

# Aesthetic Interpretation\*

Elin McCready  
Aoyama Gakuin University

November 7, 2023

## Abstract

This paper proposes a semantic framework for making sense of aesthetic communications. Literature is a paradigm example of a text type that does not aim at conveying truth-conditional content, but has other communicative goals; this paper aims toward a first step in the understanding of texts beyond the truth-conducive. The paper begins with the proposal of an ‘aesthetic interpretation,’ a semantics for literary texts that bases their impact on a generalized notion of affect rather than truth-conditions. It takes affective and emotive responses to literary texts to be derived via a function yielding the affect a particular agent derives from a text, itself derived from three functions corresponding to different sources of affect. It then proposes notions of significance and sincerity that codify some aspects of Gricean cooperation for aesthetically interpreted texts.

## 1 Introduction

This paper is about the language of art and literature, which has, in the sense it will be discussed here, not been the focus of much work in formal linguistics and philosophy. The tools of semantics, pragmatics, and analytic philosophy of language arise from the analysis of language aimed at conveying truths about the world, and

---

\*Thanks to audiences at ZAS Berlin, Humboldt Universität, Utrecht University, University of Paris 7, Tübingen University, DGfS Workshop on Linguistic Aspects of Fictionality, Pompeu Fabra, University of Buenos Aires, University of Bochum, and University of Erlangen-Nuremberg. Thanks particularly to Chris Davis, Quentin Dénigot, Daniel Gutzmann, Daniel Harris, Stefan Himmerwinter, Paloma Jeretic, Stephanie Solt. . .

from the analysis of communicative acts aimed at such communication. At this point in the development of these fields we have a pretty good understanding of a wide range of facts in these areas, and formal analyses with the capacity to ground them. This is a success story, but a limited one.

Not all language, and maybe not even most language, is aimed at communication of this kind. What, for example, is one to say about conversations involving small talk and bullshitting? What about jokes? These are normal social behaviors, but semantics, pragmatics, and philosophy of language have not had much to say about them, despite some attention (such as Frankfurt 2005 on bullshit). This situation maybe makes sense given that the roots of analytic philosophy, where in early days the motivation was not at all to understand what's happening in everyday conversation, but to make sense of 'larger' notions like truth, logical validity, and the language-world relation; semantics and pragmatics, having developed in part from this field, share these foundational assumptions. But this picture leaves out a great deal of language use. It might even leave out an aspect of language that we might think is in some ways the most important of all: the way language is used in art.

Formal semantics and philosophy of language have had little to say about the language of art. There is extensive work on truth in fiction, which is mostly a pretty straightforward extension of the formal semantic toolkit and its concern with language-world connections to the worlds developed within fiction; this line can be traced back to the classic Lewis 1978, and remains constant in (for instance) many of the recent papers on the topic in Maier and Stokke 2022. But this kind of work is not concerned with what fiction aims to do, but more with how it does it: the question of whether (1) is true is presumably not the main concern of any reader even of this kind of genre fiction, much less of work with more lofty literary goals, where the situations described are often secondary to the aims of the writer.

(1) Sherlock Holmes lives at 16 Baker Street.

The situation becomes more obviously fraught when one starts to consider poetry, where truth doesn't in general even seem an operant question, since for much or even most poetry there isn't really an attempt to construct fictional universes of the kind we see in detective fiction or the like. Maybe the closest thing on the market are theories of metaphor, but even here the goal tends to be to rationalize interpretation by bringing it closer to 'vanilla' truth-conducive communication, rather than to try to make sense of the goals and methods of poetry in a way that

takes them on their own terms.

But what kind of theory do we need for the language of art? Can't we just adopt a version of truth-conditional semantics? This family of theories has a standard answer to what it is for a sentence to be true, usually stated using disquotational definitions, which assume the existence of a semantic theory able to provide logical forms of some kind for sentences:

- (2) 'The cat is on the mat' is true iff the cat is on the mat.

The precise way in which the logical forms resulting from interpretative processes – the word, in the sense of 'direction of fit' – are related by this definition to the world depends on the background theory. Model-theoretic semantics makes reference to abstract mathematical representations of the world; for (2), there must be an object which is in the extension of the predicate *cat* and one which is in the extension of the predicate *mat*, and the two must stand in the *on* relation to one another, ie. lie in the extension of that relation. Alternatively, theories like that of Heim and Kratzer (1998) leave out the models and look directly to the world for satisfaction of truth conditions. But in each case, the sentence is meant to relate to the actual facts, or at least to our representations of them, which may of course be fallible.

The key point for our purposes is that, in theories of meaning of this kind, there is a mapping from the logical form of the sentence to the world which allows a judgement of truth or falsity. But the *point* of the language of fiction is not to communicate factual truths, but rather to do other things. This means that it's not obvious that truth conditions and the kinds of disquotational theories they stem from are the right tools to analyze fiction, or artistic language more generally. This paper aims to take a different starting point and produce a different toolkit for aesthetically intended language, rather than trying to juryrig one from tools initially developed to model information transmission. The resulting theory can run in parallel with truth-conditional semantics, but simultaneously produces a different kind of output.

The starting point for this project depends on what we take the point of literature to be, or rather what we take the content it communicates to depend on. For more standard cases of information transmission, what's communicated are facts (or purported facts) about the world<sup>1</sup> and so truth-conditions serve as a suitable theoretical base for modeling this kind of communication. For literature,

---

<sup>1</sup>The content of interrogatives and imperatives can also be made sense of in terms of facts: as sets of facts, or commitments to induce certain facts, for example.

the situation is different. What's communicated can be somewhat various. One possibility is the communication of metaphorical truths and framings, involving the author's worldview or a distillation of their knowledge of patterns and significance. Another is the production of affective value, in a broad sense, which often comes together with the transmission of framings and worldviews. Other kinds of literary and artistic practice may have different goals, at least in part, as with conceptual art and various literary avant-gardes. This paper will take affective transmission as the base case for construction of a theory of aesthetic interpretation; I will return to metaphorical and symbolic interpretation in the conclusion, where I show how it can be informed by the theory I develop.

