Aesthetic Debunking and the Transcendental Argument of the Novel

Author
ADAM GJESDAL AND
JONATHAN M. WEINBERG

Affiliation
CHAPMAN UNIVERSITY
UNIVERSITY OF ARIZONA

Abstract: Gilbert Plumer has recently argued in his Plumer 2017 that psychologically rich novels offer the reader an opportunity to draw a transcendental inference: what seems to us believable about the psychology of the characters, can be inferred to be actually true about real human psychology. We propose, first, to disambiguate a key term of art in Plumer's argument, 'believable'. Given that disambiguation, the empirically contingent nature of one of Plumer's premises comes into view. We raise two main lines of empirically-motivated debunking arguments against that premise, drawing particularly upon the psychological literatures about processing fluency, and the illusion of explanatory depth. We then conclude with some further implications for naturalistic approaches to aesthetics, and the relevance of such debunking arguments.

I.   INTRODUCTION: PLUMER'S TRANSCENDENTAL ARGUMENT

Fictional narratives are thought, in Shakespeare's words, to 'hold . . . the mirror up to nature', revealing to ourselves something about ourselves – despite the fact they largely concern people who never existed and events that never occurred. How, and indeed whether, they do this has been a ripe topic for philosophical and empirical investigation.[1] And as we learn more about how novels achieve this, we may find they do not offer such clear reflections of human nature – showing us ourselves at best 'through a glass darkly', to pivot

from Hamlet to St. Paul. Such a discovery would *debunk* the claim that novels mirror us. Debunking arguments take empirical premises about how we form beliefs and undermine the justificatory status of the beliefs in question. Our topic is the application of debunking arguments to novels' claims to reveal human nature. We start by discussing Plumer's recent proposal that novels have cognitive value – they reveal to us something about psychological reality – by expressing transcendental arguments.[2] We then outline the scheme of what we call 'aesthetic debunking arguments', providing two candidate arguments for how recent work in empirical psychology can call into question the soundness of any novel's transcendental argument. We conclude with some methodological implications for aesthetics from debunking arguments.

In his 'The Transcendental Argument of the Novel', Gilbert Plumer argues that fictional stories *as fictions* can teach us something new about the world. On his analysis, fictional narratives as a whole can express an argument, and we grasp this argument in experiencing the narrative as believable. Believability is not the same thing as realism.[3] An episode of *Star Trek: The Next Generation* may be believable even though it features technologies, like the warp-drive, that violate of the laws of physics, and creatures, like androids and aliens, unknown to modern science. What makes an episode of *Star Trek* believable is how the characters like Commander Data behave and think. (Or, if the android Data does not in fact think, we might say believability tracks how we can't help but think he must think.) Our sense of believability latches on to the psychologies of the characters, not to other features of the fictional world. As we emphasise below, this sense is part of our experience of the artwork, which we naturally slip into attributing as a property to the artwork as a whole. On Plumer's analysis, this experience has cognitive value, expressed by the transcendental argument of the novel.

A novel's argument is transcendental in a specifically Kantian sense. Transcendental arguments in the Kantian vein show that certain conditions must obtain for us to experience the world as we do.[4] For us to experience a novel as believable, Plumer argues, it must be the case that we experience the novel as 'expressing' or 'embodying' principles of human psychology, action, or society that we believe obtain in the real world.[5] As Plumer presents it, a novel's transcendental argument has the following form.[6]

(1) This novel is believable.
(2) This novel is believable only if principles p operate in the real world.
(3) Therefore, principles p operate in the real world.

The Kantian twist occurs in (2). Our experience of believability tracks certain features of the novel, ignoring others: we might experience Frodo Baggins' decision to keep the One Ring at the Cracks of Doom as believable, even though that decision results from the Ring's malevolent but seductive

promise of power to its bearers, which is decidedly fantastical. Those narrative elements eliciting this sense of believability generate the transcendental argument. Those elements, taken together, have cognitive value in that we learn something about the world by experiencing them as believable. Plumer argues that for us to experience part of the narrative as believable, and for us to be correct that these parts are believable, those parts must somehow express principles of human nature we already intuitively grasp as true. Readers might not be able to reflectively articulate these principles: that difficult task often falls to the literary critic, who uncovers and articulates principles expressed in the work as a whole.[7] 'The novel's transcendental argument is there', Plumer writes, 'whether or not anybody notices'.[8] The transcendental argument of the novel elicits these shared assumptions, making them possible objects of reflective knowledge.

Plumer doesn't do so himself, but we can appreciate how this works by comparing our grasp of truths of human nature with linguistic competence. All native English speakers can recognise that Chomsky's semantically nonsensical sentence, 'Colorless green ideas sleep furiously', is grammatical. Few native speakers can explain why that sentence is grammatical. One use of stimulus sentences like Chomsky's is that they elicit rules speakers intuitively grasp, enabling linguists to formally analyse the structure of those rules. Novels, according to Plumer, function much like Chomsky's stimulus sentence. From the latter we can learn much about syntax, even when the semantic content has no connection to reality. Similarly, novels can teach us about human nature even when the narrative concerns physical or biological impossibilities. When we experience fictional narratives as believable, those narratives engage with principles of human nature we already intuitively, if implicitly, grasp. We express our grasp of those principles by experiencing the novel as believable, much like we express our grasp of syntactical rules in judging sentences grammatical. The reader's or literary critic's task of clarifying why we experience the novel as believable is much like the linguist's task of inducing and then explicitly representing those syntactic rules. We complete this task by working through the transcendental argument of the novel.

