us his missives from a far-away country, the fictional narrator sends his reports from a fictional world. In engaging with a fictional narrative, the reader is meant to make-believe that he is learning about the characters and events in the story through reliable channels. As Gregory Currie puts it, “To make-believe a fictional story is not merely to make-believe that the story is true, but that it is told as known fact” (73).

According to Matravers, the limits of our willingness to imaginatively engage with a fictional narrative track the limits of the epistemic authority of fictional narrator. Consider first the limits to the authority of a foreign correspondent. If you read in a travel magazine that, “The food in Tamil Nadu is spicy” you are likely to take this as fact, barring any reason you have to distrust the writer, or any specialized knowledge that you have to the contrary. However, if you read, “The food in Tamil Nadu is spicy, and therefore it is not palatable,” you might raise an eyebrow. The correspondent might have privileged epistemic access when it comes to knowing whether or not the food in Tamil

Nadu is spicy. But he is no better position than you to judge whether spicy food is palatable. Similarly, when a novel’s narrator tells you that “Giselda killed her baby,” you assent to the narrator’s authority, since presumably the narrator knows better than you do whether in the fictional world Giselda killed her baby or not. However, if the narrator goes on to tell you Giselda did the right thing in killing her baby since the baby was a girl, you are likely to think that the narrator has overstepped his authority. The narrator does not have privileged access to moral facts, so when he gets them wrong in a way that is obvious, you disengage.

Matraver’s account has a couple of important advantages over Walton’s. First, Matraver’s theory does not assume the implausible impossibility hypothesis. To a certain extent, appreciators can choose whether or not to, say, allow themselves to suspend ordinary moral evaluation. In addition, it is possible for the narrator to conceal that he is fudging the moral facts in order to obviate resistance, much in the way Gendler conceals her fudging of mathematical truths in the “Tower of Goldbach”.

Second, Matraver’s theory is able to explain how imaginative resistance can be elicited by non-moral evaluations. If a narrator suggests that an ice cream sundae from the Coldstone Creamery is an epicurean’s delight, or that the novels of Michael Crichton are literary masterpieces, sophisticated readers are unlikely to go along. Indeed, any time we
think that a narrator has overstepped his authority as an expert, we are liable to resist\(^\text{64}\).

This extends to \textit{all} evaluative claims.

However, as appealing as Matraver’s theory is, I think that it is vulnerable to some damaging counter-examples. Consider: I do not think that preserving one’s “honor” is a defensible excuse for killing someone, and I do not think that a woman’s virginity is paramount to her virtue; moreover, I don’t think that Tolstoy or Turgenev has any special expertise on this matter. However, these sorts of evaluations by Tolstoy’s and Turgenev’s narrators do not cause me to imaginatively disengage and resist. Why? I do not pretend to have a well-worked out theory on the matter. But I think that part of the explanation lies in my background belief that Tolstoy and Turgenev are more or less blameless (or at least less blameworthy than a contemporary writer) for holding these views. Given mores of the authors’ society, I may even be impressed that they do not express \textit{even more} repugnant views in their work.\(^\text{65}\)

The historical or cultural remoteness of work can also have the effect that the appreciator does not feel at all \textit{implicated} in the pernicious moral views of the author. There may be less of the kind of moral queasiness that derives from the question, “What does this say

\(^{64}\) Stephen Yablo (2002) gives the following example of a text that is likely to elicit aesthetic imaginative resistance: “All eyes were on the twin Chevy 4 X 4’s as they pushed purposefully though the mud. Expectations were high; last year’s blood bath death match of doom had been exhilarating and profound, and this year’s promised to be even better. The crowd went quiet as special musical guests ZZ Top began to lay down their sonorous rhythms. The scene was marred only by the awkwardly setting sun.” (485)

\(^{65}\) Another problem with the report model is its difficulty in dealing with unreliable narrators (\textit{Midnight’s Children}), ignorant narrators (\textit{The Turn of the Screw}), naive narrators (\textit{The Adventures of Huckleberry Finn}) or otherwise confused narrators. Matravers addresses, but does not fully diffuse, this sort of objection in Chapter 4 of \textit{Art and Emotion} (1998).
about me?” The emphasis on honor and virginity is so foreign and strange that the appreciator may not sense any moral vulnerability or impropriety. At the same time, the appreciator does not defer to the “expertise” of the narrator. Rather, the narrator’s evaluations are so bizarre as to appear quaint.
Nichol’s “Single Code” Solution

Shaun Nichol’s “Single Code” solution to the puzzle of imaginative resistance suffers from similar problems as the report model. According to the Nichols and Stitch single code theory, the mind’s inferential, emotional, and moral response mechanisms work in the same way on pretense representations as they do on belief representations (Nichols and Stitch 2003). If a mechanism is activated by the occurrent belief that \( p \) it will also be activated by the isomorphic occurrent imagining that \( p \).