The paper proceeds as follows. The following section, §2, is on semantics and perspective. §2.1 begins by giving a semantics for art-text. I begin by presenting a base theory of how aesthetic interpretation proceeds, which I argue to be via the generation of affect (of a sort broader than standard notions of emotivity) in a way modelable in a quasi-compositional and dynamic manner. I then show in §2.2 that a notion of agent-relativity must be introduced to the model to make sense of the affect extracted from a text by different interpreters, including the author. This idea in turn plays into notions of aesthetic significance, which can be understood as a correspondent of truth for the literary domain. The next section, §3, builds on the semantics introduced to construct a partial pragmatics for art-text. The notion of significance is already in play, and can be viewed as corresponding to Gricean Relevance for the aesthetic domain. But can we find an analogue of trustworthiness or testimonial reliability here that can play the role of Quality? I argue that we can, and a notion of aesthetic sincerity besides, building on other work in pragmatics using the theory of repeated games. This section also discusses other aspects of pragmatic interpretation. §4 sums up and discusses some remaining issues and prospects for the future, including how other aspects of broadly Gricean cooperation might look in aesthetic pragmatics.

## 2 Aesthetic semantics

If we want a semantics for literary texts that takes something other than information transmission to be fundamental, we must move away from truth-conditions as the base of our semantic theory. Taking truth-conditions as central assumes that the central purpose of language is to describe facts about the world. We thus need something different to ground interpretations. The starting point of my analysis in this paper is that aesthetic interpretation arises by keying valuations to affect

rather than truth. This idea is spelled out in 2.1, and elaborated in 2.2, which considers the question of whose affect is relevant to making sense of aesthetic significance and meaning, concluding that the answer depends on the goals of the current interpretative project, and that no agent is privileged over any other.

## 2.1 Aesthetic valuation

According to the theory advocated for in this paper, the key to making sense of aesthetic valuation is whether the linguistic object triggers affect in the hearer. It should be clear from the outset that the notion of affect I have in mind is broader than standard emotive affect, and, although it includes positive and negative emotive responses, it also is meant to cover triggering of values and ‘emotion-neutral’ responses to technique and artifice.<sup>2</sup> Affect can be triggered by content, by form, or by both at once. The full story of how affect is assigned is far beyond the scope of this paper and indeed of semantic interpretation in general; it is a psychological question which relates to general mechanisms of emotional response and (dis)approval which in turn relate to (for the case of literature) the associations speakers have with particular words/concepts, phonetic forms, larger sound structures such as assonance, the ways that words are juxtaposed with each other in the larger form of the work, and so on.

Still, there is more to be said about where affect comes from. Recent work on affect in pragmatics that makes use of affect mostly puts this question aside. For example, Burnett (2020), who uses affect as part of an analysis of slurs, simply assumes that affect can be assigned in a way consonant with the speaker’s social persona; Henderson and McCready (2024), who also work to relate affect with social personas, go further by assuming that affect is partly determined on the basis of homophily, the degree to which the interpreter’s political ideology is aligned with the content she’s hearing. This idea will be incorporated into the analysis to follow here, but it’s too specific to politically related content to serve as an overall mechanism for affect assignment. Another approach can be found in McCready (2012), who takes affect in interpretation of underspecified expressives to be tied to knowledge about the speaker, the hearer, and to normative assumptions about

---

<sup>2</sup>A more general version would also consider changes to how the hearer assigns value and experiences emotion corresponding to the effects of interacting with a work of literature. We can understand these as changes in the affect-assigning function *Aes*, or in the functions comprising it (which will be detailed shortly), which could be thought of in something like the way that van Benthem and Liu (2007) analyze changes in preferences over update sequences. I will leave this set of issues for future work.

speech and event scripts. All these are useful in making sense of how affect works. But what more specifically is the source of aesthetic affect?

The question is related to but distinct from the ways in which affect has been explored in aesthetics as studied in philosophy or in the theory of art. Philosophical aesthetics is quite various, but is concerned mostly with the nature of art and aesthetic qualities. The theory of art and art criticism is also various but is most often concerned with the use and employment of aesthetic response. For instance, to consider a few relatively recent theories, relational aesthetics asks how collective aesthetic experience relates to community formation (Bourriaud, 1998), and investigative aesthetics considers how to recruit aesthetic responses to projects of social justice (Fuller and Weizman, 2021). In the current work, however, we are interested in the sources of response (affect) to a work as they relate to questions of semantic interpretation and how it results to the generation and evaluation of semantic content rather than in ontological or ‘procedural’ questions.

To this end, I propose that there are three basic sources of affect, which I will call somantic, ideological and formal. Somatic affect is mediated by the body and by physical, embodied experience. Ideological affect is mediated by social identities and political and ideological beliefs. Finally, formal affect is mediated by aspects of the form of a work and the interpreter’s formal preferences. Let us consider each of these in a bit more detail.

Somatic affect is mediated by the body. In some sense the body is ultimately the primary mover of affect, in that all of our knowledge and emotion is mediated by the body: there is no knowledge that we can acquire that we are not embodied at the time of acquiring it (as pointed out in a rather different context by Merleau-Ponty 1962). All our experience is embodied, which means physical sensation correlates with our evaluative experiences and our memories of emotionally and otherwise affecting events, as do whatever surrounded us at those moments: physical objects, landscapes, and so forth, to the extent that we can recall them, at a conscious or subconscious level. This view has several consequences. First is that predicates describing physical things frequently trigger affect because of the embodied experiences associated with them, though we might or might not understand why a particular predicate is affecting for us, because we might not consciously remember the relevant experience. A second consequence is that descriptions of embodied experience can trigger affect in sometimes unpredictable ways. We don’t always know why particular predicates affect us in certain ways, and much less can we project how they might affect our interlocutors, because we don’t know all aspects of the embodied experiences of the people we are speaking with or who might read the text we are writing. This observation has consequences

for how one might choose to write and to communicate which I will return to below.<sup>3</sup>

Ideological affect is induced by ideological and political beliefs, social groupings and the interpreter's stereotypes about and attitude toward them, and their judgements about good and bad actions and groups. Ideological affect is assigned on the basis of how social aspects of text affect the interpreter. It can arise on the basis of events described in the text, i.e. its at-issue content in some sense, as with literary work describing the experiences of refugees rejected at a border or put into camps, or painful experiences of misogyny; from implicitly expressed political values, framings or perspectives in the sense of Camp (2013), or, most directly, by terms carrying social meaning such as slurs, which themselves are lexical expressions of certain kind of ideologies on at least some accounts (Burnett, 2020; Davis and McCready, 2020). This kind of affect can be formally understood using personas in the sense of Third Wave sociolinguistics (Eckert, 2008, 2012) as formalized by Burnett (2018) and further explicated by Henderson and McCready (2024). On this kind of view, certain personas are related to ideologies, which have the function of indicating some aspects of the persona-holder's beliefs and the way they assign value to actions, individuals and relevant groups. I will return to the formal details below.