The form of a transcendental arguments, as described above, is transparently valid, as an instance of *modus ponens*. But we are concerned that the reading of 'believable' that makes premise (1) highly plausible at the same time renders premise (2) contentious, and more empirically fraught than Plumer perhaps realises. Let's distinguish between *experiential* and *objective* readings of that term, which could be glossed as something like, 'experienced by the reader as psychologically plausible' and 'factually accurate in its portrayal of ways real human agents actually would act, in such circumstances'. It seems that Plumer generally has the experiential reading in mind, e.g., when he describes believability as a 'simple, unanalysed measure of the novel'.[9] Yet rewritten with that reading made explicit, the resulting second premise no longer looks so intuitive:

(1') This novel is experienced by the reader as psychologically plausible.

(2') This novel is experienced by the reader as psychologically plausible only if principles p operate in the real world.

(3) Therefore, principles p operate in the real world.

That is not to say that (2') is anything like *counter*-intuitive – it just makes it usefully clear that it is an empirical claim, linking a kind of experience of the novel on the one hand with objectively true principles of psychology on the other. We would parse that connection between reader's experience and human psychological nature into two major links. First, the experience of believability must track some sort of actual psychological reality. Second, that psychological reality must be something that can be appropriately represented in terms of principles, which can then be realised and articulated, perhaps with varying degrees of reliability, by readers or critics. Let's call the former the *reflection thesis* (RT) and the latter (at perhaps some risk of confusion) the *principle thesis* (PT). The RT gets you from experienced believability to psychological reality, and the PT tells you that that reality is well expressed as principles – just what (2') requires.

Plumer does not really consider the PT thesis explicitly, beyond rejecting any sort of Churchlandian eliminativism.[10] He perhaps takes it to be beyond argument that any such psychological reality must be of a sort that can properly be translated into some sort of general, universal propositions.[11] We won't pursue the point much here, though we would note that it is not an empirically innocent presupposition. It may be that our folk psychological competence is not, after all, so much a matter of internalised, implicit principles *à la* Chomskyan rules of grammar, but more like a grab-bag of tricks and hacks that allow us to do a generally adequate job of coordinating our behaviours. For example, Matthew Ratcliffe has argued that many of these tricks – mutual gaze and use of facial expressions, among others – 'operate *through* [real-time interpersonal] interaction and are not internalised abilities that precede it'.[12] Interpersonal understanding does not result, on his analysis, from the application of general principles to specific circumstances. Nor could that understanding be translated into a set of propositions, as particular features of the second-personal interactive context are what constitute that understanding as an unfolding dynamic between conversational partners. Should the PT be false – as Ratcliffe's work suggests – (2') would be false, and the transcendental argumentative form unsound.[13]

We will focus instead on the RT, which Plumer is concerned to defend at some length, and on largely evolutionary grounds. His argument runs roughly thus: Our evolutionarily honed psychological competence in mind reading underwrites this experience of believability. Humans have lived for thousands of generations in small communities where survival and flourishing requires predicting and understanding others' behaviour. We do this by attributing to

others mental states – beliefs, desires, and intentions – on the basis of what
they say and do to formulate these predictions, the veracity of which signifi-
cantly determined our ancestors' reproductive success. As a result of sharing
this evolved competence, 'we share a significant set of fundamental assump-
tions about human nature'.[14] Of this shared set of fundamental assumptions,
Plumer notes, '[t]here is no alternative possible explanation of their existence
and entrenchment other than that they have evolved in answer to millions
of years of human needs'.[15] These shared assumptions could not have served
human needs of predicting others' behaviour unless they were 'on the whole
reflective of reality'.[16] So, on Plumer's analysis, we approach a novel already
having mastered key facts about human psychology, and our mastery of those
facts is manifest in our experiences of believability.

Granting *arguendo* that there is an underlying psychological reality, and
that as per PT it is capturable in terms of principles at least sometimes, how
can transcendental arguments go wrong? Novels hold our interest by placing
characters in unusual situations, including those totally unknown to our an-
cestors in the environment of evolutionary adaptiveness (EEA) that shaped
our competence in mind reading. For several hundred thousand years our
ancestors lived at subsistence-level in small forager bands, gradually tran-
sitioning to subsistence farming some ten to fifteen thousand years ago.[17]
Only in the past few hundred years has economic growth freed the majority
of humans from a grinding concern with day-to-day survival.[18] Setting aside
outliers like *Robinson Crusoe* (or *Clan of the Cave Bear*), the overwhelming
majority of novels place their characters in environments profoundly differ-
ent from that of the EEA, where it was a daily pressing concern to satisfy
nutritional needs. Our distant ancestors dealt with problems of social class
and courtship, and we would not deny the social world of *Pride and Prej-
udice*, say, bears *some* similarities to those of the EEA. But the differences
between the two are extreme. It hardly follows that an evolved competence
finely attuned to features of the ancestral environment will remain a reliable
predictor of behaviour under industrialised non-subsistence conditions that
are, in evolutionary terms, completely new. Thus we reject (2') as being in
any way a truth that should be obvious from the armchair.

Going forward, we offer empirical grounds for doubting (2') by debunking
the intuitive sense of believability. Just to be clear, our objective is not to
show that (2') is false, but merely that it stands as an open empirical question,
and cannot be presupposed for the purposes of arguments like Plumer's. Our
main argument here will suggest that our sense of psychological plausibility
may not be as reliable a guide to human nature as Plumer's transcendental
argument requires.