[T]he moral response mechanisms will treat pretense representations as they treat isomorphic belief representations. Having the occurrent belief that Macbeth ordered Duncan’s murder leads us to regard the action as immoral. Hence, since imagining is in the same code, having the occurrent pretense presentation that Macbeth ordered Duncan’s murder leads us to regard the action as immoral in the imagination. That would explain why we resist alien moral judgements even in the imagination.

(9, forthcoming in *Mind and Language*)

The problem with Nichol’s account is that it is unable to take into account many of the subtle variations in the imaginative resistance phenomenon. For example, a pious and devoted Catholic might experience imaginative resistance while reading Dan Brown’s *The DaVinci Code* because she thinks that the author is trying to paint an unfair, even libelous, portrayal of the Catholic Church (in the novel the Church covers up the fact that Mary Magdalene was pregnant with Jesus’ child at the crucifixion). Often when readers feel as if an author is “lying” or “slandering” through his fiction, they will not experience the patterns of emotion that the author invites them to feel. Something similar might
happen when the committed environmentalist reads Michael Crichton’s novel *State of Fear*, whose plot is built around a group of “eco-terrorists” who attempt to further their political agenda regarding global warming by conspiring with a cabal of dishonest scientists. This kind of resistance is generated by *descriptive* rather than straightforwardly evaluative claims. But in these descriptions, the appreciator perceives disguised evaluations (that the Catholic Church is secretive and corrupt, or that environmentalists are dishonest and anti-American). The outraged reader of the *Da Vinci Code* is certainly not responding in the way that he would to the *belief* that Opus Dei is an evil cabal.

This kind of imaginative resistance does not require that the reader be possessed of a strong religious or ideological conviction. Consider the film review by Charles Taylor of Ron Howard’s “*A Beautiful Mind,*” a film that profiles the life of John Nash, the MIT mathematician and Nobel Prize winner:

In this type of inspirational moviemaking, a conditional triumph is no triumph at all, and a complex or unlikable hero is unthinkable. Goldman and Howard have expurgated everything about Nash's bisexuality and the fact that he fathered a child out of wedlock. His wife Alicia (Jennifer Connelly) stays faithfully by his side in the movie. We aren't told that, though Alicia did shelter him and though they are still together, she divorced him in 1963. We aren't told that their son, also a gifted mathematician and also schizophrenic, has suffered as his father did, living a life often divided between rootless wandering and institutionalization. And there is nothing of Nash's plain unpleasantness: his racism, his snobbery, his history of violent behavior toward others. Of course, none of these things lessens Nash's accomplishment, diminishes the originality of his ideas or makes his emergence from mental illness any less remarkable. *But, by omitting them, the*
moviemakers are signaling that think they do. And they're implicitly saying to the audience that we can't be addressed as grown-ups. (italics, mine)

I think it is quite likely that the audience-member who has the distinct sense that he is not being addressed as a grown up will fail (or rather, refuse) to engage with a work and to respond emotionally in the ways that are mandated by it. The inkling of being sold an over-simplified, caricatured, or unsubtle portrayal is liable to produce imaginative resistance that is robust and persistent. Taylor’s central complaint is not that Howard played fast and loose with some of the facts of the case. Rather, his objection is that by omitting a number of complicating factors, the creators of the film fail to reckon with complicated moral truths. He is criticizing the distinctly fraudulent way in which the world is described, and it is not difficult to imagine that the reviewer remained emotionally inured throughout much of the film.
Gendler’s “Export” Solution