The final source of affect represented in the model is formal affect, which relates to formal properties of the work. Formal properties as conceived of here include all aspects of literary technique and form (compare the notion of poetic articulation in Lepore 2009).<sup>4</sup> These techniques could lie in sonic/phonological aspects of the work, such as assonance, the use of sound representation, and so on; in local structural aspects such as syntactic structures and the type of phrasing used, complex or simple; in global aspects of the text such as its overall structure and form, as in Georges Perec's *Life: a User's Manual*, in which the novel concerns the lives of the people living in a particular apartment house and is structured following its floor plan; ways in which structural/formal expectations coming from genre or other sources are satisfied or betrayed; and even the interpreter's prefer-

---

<sup>3</sup>It might be that it's this kind of somatic affect and embodied experience that is at work in cases of Fregean color such as the differences to be found between 'hound/dog', 'stone/rock', and even 'couch/sofa' (e.g. Neale 1999). I will not pursue this possibility further here.

<sup>4</sup>There is nothing foundational in the theory I will propose that limits it entirely to language-based art. As long as the work to be interpreted has some kind of separable compositional structure at an interpretational level, the theory should in principle be applicable. The issue then becomes whether or not such a structure is recoverable for (for instance) visual art or music. I will return to these questions at the end of the paper.

ences in genre and style (e.g. ‘popular’ fiction vs avant-garde literature). Any of these elements of style and technique can give rise to affective responses which form part of the reader’s response to the text.

What can we say about these three sources of affect and their linguistic realizations? Several immediate conclusions are apparent, with respect to each of the three sources.

First off, somatic aspect relates closely to at-issue content.<sup>5</sup> Our embodied experience relates to things in the world and the state our body was in when we encountered them; these are precisely the things that are described in the at-issue domain (barring derived cases like the conventional implicatures denoted by appositives: Potts 2005) using ‘vanilla’ predications and discourse-linked sequences of e.g. narrative (Asher and Lascarides, 2003). We thus can expect that somatic affect is normally triggered by individual terms, and (the result of semantic composition of) sequences of terms, with denotations in the at-issue domain.

Formal affect differs from somatic affect in that it’s mostly triggered by (as the name suggests) the form of the text and the choices made in it. It also differs in that it’s less individual lexical choices and more longer sequences of text (or even of hierarchial structure) that induce it. The reason is that formal play with textual elements requires patterns to be set up within the text in a way not often amenable to triggering by single elements. It also differs in that it is only sometimes triggered by content, but often by form: this is one reason to take the inputs to the affect-assigning functions that will comprise the core of the semantics discussed below to consist of both linguistic elements and their denotations rather than denotations alone. The text itself is required as the trigger of formal affect.

Ideological affect has properties very different from both of the above. It differs from somatic affect in that it’s very often triggered by not-at-issue content. The examples raised above involved slurs, dogwhistles, and sociolinguistic indices, all of which exemplify the kind of insinuating content discussed by Camp (2018), which is introduced by ‘back-door’ means that are difficult to challenge, ie. via not-at-issue content. Ideologically affective content is however often also introduced by at-issue description and narrative: it’s easy to find descriptive content that resonates with ideologies of various kinds, as seen in recent years in media strategies by people inhabiting mostly ideological positions on the right.

Ideological affect is also the most predictable across interpreters of the three

---

<sup>5</sup>This despite footnote 3, where I suggested that somatic affect might relate to the analysis of Fregean color; roughly, Fregean colors would correspond to distinct somatic affects and embodied experiences with respect to instances of the relevant predicate.



types. One can reasonably expect that agents will assign positive and negative affect to triggering content depending on how it aligns with their own beliefs and value systems, ie. with their own ideologies. Henderson and McCready (2024) used the notion of homophily to assign affective value to certain sociolinguistic personas associated with ideological content; the idea was that positive affect is assigned to the degree that ideologies align, in a similar way to the manner in which social media algorithms filter content for their uses (Chun, 2021).<sup>6</sup> We can also adopt this idea here; doing so means that some aspects at least of ideological affect can't be assigned without the interpreter recognizing the ideology expressed by the author, or the agents of the narrative, or the situations described and how it relates to their own. But without relation to hearer ideologies, whether by content or by author presentation, no affect will be triggered in any case. This relative predictability is not present in the other two types, for aesthetic tastes (formal affect) and backgrounds (somatic affect) will differ extremely and radically across the set of all possible readers of a text.

Let us turn now to formal implementation. Recall that aesthetic interpretation is based on the recovery of affect. Affect is recovered in the formal model of the interpretative process via three functions  $\mathcal{S}, \mathcal{I}, \mathcal{F}$  corresponding to the three sources of affect discussed above (details of which will be provided momentarily); the result of applying each of these functions to a term  $Term$  is assembled into a tuple, which is identified with the aesthetic interpretation of the term:

$$(3) \quad Aes(Term) = \langle \mathcal{S}(Term), \mathcal{I}(Term), \mathcal{F}(Term) \rangle$$

Each individual function assigns affective values to textual elements, where affect again is broadly conceived as emotive or aesthetic responses to linguistic objects and modeled using real numbers in the range  $[-1, 1]$ . Here -1 indicates a very strong negative response and 1 a very strong positive one, with intermediate values interpreted in the obvious manner.

The inputs to these functions are pairs of expressions of natural language  $E$  and their interpretations  $[[E]]$ . The reason for using pairs is that aesthetic reactions can arise on the basis of the form of the expression as well as its meaning especially in the context of formal affect, so it's necessary to take the expressions themselves into account in addition to their interpretations.<sup>7</sup>

---

<sup>6</sup>Note also the relationship to the results of Farrell (1993) on credibility in cheap-talk games, signaling games where senders are credible to precisely the degree to which their interests align with those of receivers.