II. AESTHETIC DEBUNKING ARGUMENTS & THE FALLIBILITY OF OUR AESTHETIC SENSIBILITIES

In "Aesthetic Debunking Arguments", we offer an initial survey of the prospects for *aesthetic debunking arguments*, in which learning about the nature of the psychological processes underlying our aesthetic engagements threatens to diminish our satisfaction in those engagements.[19] One such form of debunking we consider there regards authorial techniques to gin up an impression of truthful realism:

> Consider James Wood's argument in *How Fiction Works* that writers including Balzac and Flaubert in the 19th century wrote in a realist style by, among other things, filling their fictions with details seemingly unnecessary to advancing the plot. For example, Emma Bovary doesn't simply touch satin slippers, but slippers 'the soles of which were yellowed with wax from the dance floor.' This accumulation of detail gives the novel a 'truthful' feel, which, for some readers, may lead them to treat the novel as a source of truth about the actual world: say, about the monotony of marriage, in the case of *Madame Bovary*. In this case, we come to acquire a belief . . . through the process of experiencing a realist narrative – despite the fact that this realism is just a stylistic trick of Flaubert's. Surely, the accumulation of unnecessary details in a prose work has nothing to do with that work's ability to convey deep truths about the actual world.[20]

In this example, should we come to believe that our impression of veridicality in *Madame Bovary* may be merely the consequence of a Flaubertian layering on of details, we may find ourselves appropriately less inclined to credit the apparent insights of the novel. (We may also find ourselves valuing our experience of the novel accordingly less as well, although questions of aesthetic value are mostly not the topic here.)

Borrowing our argument schema from Nichols 2014 and Kumar and May 2019 debunking arguments have the following form for a belief B and a cognitive process P.

(D1) B is based mainly on P. (empirical premise)
(D2) P is epistemically defective. (normative premise)
(D3) So, B is unjustified.

Debunking arguments have a clear application to Plumer's transcendental argument schema, once we acknowledge his (2') presupposes the truth of the RT, which is itself an empirical claim. When we believe *Madame Bovary* reveals aspects of psychological reality on the basis of our experiencing it as believable, that experience of believability has a certain etiology in the reading experience. Sound transcendental arguments require that this etiology results from epistemically reliable processes for forming beliefs about

psychological reality. One species of aesthetic debunking arguments uncovers the actual cognitive processes generating the relevant belief (D1), and then shows these processes to be unreliable belief-forming mechanisms (D2). Aesthetic debunking arguments show our beliefs in propositions about the world, which the novel purportedly reveals as true, to be unjustified – insofar as those beliefs rely for their justification on the reading experience.

Now, as already noted above, believability and realism are distinct characteristics, and we are not suggesting that the exact same concern from Wood will apply to Plumer's argument. Nonetheless, it opens the door to what we take to be the right kind of question to ask, in considering Plumer's RT: what kinds of potentially misleading cues, nudges, or heuristics might drive our impressions of believability, such that a novel may be *experienced* by a reader as psychologically plausible, when the portrayal of human agency is, in fact, either inaccurate *tout court*, or when we are unjustified in deeming it either objectively plausible or implausible? In short, how might the experiential and objective notions of plausibility come unstuck from each other, where novels can give rise to the former without truly possessing the latter?

Before we try to articulate a few such candidate ways in which these notions can become unstuck, we would offer two caveats about what we take the rules of the dialectic to be here. First, we are not affirming that any of these hypotheses are true, or even more likely to be true than the RT. Our objective is only to show that the RT is not as empirically trivial as Plumer hopes, and that it does not follow as trivially as he would like from broad evolutionary considerations. Second, however, it's not enough for us to just make up some merely metaphysically or logically or conceptually possible scenario. We take it that our hypotheses need not be shown to be true, but nonetheless they need to be empirically live, and consonant with good scientific evidence.

III.   DON'T BELIEVE EVERYTHING THAT SEEMS BELIEVABLE: PROCESSING FLUENCY

In this section and the next, we identify two ways, well-motivated by the empirical literature, in which the RT can turn out false, causing the transcendental argument scheme to be unsound. First we consider how our sense of the psychological believability of characters in novels is subject to distortion from the *cue of processing fluency*. Next we consider how the *illusion of explanatory depth* could drive literary critics to unjustifiably treat believable fictional characters as instantiating general truths of human nature.

The *cue of processing fluency* turns on our feeling of how easily a well-wrought narrative is to follow, digest, and enjoy. Contrast the feeling of how easy it is to read a phrase when written in a simple sans-serif font, contrasted with one written in a more challenging *script-like font*. That feeling turns out to enter into a large number of cognitive transactions, and our minds

recruit that feeling as a cue that, roughly speaking, things are going well – that feeling of easy thought is used as a positive heuristic for all sorts of evaluation. The fluency heuristic has a distinct, and sometimes significantly stronger, influence over what we choose in, say, consumption bundles, than do the actual substantive features of the selected options.[21] While, as you might imagine, there is a vast literature on these effects in the consumer choice literature, it has also been of interest in more aesthetic domains as well. For example, increasing feelings of fluency can augment the humour of a joke.[22] Moreover, processing fluency has been hypothesised to explain the 'mere exposure effect', in which people's preferences for one painting or another can be manipulated by their simply having seen the preferred one more often in the past;[23] though see Meskin et al 2013, and Liao and Meskin (this issue) for some recent aesthetics x-phi that complicates that story.[24] Processing fluency has even been posited as playing a central role in aesthetic responses such as judgments of beauty, though this is perhaps still more conjectural than many other psychological effects of fluency.[25]