Tamar Gendler’s “export” solution to the puzzle of imaginative resistance accounts for some of these kinds of resistance phenomena quite naturally. Gendler maintains that non-distorting fictions (fictions that share a background with the actual world) ask appreciators to “export” from the fictional world fictional truths that are not merely truths in the story (77). The way in which something is described in a fiction can often be the author’s means of saying something about the real world (this is most obviously the case with fables). Gender’s view is that in certain instances, the appreciator refuses to perform the exportation, and this is what accounts for the resistance phenomenon: “So my hypothesis is that cases that evoke genuine imaginative resistance will be cases where the reader feels that she is being asked to export a way of looking at the actual world which she does not wish to add to her conceptual repertoire” (77). When the appreciator is unwilling to perform the relevant exportation, she experiences what Gendler refers to as “pop out”: she no longer experiences the emotions that she is invited by the author to experience, and she is no longer immersed in the fictional world.

According to Gendler, appreciators are particularly liable to interpret moral claims as to be exported. The reason for this is two-fold. On the one hand, moral claims are categorical: if true, they are true in all possible worlds. On the other hand, unlike the truths of mathematics or of logic, there are salient instances of real moral disagreement.

---

66 Gendler’s examples of non-distorting fictions are realistic as opposed to absurdist or surrealist.

67 Gendler does not elaborate on what it means to “add” something to one’s “conceptual repertoire”. Perhaps it means to agree to perceive and evaluate something in the real world in a way that one did not previously.
So, when we encounter false moral claims in a fictional work, we interpret the author of the work as intending for them to be exported (78). In this way, moral evaluations clamor for exportation in a way that claims about, say, the existence of the fountain of youth, do not.

I am largely in agreement with the spirit of Gendler’s account. However, there are some refinements to be made. First, it is interesting to note that a subject’s unwillingness to engage with a work is often mediated by a concern with what other appreciators will add to their “conceptual repertoires” rather than a concern with what will be added to his own. Environmentalists who experience imaginative resistance while reading *State of Fear* are probably most concerned that other people might “get the wrong idea” about global warming. The pattern is similar for the devout Catholic who reads *The DaVinci Code*.

The tendency of appreciators to “pop out” of a fictional narrative because of their concern with the effect that fiction might have on others is not limited to cases of perceived slander or lies. The television series *24*, whose protagonist Jack Bauer protects America from terrorists, sometimes by resorting to torture, has been heavily criticized by liberal critics who fear that it helps to justify the policies of the Bush administration in the minds of the voting public. A philosophy professor who has no problem discussing the “ticking time bomb” thought experiment in an ethics class on utilitarianism might well balk at the invitation to imaginatively engage with *24*. In cases like this, it is at least in part the
spectator’s concern with the moral opinions of others that causes the imaginative resistance.⁶⁸

We have already seen that whether an appreciator resists and how much can depend on a number of factors, including:

(1) her moral and aesthetic opinions;

(2) the degree to which she is concerned with these evaluations;

(3) the vulnerability she feels to being tricked or manipulated into holding an opposing view;

(4) the degree to which she feels implicated in the evaluative assumptions of the author by engaging with the work;

(5) whether she holds the author blameless for harboring a false view; and

(6) the degree to which she imagines that others will be influenced by the possibly corrosive view.

But even if we hold all of these factors constant, I suspect that individuals will still vary in their tendency toward imaginative “promiscuity”. For some people, musicals like Showboat or westerns portraying dishonorable “Injuns” are impossible to engage with. But others who have a very similar moral outlook will have no difficulties. Still others will initially resist, but eventually “capitulate” because of the artfulness or persuasiveness of some aspect of the work.

⁶⁸ A article by Dahlia Lithwick Slate magazine revealed that no less than Michael Chertoff, the homeland security chief, John Yoo, the former Justice Department lawyer who produced the “torture memos”, and Supreme Court Justice Antonin Scalia have cited the Jack Bauer character as inspiration (2008).
This is not to say that there are no moral stakes in imagining. Richard Moran rightly points out that one’s emotional responses and involvement often say something about oneself; they can’t just be written off as “fictional”.