<sup>7</sup>Note that using phonetic forms directly in the input to each affect-assigning function opens

To generalize this idea to the interpretation of texts, we first must introduce a method for applying the functions which yield the various affective characters of terms to larger textual units. I do this by applying the functions to each subpart of a text:

$$(4) \quad \mathcal{S}(\langle E, [[E]] \rangle) = [\langle \langle E', [[E']] \rangle, n \rangle | E' \sqsubseteq E \wedge n = \mathcal{A}_S(\langle E', [[E']] \rangle)]$$

Here, ‘ $\sqsubseteq$ ’ is a relation holding between two elements when one is a subexpression of the other (McCready, 2012).  $\mathcal{A}_S$  is a function which yields real numbers in the interval  $[0,1]$  when applied to an expression, so  $\mathcal{A}_S =_{df} f : \langle E, [[E]] \rangle \mapsto [0, 1]$ ; this function retrieves the degree of somatic affect of each textual unit.  $\mathcal{S}$  thus yields real-numbered outputs indicating degree of somatic affect, whether positive or negative, paired with each subexpression of the text. Thus, the function  $\mathcal{S}(\langle E, [[E]] \rangle)$  yields a multiset corresponding to the interpretation of each subunit of the text it applies to, from largest to smallest.<sup>8</sup> I use multisets because a given expression can appear many times over a text, and needs to be treated as introducing affect on each appearance.  $\mathcal{I}, \mathcal{F}$  (for ideological and formal affect respectively) can be constructed similarly, except of course for use of  $\mathcal{A}_I / \mathcal{A}_F$  corresponding to the different sources of affect.

The output of  $\mathcal{S}, \mathcal{I}, \mathcal{F}$  are unified as follows under the general rubric *Aes*(thetic response), for  $T = \langle Text, [[Text]] \rangle$ .

$$(5) \quad Aes(T) = \langle \mathcal{S}(T), \mathcal{I}(T), \mathcal{F}(T) \rangle$$

This basic representation is information-rich, which is useful in that it allows us to recover different ways in which a text is affecting due to the possibility of examining the distinct elements of the tuple comprising  $Aes(T)$ ; that possibility also makes available analyses which distinguish types of texts via their sources of affect, in ways corresponding to genre or even textual purpose. For instance, a highly ideologically affecting text which has little other affective quality could be classified as a piece of propaganda, and a formally affecting text with little other emotive resonance might fall into the traditions of certain formally oriented

---

the door to interpretative agents having strong reactions to particular lexical choices. This opens the possibility of using the proposed system to make sense of theories of slurs where reactions are mediated by specific lexical expressions rather than meanings (e.g. Anderson and Lepore 2013; Davis and McCready 2020). Doing so also makes sense given that ideological affect also plays a role in aesthetic interpretation itself.

<sup>8</sup>Recall that the square-bracket notation above indicates a multiset. Multisets are like standard sets except that distinct instances of identical content have a separate existence in them:  $\{a, a\} = \{a\}$ , but  $[a, a] \neq [a]$ .

literary avant-gardes. Further, particular critical traditions might make reference to these sources as well; ecocriticism, for example, has been characterized as having ‘one foot in literature, the other on land’ by Glotfelty (1996), which in this context can be parsed as a critical stance that makes reference to specific aspects of somatic experience.

We would like to have a method of determining the affective value of a text in a way that makes it comparable to others, and which makes it available for applications involving general considerations about affective value and significance, to be discussed below. This can be done by extracting total values from the text. As things stand, we have multisets corresponding to how each textual unit corresponds to affect generation, for each source of affect. These can be unified as follows (here,  $\Pi_i$  is the projection function picking out the  $i$ th element of the tuple):

$$(6) \quad \mathcal{S}_{Total}(T) = \frac{\sum_{x \in \mathcal{S}(T)} |\Pi_2(x)|}{Card(\mathcal{S}(T))}$$

The total somatic affect yielded by  $\mathcal{S}_{Total}$  is then the sum of the absolute values of all textual units of  $T$  divided by the total number of such textual units. The use of absolute value is important here: we would not want to say that a positive and negative reaction toward two parts of the same text cancel each other out, but rather that they both play into a more intense affect (compare Neill 1998). Defining similar functions for ideological and formal affect yields the following as the total affect of a text  $T$ :

$$(7) \quad Aes(T) = \langle \mathcal{S}_{Total}(T), \mathcal{I}_{Total}(T), \mathcal{F}_{Total}(T) \rangle$$

$Aes(T)$  can in turn be transformed into a single value by averaging over the three elements:

$$(8) \quad Aes_S(T) = \frac{\mathcal{S}_{Total}(T) + \mathcal{I}_{Total}(T) + \mathcal{F}_{Total}(T)}{3}$$

Each of these representations has its own applications, as we will see below. Before turning to applications, though, I want to briefly consider how the static approach introduced here might relate to the dynamic processing of literary text.

There are two ways in which we can think of the aesthetic value of a text in terms of generalized affect. The first is as a kind of quality possessed by the text; in a way, this is close to traditional, static conceptions of semantic meaning, where the meaning of a text corresponds to its truth conditions. In this sense, the affect produced by a text corresponds to a kind of abstract valuation of aesthetic content.

Alternatively, we can ask about the effect on a reader/interpreter of processing the text: what happens to a reader as they go through the text?<sup>9</sup> How would they currently evaluate it? This is a dynamic picture corresponding roughly to the kind of update-based semantics employed by for instance Groenendijk and Stokhof (1991), among very many examples. Just as with static and dynamic semantic theories, the conceptual nature and empirical application of the two is distinct, and it is worth exploring both. Above I presented a static theory which results in a record of the aesthetic content of texts. How can this static theory form the foundation of a more dynamic picture?<sup>10</sup>

Unfortunately, the project is not entirely straightforward. Classical dynamic semantics proceeds by update on propositions, which may or may not be further articulated (Groenendijk and Stokhof, 1991; Veltman, 1996); in more recent versions, update proceeds incrementally by update with single predicates and discourse referents (Bittner, 2014; Murray, 2015). What these frameworks have in common is that update is monotonic, in that previously introduced information is added to rather than destroyed. There is of course also a large literature on destructive update (often under the rubric of belief revision: Gärdenfors 1988), but it hasn't been particularly influential in linguistics outside the analysis of a small set of phenomena focused on discourse particles (McCready, 2008; Davis, 2011). But in a way this is just what we get with aesthetic update: not a destructive update exactly, for information is not destroyed, but one in which the value of what comes now is not monotonically correlated with the value of what came before.