Given our topic here, we are interested not so much in forming preferences or making choices, as we are in judgments about what is true or false. And indeed fluency can play a substantial role in our metacognitive evaluation of how true some message is, how credible the argument, how trustworthy the source.[26] A lovely illustration is that aphorisms that rhyme – and are thus phonologically fluent – are perceived as more insightful than aphorisms that do not; e.g., 'Woes unite foes' vs. 'Woes unite enemies').[27] In a classic review paper, Alter and Oppenheimer write: 'If the experience of fluency is important regardless of the associated content, then the ease or difficulty associated with a broad range of cognitive processes should have similar effects on judgment. Indeed, Schwarz later noted that, 'theoretically, any . . . variable that increases processing fluency should have the same effect'.[28] Supporting this claim, researchers have manipulated processing fluency using numerous approaches, including semantic priming, phonological priming, or visual clarity.[29] The authors list an example of each type of study, and then go on to argue that although each 'manipulation influences a different proximal cognitive system', nonetheless, 'all three studies found identical effects: Participants rated fluent stimuli as more true or accurate than similar but less fluent stimuli'.[30]

There are many ways of generating feelings of fluency or disfluency, because there are many aspects of engaging with a stimulus that can be comparatively easy or hard to do. Just in terms of reading a text, we gave the example of how easy or hard a font may be to read, but also one could manipulate other visual aspects of a reading task, such as whether the letters and the background contrast sharply or are in vexingly nearby shades of the same hue.[31] The words themselves could be more or less familiar, and the concepts appealed to could be more or less abstruse. Outside of what one might think of as belonging to the task of the reading proper, other aspects

of the setting can make reading feel easier or harder, for example, if there are distracting auditory stimuli as well.

So, the empirical literature is fairly clear at this point that the processing fluency cue can increase how persuasive we find a text; and it can be triggered by a wide range of manipulations. We need two more pieces in place, to produce a debunking hypothesis for Plumerian believability. First, our hypothesis will include that experiential believability is itself something that is susceptible to this cue – that more fluency would mean, on average, more experiential believability. This seems to us not much of a stretch, given the sheer breadth of documented fluency effects, and that, again, we only need our hypotheses to cross a bar of well-motivated empirical plausibility. So we will simply conjecture that the experience of psychological believability is indeed susceptible to fluency cues.

Second, even granting ourselves that connection, we still need to offer a decent empirical reason to suspect that novels may be particularly susceptible to this cue. Is there something about a novel, at least a well-written and, indeed, believable one, that should lead us to predict it will score highly in terms of fluency? Without this last piece, then even if the fluency cue could in principle create illusions of increased believability, we wouldn't have grounds to suggest that such illusions could afflict the novels from which Plumer wants us to extract transcendental arguments.

This last missing piece is provided by *narrative.* Expert novelists engineer their plots to be engrossing, even transporting, and they have an array of tools to bring about that effect.[32] And not least among those tools is a well-crafted plot that holds the reader's attention without frustrating it. In their synopsis of the literature, Vaughn et al. state that '[m]any models of narrative impact posit that transportation is a key factor in narrative persuasion . . . . Studies often show a positive relationship between transportation and subsequent story-consistent attitudes and beliefs'.[33] Through some clever experimental manipulations, they found that in general, increasing the processing fluency of a story even further intensified its persuasive power. The more fluent our experience of reading the text, the more likely we are to agree with its contents.[34]

Now, the reader may well object that novels from the postmodern or even just late modern period typically frustrate the reader, and works such as *Finnegans Wake* or *Gravity's Rainbow* strike the reader as having been exquisitely crafted to *avoid* any easy digestion. Yet we think this phenomenon rather strengthens our case here, for such novels are not among the ones that would rate highly in terms of psychological believability. Whatever we are to take away from the adventures of Pynchon's Tyrone Slothrop, our appreciation for the fine points of his psychology does not seem among them. Books that are meant to be difficult will not generally be ones that provide possible sources for a Plumerian transcendental argument. Among other books that may be narratively shambolic, but still present as psychologically believable,

we should also consider whether their authors have tapped into other sources of fluency, such as clever writing and humour. (*Tristram Shandy*, perhaps?)

We thus have all of our pieces in position for this first debunking hypothesis. Plumer claims that novels that occasion experiential believability must be doing so because our minds really do work like they are shown to work in the novels. Yet perhaps experiential believability is driven in no small part by our processing fluency in reading the text, and that fluency is typically amplified by all the tricks and tactics of the writer's craft and genius, not least of all, their capacity to spin a good, absorbing yarn. Even if sometimes great novels will be experientially believable because they embody insights about human cognition, sometimes they will be experientially believable *simply because they are great novels.*

IV.   DON'T BELIEVE EVERYTHING THAT SEEMS BELIEVABLE: EXPLANATORY DEPTH

Our second debunking hypothesis challenges the move from a novel's experienced believability to viewing that novel as capturing truths of human nature. For our second debunking hypothesis, we suggest that philosophers and literary critics, when they regard a work of fiction as expressing general truths of human nature, may be in the grip of the *illusion of explanatory depth.*

The illusion of explanatory depth is a general tendency for people to feel they understand the world better than they actually do. Rozenblit and Keil have documented this tendency in subjects' self-reported understanding of the internal workings of artifacts and natural processes.[35] And the effect can be found when we deploy too-simple causal models not just in mechanical but in psychological domains as well, such as the workings of public policy,[36] belief in conspiracy theories,[37] and the workings of mental disorders.[38] These studies elicit subjects' self-reported understanding of causal mechanisms before and after exposure to expert explanations of those mechanisms, consistently finding that subjects were initially overconfident in their explanations.