When we think of what the audience feels as falling within the scope of what is fictionally true, this suggests an innocence or lack of real-world accountability in those responses that doesn’t fit how we actually think or talk about what we feel in the theater or in other fictionalized contexts. Rather, the responses of laughter, lust, indignation, relief, delight in retribution, etc. are normally treated as expressions of genuine attitudes that we actually have, and are esteemed or repudiated accordingly. […] [T]he ordinary practice of eliciting, sharing, and criticizing such emotional responses to fictions would be a quite bizarre and pointless exercise if these responses and the attitudes they express were not located on this side of the counterfactual divide. (93-94)

I certainly agree with Moran that our emotional response to fictional narrative must be located “on this side of the counterfactual divide”. Given this, it is not surprising that imaginative engagement can be the source of anxiety. In some instances one is imaginatively engaged with a work despite one’s moral and aesthetic aversion, and despite one’s conscious and deliberate effort to circumvent engagement. For instance, you might be the sort of person who reviles romantic comedies, who considers them syrupy, maudlin, and sentimental. Indeed, it might even be important to your self-conception that you view romantic comedies as a corrupted and fraudulent genre that
perpetuates a corrosively patriarchal vision of romantic love. But one unfortunate evening you find yourself dragged along, kicking and screaming, to partake of the latest offering of Hugh Grant dreck. Exasperated, you are prepared to silently wretch the whole way through. As the movie progresses, the Hugh Grant character does something characteristically foppish and charming, and the emo music starts to play. The Julia Roberts character blushes and flashes a wan smile. Much to your disgust, you develop a distinct and unpleasant sensation of – egad! - a lump in your throat.

Is this lump evidence of an inconsistency in your values? Is it evidence of uncertainty in your judgment? I think we would be hasty to draw such conclusions. Your capitulation could be chalked up to a number of things: the alluring prettiness of Hugh Grant or Julia Roberts, the flattering camera angles, the heart-string-tugging music in the background, the extra glass of red wine you had after dinner, the fact that you were recently dumped. A variety of factors, many of them non-mimetic, can contribute to an emotional response. Sometimes this response can be both alienating and, to the extent that one values control, embarrassing.

Conversely, there are instances in which one experiences resistance, rather than capitulation, as being out of one’s control. Sometimes it is difficult to appreciate allegorical writing once we learn what the work is an allegory of. A staunchly atheist reader who as a child loved C.S. Lewis’s tales of Narnia might be put off to discover that Aslan was intended to represent Christ. Even though this reader wants to “bracket” this
annoyance and immerse herself in the absorbing fantasy, she finds that she cannot. In the same vein, it makes sense to say things like, “I wish I could get into nineteenth century novels, but I just can’t get over the sexism.” That is, you might perceive your resistance as something of a “hang up”, even though your do not at all perceive your moral views about sexism as a “hang up”. Someone could also say, “I know that Dostoyevsky has that stupid virgin/whore complex, which is very annoying, but when I read his novels I just ignore that and focus on the incredible psychological realism with which he portrays his protagonists.” So, it seems that we cannot read off someone’s moral convictions just by knowing whether they will resist or engage to particular works. Imaginative capitulation is not always morally worrisome. Just as you might regard your imaginative resistance as a hang up, you might regard aspects of your moral emotions as hang ups, and inconsistent with your settled moral opinions. In cases like this, it may be an imaginative achievement on your part to overcome imaginative resistance that derives from these emotions. Say, for instance, you hold liberal moral views on same-sex romance and sex. However, you find that you just don’t get the same “butterflies in the stomach feeling” you get from narratives of opposite sex romance. You wish you did. And you cannot identify any good reason why you do not. But you don’t. Finally a work comes along that is so beautiful, so persuasive, and so authentic that you find yourself able to emote in the way that you had always wished you could. Your “capitulation” would be triumph.

---

69 This example was suggested to me by Nomy Arpaly.
Many of these complexities arise from the fact that imaginative resistance is in some measure something that we do, and in some measure something that happens to us. Any general account that does not take this into consideration is bound to be explanatorily inadequate. In fact, I am skeptical that any unifying theory will be up for the job. The terrain is simply too uneven.
Many of these complexities arise from the fact that imaginative resistance is in some measure something that we do, and in some measure something that happens to us. Any general account that does not take this into consideration is bound to be explanatorily inadequate. In fact, I am skeptical that any unifying theory will be up for the job. The terrain is simply too uneven.
Works Cited


Radford, Colin. 1975. “How Can We Be Moved by the Fate of Anna Karenina?” *Proceedings of the Aristotelian Society*. Suppl. 49