Consider the aesthetic value of a segment of text. On the current theory, it can be viewed in three ways: as a number in  $[0,1]$ , or as a sequence of such numbers, or, finally, as a sequence of multisets corresponding to all the textual elements together with the affects of various kinds they yield. What happens when we add a new segment? The last of these is updated in a clear manner: we must add the new segment with its value to each multiset, and also all the new segments in which it participates together with their values. This is straightforwardly implemented. The others, however, are not, at least directly: they can be derived from the new multisets in the manner described above, but there is no simple technique which

---

<sup>9</sup>Thanks to Paloma Jeretic for pressing me to more deeply consider dynamic aspects of aesthetic interpretation.

<sup>10</sup>One could ask whether any dynamic theory of aesthetic interpretation would actually be pragmatic rather than semantic. This question closely relates to similar questions addressed to dynamic theories of meaning, for instance in debates around the meaning of conjunctions like *and*; I will not enter deeply into this discussion here, instead only pointing the reader to the more specifically pragmatically oriented discussion in §3.

allows them to be directly updated. The reason is that adding a new segment might cause the value assigned to sequences in which it participates to fluctuate wildly.

For an example, imagine a Shakespearean sonnet in which a random word is substituted by ‘fart’.<sup>11</sup> The presumably high aesthetic quality of the text up to that point will drop off quite suddenly when this new word is processed (for most readers, at least). Conversely, consider a case in which a new word suddenly makes clear a complex structure in the text, for example the presence of a hitherto invisible metaphor or symbolic structure. Once that is processed, the aesthetic value of the whole increases dramatically. One could compare this situation to one in which the discourse structure of a text suddenly comes clear, and the connections between seemingly unconnected statements become transparent; this is something that of course happens in literature, but also in more standardly informative texts, where inference of discourse structure is usually farmed out to a process independent of dynamic update, as in Asher and Lascarides (2003), where it happens in a separate logical space. The upshot is that a simple dynamics – for example, one based on a standardized operation on the values of the new element and the previous context, as defined by McCready (2019) for the case of honorifics – won’t do.

I therefore will not attempt the project of providing a direct system of dynamic update of aesthetic value. The dynamic update of the multiset construction produced by the combination of (4) and (5) is nothing more than the reapplication of these two constructions (here, ‘ $A \bowtie B$ ’ indicates the concatenation of  $A$  and  $B$ ).

$$(9) \quad T[Wd]_{Aes} = Aes(T \bowtie Wd)$$

This is entirely expected on the current theory. It is possible to complexify the situation by considering decay of previous affect over time, so that the more distant a particular part of the text the less the affect that remains active from that textual segment, or, similarly, by magnifying the effects of more recent text by some method (e.g. multiplication by a constant); I will not further consider these attentional effects here, instead moving to other aspects of aesthetic interpretation which involve applications of the theory developed so far.

The first application involves the significance to the interpreter of a literary text. A text can be viewed as aesthetically affecting if the affective response that it induces exceeds a certain threshold, just as on theories analyzing gradable adjectives using the (often-covert) operator *pos*, which introduces such a standard (Kennedy, 2007), or the theory of discourse-level reliability and sincerity argued

---

<sup>11</sup>This example is due to Daniel Harris (p.c.).

for in McCready (2015). Such thresholds can be set in various ways corresponding to ways of thinking of (intensity of) aesthetic experience. This significance can be measured on individual values  $\Pi_1(Aes(T))$ ,  $\Pi_2(Aes(T))$  or  $\Pi_3(Aes(T))$  (where  $\Pi_i$  is again a projection function picking out the  $i$ th element of the tuple) or across the whole sequence of values; this gives the possibility of measuring significance for particular kinds of affect, or globally, or in ways determined by text genre.

$$(10) \quad Aes_S(T) \geq t, \text{ for } t \text{ an affective threshold}$$

Alternatively, one can take the full text to be the vehicle for significance, as in (11), where significance is defined on the maximally-sized textual unit, ie. the entire work. Which kind of definition is optimal is somewhat application-oriented, but in general the maximal-size method should be preferred; consider, for an example (overheard by the author in eavesdropping on a Twitter conversation), a wonderful novel which can, however, be recommended only if the reader ignores the epilogue, which is tacky and destroys the whole effect of the rest of the work. This means that the whole work has to be taken into account when assigning value and significance. The Shakespearean sonnet substitution described above is another kind of example in which the maximal unit seems optimal for determining significance, differing from the novel case in that the affect-destroyer is medial rather than final in the text.<sup>12</sup>

$$(11) \quad Aes_S(\max(\langle T, [[T]] \rangle)) \geq t$$

Let us briefly look at one example of the application of this theory before moving on. For reasons of tractability, I'll use a very short text: the internet-famous 'best extremely short story in the world' (according to some sources on Google, at least), supposedly though probably not actually authored by Hemingway:

$$(12) \quad \text{For sale: baby shoes, never worn.}$$

How can affect be assigned to this brief but evocative story?

Consider two interpreters. The first interpreter is the ideal audience for this particular story. They are a parent of a young child, and a child who was an 'easy

---

<sup>12</sup>I have only considered cases where texts trigger existing affective structures. But sometimes encountering a text changes affective response, enabling new emotive triggers or alterations in the values assigned by *Aes* to the input. Such responses correspond to changes in the values assigned by *Aes* to its inputs, and can be modeled in ways similar to changes in utility functions within decision and game theory (with similar complications, such as the problem of decision-making under uncertainty about future utility assignments discussed by Paul (2014).

baby’ that was quickly and easily delivered, slept easily and deeply, and had no serious illnesses. Their relationship with their partner is good and they have a happy family life. They love babies and their life is lived in an unquestioned normative pattern where child-raising is centered. Further, they are a person whose emotional buttons are easily pushed. On top of all that, they love to read but have very little time, so short shorts are their favorite genre. As a result, this story is very affecting for them: it has many somatic resonances, down to the individual word ‘baby’, which is affecting for ideological reasons as well; the phrase ‘baby shoes’ also induces somatic affect, because they also remember their child’s first pair of shoes. Even formal affect is induced because of their preference for short-shorts. This story is thus highly significant for them, as  $Aes_S(T) >_! t$ .