Importantly, Rozenblit and Keil's study found no evidence of overconfidence in an understanding of narrative plots.[39] They suggest this may be due to different ways we represent narratives from how we represent causal systems. Our mental representations of causal systems are low-fidelity and miss many crucial components, which becomes increasingly likely as the system's internal complexity increases.[40] In contrast, narratives are crafted, for the most part, so that consumers can follow them. Some may be difficult to follow, but they are rarely designed to be inscrutable (the probable likes of *Finnegans Wake* and *Gravity's Rainbow* excepted, as noted above). Whatever features of our mental representations generate the illusion of understanding, those features are not necessarily shared by our representation of narratives.

Rozenblit and Keil do not predict explanatory overconfidence across the board, but they do expect it in the situations we are in as readers of psycho-

logically rich novels. In their experiments, the strongest predictor of overconfidence is a high ratio of visible to hidden parts in the thing explained. They speculate that 'the prominence of visible, transparent mechanisms may fool people into believing that they have understood, and have successfully represented, what they have merely seen'.[41] And while minds are complex mechanisms with causal components not available to immediate scrutiny, novels wonderfully make many of those parts visible to the reader, and in a way that can give a strong impression of 'watching the gears turn', as it were. The form of the novel gives us intimate access to the minds of the characters by directly representing their inner dialogue. It's as if we are witnessing when one course of action pivots on a fulcrum into a new behavioural direction, and can see the mechanisms bringing about that change. In addition to our direct access to their stream of consciousness, fictional characters' minds have a neatness to them missing in actual minds. Our minds are evolved biological systems, a mess of kluges with many drives and motives hidden from scrutiny. Fictional characters, in contrast, are made out primarily of words chosen by the author. The author may stipulate what her characters' drives and motives are, giving us access to them as well as to the character's personal history as it has shaped her motivational set. Through the novel's lens, everything about the character's mind can be placed in our view.

To be clear, our suggestion is not that readers are overconfident in their experience of understanding a character's mind. A good author of the sort of novels Plumer discusses will typically succeed in giving the attentive reader a highly accurate sense of their creation's psychology. Rather, the high ratio of visible to hidden parts in the novel's depiction of mental life primes readers to have an illusory sense of understanding *actual psychological reality* on the basis of their experience of the fiction. Much as Rozenblit and Keil's subjects are overconfident, when first pressed, that they understand a crossbow's workings, so, we suggest, are readers when they infer truths of human nature from experienced believability.

But Rozenblit and Keil's subjects recognise the simplicity of their mechanistic models when offered an expert's description of how a crossbow works. For readers, it is not clear what counts as the 'human nature expert', of a sort whose proposed account of actual human psychological working we would allow to trump our experiential, intuitive understanding as easily as the expert in crossbows or bicycles or microwave ovens. After all, even highly successful psychologists may be poor students of human nature as such – perhaps possessing a deep grasp of this or that specific mechanism while missing the larger picture – and expert observers of human nature may have a poor grasp of those sub-personal processes that drive much of our behaviour.

Perhaps one might think the very best novelists come as close as anyone to having expertise in human nature. Yet we should not uncritically treat their novels as displaying that expertise. As Gregory Currie observes, novelists do not solely pursue psychological verisimilitude in representing their characters.

Their other goals include shaping characters to respond to dramatic necessity. He writes, 'a great deal of speculation and debate goes on about the extent to which these characters are motivated from within – by plausibly human models of deliberation and feeling – or from without, by dramatic necessity, or poetic inclination, or the need to make the always difficult fourth act work, or whatever'.[42] Novelists simultaneously pursue all these aims at once, making it impossible to fully untangle those parts of the fiction that aim at proper psychological verisimilitude from those that don't. Careful readers and expert literary critics sometimes fail to reach any consensus on when one aim is sacrificed for another; Currie cites the different positions Coleridge, Empson, and Bradley adopt on the plausibility of Iago's scheming malignity as an example.[43] Because we cannot fully untangle these aims, we shouldn't treat great novels, which tell great stories, as though they were performing the task of explaining how minds work by highlighting a subset of causal factors.[44]

The illusion of explanatory depth is one of many phenomena that challenge Plumer's rationale for the RT. Recall that rationale: mindreading competences become entrenched by serving human needs, and that they could not do this unless those competences delivered predictions of human behaviour that were 'on the whole reflective of reality'.[45] This view presumes our cognitive capacities evolved to deliver increasingly reliable judgments about the world. Hugo Mercier and Dan Sperber offer an alternative story of how our cognitive capacities became evolutionarily entrenched – a story that, while compatible with the RT's truth, leaves its truth an empirically open question. On their 'argumentative theory of reasoning', cognitive mechanisms, including mindreading capacities, evolved for 'justifying oneself, and ... producing arguments to convince others'.[46] Cognition in general, including mindreading competences, they argue, developed for use in an interpersonal setting where participants aimed at coordination and persuasion, with these aims often conflicting with that of truth-tracking. Mercier and Sperber's framework predicts a general 'myside bias' in reasoning towards justifications supporting the reasoner's point of view.[47] This bias is compatible with reasoners accounting for, and changing their views in response to, counterevidence when it is presented. But as a design principle of the mind, this bias predicts many of the quirks in our representational and reasoning systems empirical psychologists have recently uncovered, including the illusion of explanatory depth.[48]

Our second debunking hypothesis can stand (or fall) independent of Mercier and Sperber's argumentative theory of reasoning. But that theory, assuming its truth, provides an alternative evolutionary rationale to the RT for why many readers feel confident novels convey truths of human nature. We've argued it is a live hypothesis that this feeling of confidence may often be a product of the illusion of explanatory depth. While Mercier and Sperber do not explicitly present it as such, the illusion of explanatory depth can be construed as a manifestation of the general myside bias they describe.