Now for a second interpreter. This person is hard-left and supports abolishment of the nuclear family, and further is antinatalist. As a result the terms ‘baby’ and ‘baby shoes’ also deliver a strong affect for them, but in this case it’s a negative one; further, because of this interpreter’s anticapitalist stance, the phrase ‘for sale’ also delivers a highly negative ideological affect. They also hate easy emotional appeals, especially ones grounded in cisheteronormativity. Consequently, they have a strong negative response to this aspect of the text as well, which I will take to have its source in formal affect as a content preference with respect to literary and artistic work. For this interpreter, too,  $Aes_S(T) >_! t$ , but this time because of a strong negative response rather than the strongly positive one of the previous interpreter. We thus see the utility both of having access to a record of the different kinds of affect induced by the text – via  $Aes(T)$  – and the absolute values of that affect, as used in  $Aes_S(T)$  to compute the text’s significance for the interpreter.

The next section will turn to a consideration of what interpreter’s responses are relevant for determining the value of the text. First, though, I consider a few more abstract points, one involving distinctions between different textual types and one considering the prospect of completely replacing truth-correspondence with affect, before turning to questions of perspective and interpretation.

The reader may have the impression that this way of assigning significance is a bit of a blunt instrument. Should we distinguish between affective responses which have relevance to aesthetic judgements and those tracking eg. political solidarity? The answer depends on our other assumptions. Allowing e.g. political propaganda as aesthetic communication makes sense on a universalist view of literary text, on which any text can be read in a literary mode; I myself am sympathetic to this view (which I will say more about in the conclusion), and see no special need to strictly separate ‘literary’ texts and interpretation from

information-conveying texts and truth-conditions. Still it remains possible to distinguish aesthetic and other responses in a principled way. We need only return to our ‘official’ representations of aesthetic content, where it is modeled as a 3-tuple of multisets corresponding to aesthetic value derived from the three sources of affect. Thus, a text inducing aspect only from ideological sources might be categorizable as propaganda, while one which only induces formally sourced aspect might be taken to stem from certain kinds of literary avant-gardes which are concerned with formal aspects of literature to the exclusion of appeals to other kinds of affect.

Can aesthetic interpretation replace truth-conditions entirely? The two have completely different functions and it seems pointless to even make the attempt.<sup>13</sup> Formally speaking, in any case, *Aes* operates independently from ordinary truth-correspondence, which can proceed as usual in tandem with the style of interpretation modeled by *Aes* (and indeed, in most cases, must, given that *Aes* takes the result of compositional semantics as its input in the case of textual units larger than a single word). This fact means that a communication can be taken both as an aesthetic and truth-oriented. Indeed, we often find cases where texts read as fiction also have the capacity to teach us facts about the world; the attentive reader of Sherlock Holmes likely learns something about the geography and social mores of London of the era, for example. Still, it is probably usual to take a particular (sequence of) utterance(s) to be intended either primarily aesthetically or as information-conveying; but dual interpretation is certainly commonly used for such genres as poetic nonfiction, and in fact for any kind of prose when one starts to consider its ‘quality’ as opposed to just its information content. Thus, the line between aesthetic text and non-aesthetic text is extremely blurry, as reflected in possibility of applying the two interpretative styles together.

## 2.2 Aesthetic perspective

Let us now turn to a question I have put aside so far. The theory as stated relies on a set of functions that yield affect of various kinds when applied to parts of a text when interpreted in an aesthetic mode. But there is no sense in which the same affect is yielded for every interpreter. We saw this above already in the baby shoes example, where two possible interpreters with different experiential backgrounds

---

<sup>13</sup>One could perhaps try to derive truth from affect; maybe true statements are more affecting than false ones, on some definition of ‘affecting,’ just as they might be thought to be more relevant, on some definition of relevance. I will not pursue this project here as it seems to me quite unnecessary.



and political and moral views had quite different reactions to the work, as reflected in the different affects it yielded for them. This section considers this issue and concludes that affect assignment is agent-dependent, meaning that we are actually dealing with a family of functions, not a single one. As we'll see, this opens up other interpretative and literary modes as well, from the perspectives of both writer and reader; commonalities also appear with other phenomena in semantics, notably subjective predicates and personal taste.

Above we saw how two different interpreters derived affect from a brief story probably not authored by Hemingway. Their ways of being affected were very different because of their differing backgrounds and beliefs, but both found the story affecting, whether positively or negatively. But the case raises a more general question: whose affect is at issue here? In the theory as presented so far,  $Aes_S(T)$  yields a real number depending on whether the text  $T$  induces affect. But for who? In the case above: for one of two different interpreters. But is it always the interpreter? Could it also be the speaker, otherwise known in this context as 'the author'? This kind of question is already raised by Dénigot (2022) in his attempt to use game theory to provide interpretations of literature, though with the goal of determining how an interpreter is to recover the author's intended interpretation. The question is, of course, also not unique to literary interpretation. It has parallels across many other domains of semantics/pragmatics and philosophy of language: with subjective predicates (Lasnik, 2005), with modals (Weatherson and Egan, 2011), even in the comparison of individual information states with the common ground.

How should we make sense of a statement aimed at conveying truth-conditional content when conversational agents disagree about whether it is true? This can be viewed as a problem of imperfect information, where the agents aren't equal in their access to the facts. Research in linguistics and philosophy treats this in terms of information states, which are private, and common ground, which is both public and shared; there is a complex literature on how information states relate to common ground, and how (and if) common grounds can be constructed from private information (see Stalnaker 1979; Muskens et al. 1997; Farkas and Bruce 2010, among many others). The fundamental problem is that judgment about truth is subjective in a way that truth is not. Given this, it is possible to talk about sincerity rather than truth, or objective truth vs subjective falsehood. Each of these strategies have analogues in the current domain; sincerity will be explored in the next section, and subjectivity will, in a sense, be the core of the agent relativity-based proposal I will make now.

One style of traditional literary interpretation takes the goal to be understand-

ing what the author meant to communicate.<sup>14</sup> This is of course a powerful and interesting way of thinking about literary texts, but it is far from the only one, especially in work on literature subsequent to Barthes (1977). On this ‘death of the author’ style of interpretation, the reader need not try to figure out what the author had in mind, but instead can construct their own interpretation, or leave the text without any particular symbolic interpretation at all. Of course, this work is about how one assigns meaning to a text in the sense of symbolism and theme (an issue I will return to at the end of the paper), rather than the object-level interpretation we are concerned with in semantics and pragmatics, whether in standard truth-conditional interpretation or the sort of aesthetic interpretation proposed here. But the main question is similar: whose intentions, and whose reactions, are at issue when we think about a text? Following Barthes, it can be the author, the reader of the text, or, really, anyone at all.

I think the solution is just to relativize interpretation completely. For aesthetic interpretation, what the interpreter can extract from the text matters more than what the author intended, because significance is determined on the basis of experienced affect rather than recovery of authorial intention. In this sense: the problem of literary interpretation isn’t a problem of intention recovery, but instead is a problem of how agents find meaning in the text, in line with the ideas of Barthes (1977). Formally, this view can be realized by taking *Aes* to be relativized to agents, and each *Aes<sup>a</sup>* to be a member of a larger family of agent-indexed affective functions *Aes<sup>u</sup>*. None of these is privileged above the others in a deep sense, though authorial intent, and the significance a text has or had to the person that wrote it, can be derived by setting *a* to the author. This will be important in §3, where it will be used to define a notion of sincere creation.

But this doesn’t mean we should put questions of interpretation aside. A key question for a semantics of literary texts is how the expressions of a text should be interpreted in literary contexts. As discussed, authorial-death style interpretation is made available by the use of agent-relative affect. But, further, some disambiguations or ways of resolving underspecified content yield ‘*Aes<sup>a</sup>*-better’

---

<sup>14</sup>The work of Dénigot (2022), for example, aims to model this way of thinking: the idea is to use one’s information about the author’s likely communicative intentions (in the form of prior probabilities) to determine what vocabulary she is using, and thereby to make sense of what interpretation is intended. More specifically, in this proposal there are multiple lexica in play, and the interpretation of terms in literary contexts is mediated by priors about author personas. The identification of personas in turn induces changes in priors about the lexica in use. Here, the lexicon that should be selected is just the one the author was using, so the result is the recovery of the author’s intention.

interpretations than others in the sense of aesthetic value. These are interpretations which have a higher *Aes*-value for *a* than their alternatives. Arguably, an agent is best served to pick an interpretation that maximizes affect, for doing so will maximize the text's significance. We can define a notion of strength relevant for this purpose as follows.

- (13) *(STR)*  $Aes^a(Int) \leq Aes^a(Int') \rightarrow Int \preceq_a Int'$ , where *Int* and *Int'* are pairs  $\langle E, [[E']] \rangle$  and  $\langle E, [[E'']] \rangle$  such that  $[[E']]$ ,  $[[E'']]$  are disambiguations of *E*.

Given this principle, an affect-maximizing agent will select *Int'* over *Int* given the option.

Let's look at an example of how an affect-maximizing agent might behave. The following text is a constructed example, which is both rather artificial and without much if any literary merit. Still, it will suffice to show how the interpretative strategy works. The key ambiguity is the resolution of the pronoun *it*, which has two possible text-internal antecedents: *the moon* and *the sky*.<sup>15</sup>

The moon<sub>1</sub> was big in the sky<sub>2</sub> above her. It<sub>{1,2}</sub> shone blue-white against the bare tree branches.

Let's assume a first interpreter whose only relevant affect induced by this text is somatic: when they were a teenager, they had a peak experience of some sort (whether positive or negative) on a moonlit night. The text lacks, for them, relevant ideological and formal affective value. For this interpreter, it is optimal to resolve  $it_{\{1,2\}} = it_1$ ; doing so yields higher *Aes*-value due to the somatic affect induced by the embodied experience of the agent.

Consider now a second interpreter who has no relevant experience with moonlit nights, no ideological stakes in this content, but has a strong genre preference for science fiction and fantasy. For this interpreter, it's going to be optimal to resolve  $it_{\{1,2\}} = it_2$ , because doing so yields a higher aesthetic value; if the sky is shining blue-white, we likely read the setting as an unearthly one, and consequently the genre of the text as fantastic or science-fictional.

The reader might wonder whether this is something that readers of literature actually do. The formalism makes available the possibility of reading for affect maximization: but is this an empirical result, or is it something that's only an unrealized possibility, and possibly even a suspect aspect of the theory? As it turns

---

<sup>15</sup>I put aside for the purposes of this toy example questions of prominence, salience and how this pronoun might most naturally be resolved in an affect-free vacuum.

out, this style of interpretation – or a variant thereof – is already discussed in the theory of aesthetics as the value-maximizing strategy for literary interpretation advocated by Davies (2006). Though this strategy doesn't directly follow from the theory, which is perfectly compatible with other metrics for determining resolutions of underspecified content, the theory does make it available, again showing its flexibility, for better or worse.

It seems to me that the empirical question of whether a specific interpretative strategy is used is not actually a sensible one for literary texts in particular. Choosing to resolve ambiguities and underspecifications to maximize the interest of the interpretation – by whatever metric – is not something we do for information-conveying texts. In truth-conducive texts, the goal is to recover the intended content of the speaker, as modeled directly via signaling games, in which sender and receiver only receive positive payoffs only if the receiver can pick the actual state of the world based on the signal (Lewis, 1969; van Rooij, 2008). Given this, it is completely irrational for a receiver to try to make sense of the sender's signal in a way that potentially pushes them away from recovering the actual state. Convention, therefore, is highly important, both in terms of conventions of meaning and conventions of more pragmatic interpretation.

But for aesthetic interpretation, this anticonventional or even antinormative strategy is less obviously wildly questionable. The goal in reading literary texts is not necessarily to recover the author's intended content, perhaps because, unlike in truth-conducive communication, there's not an obvious sense in which the author has privileged access to information that the reader doesn't.<sup>16</sup> Instead, it might be an optimal strategy for the reader to maximize what they get out of a text.

The availability of this strategy also opens up possibilities for literary writing. Consider a technique in which the author chooses to write to this interpretative strategy. One would go about doing so by maximizing chances for the reader to maximize the affect they derive from the text via underspecification resolution; one would do that by maximizing the quantity and variety of ambiguous and underspecified elements in the text. Some current experimental literary writing has something resembling this character, though it's not clear whether this writing is consciously responsive to the reading strategy following from (STR) as stated in

---

<sup>16</sup>One way to make sense of this observation is that classical signaling game models aren't the right way to model literary interpretation either. Perhaps one can get closer by looking to modifications of such games for social meaning like those made by Burnett (2018) or Henderson and McCready (2024), where the sender sends a signal aimed at maximizing the likelihood that the receiver recovers the social persona the sender aims to convey, as opposed to some external state of the world. I leave further exploration of this idea for another occasion.