Overconfidence helps us sell ourselves to others in interpersonal contexts as grasping a mechanism's workings: thus we come off as well informed and promising potential cooperative partners. Although reading is a solitary activity, we propose that extracting general principles of human nature from fiction activates our social tendency to 'sell ourselves' as knowledgeable. This general disposition, which operates in us below the threshold of conscious awareness, would have been fitness-enhancing for our ancestors, who never experienced, even in communal storytelling sessions, anything like the rich interiority of a great novel. The experience of believability may have less to do with accessing an evolutionarily entrenched shared 'set of fundamental assumptions about human nature' than with activating an atavistic tendency, which misfires in a solitary context, to make ourselves look good to others.[49]

V. SOME LESSONS FOR A NATURALISTIC ACCOUNT OF AESTHETIC EPISTEMOLOGY

We will conclude with two brief methodological lessons, and a highly speculative suggestion.

First, we take ourselves to be reinforcing here a lesson that Currie in particular has been interested in pressing upon the community of aestheticians: namely, that the work of scientific psychology should give us serious pause – should make us slam on the brakes, in fact – whenever we look to transition from experienced psychological confidence to endorsed epistemic credence about the nature of the mind. Samuel Johnson, Currie writes, 'was delightfully confident in his opinions of many things, and rated himself a great observer of his fellow creatures'.[50] Like Johnson, we generalise about human nature with little more support for our claims than our own confidence in their truth. Currie observes that this hardly seems a good epistemic practice given 'the last 50 years of psychological investigation'.[51] Our arguments here build on Currie's observation.

Second, and relatedly, we would note that Plumer's argument is at heart an abductive one. He invites us to ask the question, *what else* could explain our sense of believability in these works, if not their real, objective believability? We take this to be a case study in the importance of letting the sciences help guide our sense of what possible answers there could be to such a 'what else?' question. To be clear, we are not objecting that Plumer's own preferred hypothesis is itself in any way *un*scientific. It is an empirical hypothesis in good standing, and he musters various considerations that we agree should motivate its consideration, such as his appeals to evolution. But it is a yet further step, and a large one, from hypothesis to truth, and our concern is that his inference to the truth of that one scientific hypothesis depends on keeping other equally well-motivated scientific hypotheses out of view.

Finally, Peter Kivy was fond of emphasising the importance of attending to the differences across different art forms, and avoiding overgeneralising

about all of the arts on the basis of too narrow a sample.[52] We will close by speculating on the consequences for Plumer's argument if we pan out from novels of a psychologically rich type, and cut over to a different medium altogether that may offer divergent psychological lessons. We concur with Plumer's critical observation that the greatest exemplars of such novels do, for certain, give us a strong feeling of insight into the nature of their characters' minds, and thereby into human psychology on the whole. Yet there are other artworks, equally great in their own categories, that seem to offer a contradictory lesson about the human mind: that we are all fundamentally unknowable to each other. Cinema is an especially good vehicle for conveying this sense of psychological opacity, as the medium does not give viewers direct access to characters' minds (setting aside the occasional use of voice-over). Consider for example *Vertigo*'s Scottie, who is motivated by strange, all-encompassing obsessions and passions that, while believable, are inscrutable to the audience and, presumably, to the characters – including Scottie – in the fictional world as well.[53]

If this observation upon cinema is correct, then at best, Plumer's argument would be transposed from the Kantian tune of the transcendental argument, into a different and rather more minor key of the antinomy. If a masterwork novel can teach us how we most fundamentally make sense to each other, and at the same time, a brilliant film can teach us that we are fundamentally, always already estranged from each other ... then perhaps each kind of work should serve as a *reductio* of the other, and we should abandon the idea of learning through the arts.

And yet! *Something* does *seem* right, and deeply so, in all of these great works, does it not? While our debunking arguments here should make plausible the possibility that this seeming is delusional, nonetheless, we do not at all think that such a conclusion is entailed by these arguments. We are advocating that a naturalistic perspective on our aesthetic lives poses puzzles about how to reconcile a scientific self-understanding with the kind of epistemic and aesthetic promises that the arts have seemed to make to us. There is a tension between what the sciences tell us we are, and what the arts tell us to hope for. We suspect that the best way forward is not to plump for either side of that push-and-pull, but rather, to articulate a framework that ultimately allows us to reside, even if not always entirely comfortably, within that tension. Reckoning with Plumer's transcendentally argumentative novels may be an excellent starting point for such a project.[54]

adamgjesdal@gmail.com

jmweinberg@email.arizona.edu

NOTES

[1] See, e.g., Lamarque and Olsen 1994 for one classic interrogation.

[2] Plumer 2017.

[3] One should distinguish 'realism' as a term for a literary style, from what Plumer has in mind as synonymous with 'being realistic', as in keeping true to the particulars of the actual world. 'Believable', in terms of psychological plausibility, crosscuts both, in that a genre with fantastical or science-fictional elements cannot be considered realistic, in Plumer's sense, although it may or may not be believable. And a work in the genre of 'magic realism' will by definition not be realistic *sensu* Plumer, yet typically its characters act highly believably in the, to our eyes, strange worlds they inhabit.

[4] Stroud 1968, 242.

[5] Plumer 2017, 153.

[6] Plumer 2017, 152.

[7] Plumer 2017, 155.

[8] Plumer 2017, 155.

[9] Plumer 2017, 155.

[10] Plumer 2017, 162-163.

[11] He is particularly concerned that the principles might be merely relative; see his section 5.

[12] Ratcliffe 2006, 38, italics original.

[13] See also, e.g., Spaulding 2018 for another picture of mindreading that may not be amenable to PT, in particular her focus on *models* over *theories*.

[14] Plumer 2017, 162.

[15] Plumer 2017, 162.

[16] Plumer 2017, 163.

[17] Morris 2017, 11, 28.

[18] Jones 2016.

[19] Gjesdal and Weinberg, forthcoming.

[20] Wood 2008, 69.

[21] E.g., Lee and Labroo 2004.

[22] Topolinski 2014.

[23] Cutting 2003; Oppenheimer 2008.

[24] Meskin et al. 2013, and Liao, Meskin, and Fletcher 2020.

[25] Reber, Schwarz, and Winkielman 2004.

[26] Experiences of fluency can also serve directly as inputs into the contents of judgments themselves. For example, turning up the fluency can turn down our evaluation of the moral wrongness of a transgression (Laham, Alter, and Goodwin 2009). And there is at least some evidence that it can shift our attributions of when an agent performs a free action, where an easier scenario to think about yields an increased rating of how easily an agent could have done otherwise – even when the fluency manipulation is something utterly incidental to the scenario's content, like choice of font (Weinberg et al. 2012). But we will be more interested here in the metacognitive use of the cue, to treat a message or source as more credible, than in its potential role as a first-order heuristic in the content of a judgment.

[27] McGlone and Tofighbakhsh 2000.

[28] Schwarz 2004, 338.

[29] Alter and Oppenheimer 2009, 219-220.

[30] Alter and Oppenheimer 2009, 220. A recent paper suggests a different role for this variety of cue, in which a feeling of *dis*fluency could lead participants to be more attentive and reflective, more likely to pick up on subtle errors than those who were experiencing a task more fluently. These results would suggest that disfluency could lead to greater cognitive accuracy (Hyunjin and Schwarz 2008; Diemand-Yauman, Oppenheimer, and Vaughan 2011). Unfortunately, it seems that these findings have not replicated well (Kühl and Eitel 2016). To our knowledge, however, there have not been similar replication concerns raised about the main mechanism we are relying on in our arguments here, to the effect that a greater feeling of fluency could lead to a greater sense of believability.

[31] Reber and Schwarz 1999

[32] One trick is using the environment as an 'objective correlative' to create a general atmosphere expressing characters' inner states. Consider Dickens' use of the thick London fog in *Bleak House* to emphasise the characters' moral corruption and confusion. Or, switching to plays, Shakespeare's use of the storm in *King Lear* to capture Lear's inner state. These metaphorical devices, we suspect, aid processing and enhance characters' believability by giving us environmental cues to their

psychological states, making those states creep out from the characters' minds into the atmosphere itself.

[33]Leigh et al. 2010, 1183.

[34]We should caution that we have, of necessity, presented a rather simplified picture of this corner of the empirical literature. There is a long and varied set of interactions that can complicate the effect of different manipulations. For example, somewhat akin to the Meskin et al. finding about mere exposure to *bad* art, it looks like *un*successful stories can be made even less transporting by increasing fluency. Basically, the increased fluency appears to make readers more confident in their judgments that the stories are not so good, and leads them to disengage. See Leigh et al. 2010, 1188.

[35]Rozenblit and Keil 2002.

[36]Fernbach et al. 2013.

[37]Vitriol and Marsh 2018.

[38]Zeveney and Marsh 2016.

[39]Rozenblit and Keil 2002, 543-544.

[40]Rozenblit and Keil 2002, 555-557.

[41]Rozenblit and Keil 2002, 552-554.

[42]Currie 2012, 28.

[43]Currie 2012, 28.

[44]Lombrozo 2006

[45]Plumer 2017, 163.

[46]Mercier and Sperber 2017, 8.

[47]Mercier and Sperber 2017, 220.

[48]Mercier and Sperber 2017, 102-103.

[49]Plumer 2017, 162.

[50]Currie 2012, 30.

[51]Currie 2012, 30.

[52]As in, e.g., his Kivy 1993 'Differences'.

[53]For a general discussion of these issues in Hitchcock's films, see Pippin 2017.

[54]The work is the equal responsibility of the co-authors, and author order is alphabetical.

REFERENCES

Alter, Adam L., and Daniel M. Oppenheimer. 2009. "Uniting the tribes of fluency to form a metacognitive nation." *Personality and social psychology review* 13 (3): 219–235.

Currie, Gregory. 2012. "Literature and truthfulness." Chapter 3 of *Rationis Defensor: Essays in Honour of Colin Cheyne*, edited by James Maclaurin, 23–31. Dordrecht: Springer.

Cutting, James E. 2003. "Gustave Caillebotte, French impressionism, and mere exposure." *Psychonomic Bulletin & Review* 10 (2): 319–343.

Diemand-Yauman, Connor, Daniel M. Oppenheimer, and Erikka B. Vaughan. 2011. "Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes." *Cognition* 118:114–118.

Fernbach, Philip M., Todd Rogers, Craig R. Fox, and Steven A. Sloman. 2013. "Political extremism is supported by an illusion of understanding." *Psychological science* 24 (6): 939–946.

Gjesdal, Adam, and Jonathan Weinberg. (forthcoming). "Aesthetic Debunking Arguments." In *Art and Philosophy*, edited by Alex King and Christy Mag Uidhir. Oxford: Oxford University Press.

Hyunjin, Song, and Norbert Schwarz. 2008. "Fluency and the detection of misleading questions: Low processing fluency attenuates the Moses illusion." *Social Cognition* 26 (6): 791–799.

Jones, C. I. 2016. "The facts of economic growth." Chapter Chapter 1 of *Handbook of Macroeconomics*, edited by Harald Uhlig John B. Taylor, Volume 2, 3–69. Amsterdam, The Netherlands: Elsevier.

Kivy, Peter. 1993. "Differences." *The Journal of Aesthetics and Art Criticism* 51 (2): 123–132.

Kühl, Tim, and Alexander Eitel. 2016. "Effects of disfluency on cognitive and metacognitive processes and outcomes." *Metacognition and Learning* 11 (1): 1–13.

Kumar, Victor, and Joshua May. 2019. "How to Debunk Moral Beliefs." In *Methodology and Moral Philosophy*, edited by Jussi Suikkanen and Antti Kauppinen, 25–48. New York: Routledge.

Laham, Simon M., Adam L. Alter, and Geoffrey P. Goodwin. 2009. "Easy on the mind, easy on the wrongdoer: Discrepantly fluent violations are deemed less morally wrong." *Cognition* 112 (3): 462–466.

Lamarque, Peter, and Stein Haugom Olsen. 1994. *Truth, Fiction and Literature: A philosophical perspective*. Oxford: Clarendon Press.

Lee, Angela Y., and Aparna A. Labroo. 2004. "The effect of conceptual and perceptual fluency on brand evaluation." *Journal of Marketing Research* 41:151–165.

Leigh, Ann Vaughn, Kathryn E. Childs, Claire Maschinski, N. Paul Ni no, and Rachael Ellsworth. 2010. "Regulatory fit, processing fluency, and narrative persuasion." *Social and Personality Psychology Compass* 4 (12): 1181–1192.

Liao, Shen-yi, Aaron Meskin, and Jade Fletcher. 2020. "The Vanity of Small Differences: Empirical Studies of Artistic Value and Extrinsic Factors." *Aesthetic Investigations* 4:412–427.

Lombrozo, Tania. 2006. "The structure and function of explanations." *Trends in cognitive sciences* 10 (10): 464–470.

McGlone, Matthew S., and Jessica Tofighbakhsh. 2000. "Birds of a feather flock conjointly (?): Rhyme as reason in aphorisms." *Psychological science* 11 (5): 424–428.

Mercier, Hugo, and Dan Sperber. 2017. *The enigma of reason*. Cambridge, Massachusetts: Harvard University Press.

Meskin, Aaron, Mark Phelan, Margaret Moore, and Matthew Kieran. 2013. "Mere exposure to bad art." *British Journal of Aesthetics* 53 (2): 139–164.

Morris, Ian. 2017. *Foragers, farmers, and fossil fuels: How human values evolve*. Princeton, New Jersey: Princeton University Press.

Nichols, Shaun. 2014. "Process debunking and ethics." *Ethics* 124 (4): 727–749.

Oppenheimer, Daniel M. 2008. "The secret life of fluency." *Trends in cognitive sciences* 12 (6): 237–241.

Pippin, Robert B. 2017. *The Philosophical Hitchcock:'Vertigo' and the Anxieties of Unknowingness.* London: University of Chicago Press.

Plumer, Gilbert. 2017. "The transcendental argument of the novel." *Journal of the American Philosophical Association* 3 (2): 148–167.

Ratcliffe, Matthew. 2006. "'Folk psychology' is not folk psychology." *Phenomenology and the Cognitive Sciences* 5 (1): 31–52.

Reber, Rolf, and Norbert Schwarz. 1999. "Effects of perceptual fluency on judgments of truth." *Consciousness and Cognition* 8 (3): 338–342.

Reber, Rolf, Norbert Schwarz, and Piotr Winkielman. 2004. "Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver's Processing Experience?" *Personality and Social Psychology Review* 8 (4): 364–82.

Rozenblit, Leonid, and Frank Keil. 2002. "The misunderstood limits of folk science: An illusion of explanatory depth." *Cognitive science* 26 (5): 521–562.

Schwarz, Norbert. 2004. "Metacognitive experiences in consumer judgment and decision making." *Journal of Consumer Psychology* 14 (4): 332–348.

Spaulding, Shannon. 2018. *How We Understand Others: Philosophy and Social Cognition.* New York: Routledge.

Stroud, Barry. 1968. "Transcendental arguments." *The Journal of Philosophy* 65 (9): 241–256.

Topolinski, Sascha. 2014. "A processing fluency-account of funniness: Running gags and spoiling punchlines." *Cognition and Emotion* 28 (5): 811–820.

Vitriol, Joseph A., and Jessecae K. Marsh. 2018. "The illusion of explanatory depth and endorsement of conspiracy beliefs." *European Journal of Social Psychology* 48 (7): 955–969.

Weinberg, Jonathan M., Joshua Alexander, Chad Gonnerman, and Shane Reuter. 2012. "Restrictionism and reflection: Challenge deflected, or simply redirected?" *The Monist* 95 (2): 200–222.

Wood, James. 2008. *How fiction works.* New York: Farrar, Straus & Giroux.

Zeveney, Andrew S., and Jessecae K. Marsh. 2016. "The Illusion of Explanatory Depth in a Misunderstood Field: The IOED in Mental Disorders." Edited by A. Pagafragou, D. Grodner, D. Mirman, and J. C. Trueswell, *Proceedings of the 38th Annual Conference of the Cognitive Science Society.* Cognitive Science Society, Austin, Texas, 1020–1025.