(13). Still, this aesthetic space (or at least its observation) is opened by aesthetic interpretation. Is the question of whether people actually employ it an empirical question, or a question about what kind of artistic activity writer and reader are engaged in? What indeed is artistic communication, and what are its possibilities? These questions don't seem to be, precisely, empirical ones of the kind usually addressed by linguistic theory. In this sense, the theory is circumscribing a space of possibility, only some of which might be currently realized in artistic practice.

### **2.3 Semantics: summary**

In this section, I proposed a theory of the semantics of texts interpreted aesthetically, founded on the notion of affect rather than truth. The leading idea is that embodied experience, ideological belief, and formal preference lead to three kinds of affective response to linguistic elements, which are extracted from individual elements on up to the full text and collected into tuples from which single numerical values can be computed and used in calculations of text significance and other things. All this was argued to be agent-relative, as formally realized by the use of a family of aesthetic interpretation functions indexed to individual agents. I gave several examples of the application of the theory, and indicated several additional ways in which it can be used: in the understanding of genre and preference, and in the interpretation and production of novel kinds of literary texts. But all this has been about semantics. What about the pragmatics of aesthetic/literary texts? What does a pragmatics built on affect look like? This is the topic of the next section of the paper.

## **3 Aesthetic pragmatics: trust and sincerity**

We have now got a semantics for aesthetic interpretation of texts which is grounded in affect. But what about pragmatics? Classical approaches to pragmatic phenomena assume notions of truth and informativeness, where the latter is itself grounded in the communication of true and sufficient content. This kind of pragmatics can, of course, work at the same time as an 'aesthetic pragmatics,' just as standard truth-conditional interpretation can operate in tandem with aesthetic interpretation, as discussed in the last section. But what, fundamentally, does pragmatics look like for aesthetic communication?

I cannot give a full answer to this question here; doing so would likely require a monograph-length work. In part, the reason is purely practical: pragmatics for

many years functioned as a ‘wastebasket’ for semantics (Bar-Hillel, 1971), into which everyone shoveled the phenomena that either were too messy or that they otherwise weren’t quite sure how to deal with. This means that pragmatics is already an extremely heterogeneous empirical domain, and one that requires quite heterogeneous theoretical moves. Theories of presupposition, for example, don’t look much like theories of conversational implicature, though both are theories of aspects of the pragmatics of truth-conditional interpretation.

To take one example, for the case of aesthetic communication, it’s not even entirely clear that presupposition as such is a sensible notion: presupposition is about what information is assumed in conversation, but there doesn’t seem to be anything completely similar in the domain of affect. The closest phenomenon that I see now is one in which the significance of a particular text depends on a particular experience, such as a certain embodied experience, or having engaged with some other text. A particular event or linguistic usage in a novel might not make aesthetic sense without familiarity to what came before it in the text, or even familiarity with a whole genre of writing, or without having a specific sort of life experience.

This kind of view links presupposition in the aesthetic domain to the possibility of extracting significance, and therefore differs from what we see in presupposition in the truth-conditional domain on at least many theories of presupposition, where interpretation and evaluation aren’t even possible (in the sense of assignment of a truth-value) without the presupposition being satisfied; here, instead, it’s just that a threshold of significance won’t be reached without the requisite background. This might be thought of as corresponding to a truth-value judgement, though the parallels are not at all exact, for several reasons. For example, this comparison has the consequence that every expression of an aesthetically interpreted text is at least potentially an ‘aesthetic presupposition trigger,’ for any expression could be associated with requirements of this sort; it also means that triggering presuppositions is no longer a property of expressions or constructions, but of more global aspects of texts or interpreter affective backgrounds. I am not sure at present whether these observations mean that the comparison is simply not working, or if it instead points up a difference in how ‘presupposition’ might work differently across these domains. In any case, I will be picking up a more straightforward test case here: an aspect of Gricean cooperative communication.

I am going to look at a particular aspect of Gricean communication: Gricean Quality, which is meant to underlie truth-conducive communication. The reason for choosing this domain is its generality. Cooperative behavior is found across all domains of human and nonhuman activity, and there are very many theories

of how and why it arises across these domains, for example in economics and in theoretical biology; Gricean cooperation is just one special case. Further, Gricean cooperation as such doesn't have to be understood as limited to the precise domain of truth-conducive communication, but rather in terms of broader views of sincere communicative behavior. I therefore will take this as my initial case for analysis.

The official statement of Gricean Quality is as follows (Grice, 1975):

- Try to make your contribution one that is true.
  1. Do not say what you believe to be false.
  2. Do not say things for which you lack adequate evidence.

The intent here is to require cooperative speakers to speak truly, while acknowledging that people are fallible epistemological agents and may fail to do so while still being cooperative in that they are aiming at truth-conducive statements. The basic imperative, though, is to be truth-tracking in what we say. Can we find an analogue of Quality for aesthetic communication, given that truth isn't involved? In the rest of this section, I will explore one such analogue, which I'll arrive at by looking in the domain of sincerity and trust. Since these notions are more general than truth, they can more easily be extended to aesthetic communication.

How do we judge whether someone is worthy of our trust, whether it be trusting what they say or trusting them in other ways? It has been shown that reputation is key in making this decision; deciding whether to trust based on someone's reputation as to cooperativity has been shown to be an extremely effective method. Much of this research has been carried out in game-theoretic contexts, such as in the repeated Prisoner's Dilemma, where a number of different strategies have been explored (see e.g. Nowak and Sigmund 1998a,b). One way to be judged communicatively trustworthy, then, is to be someone judged reliable with respect to speaking truth, in the sense of having a reputation for truth-tracking communication.

Having access to reputations helps us determine who's worthy of our trust. In the work of Nowak and Sigmund on reputations and cooperation, they look at player reputations in the repeated Prisoner's Dilemma; agents, in this model, have 'reputation indices' which are upgraded and downgraded depending whether an agent plays a cooperative move in a particular game iteration. Given such an index, they show that optimal ways for an agent  $i$  to choose whether or not to play a move that requires a cooperative play by the other agent  $j$  is to examine  $j$ 's reliability score and play a cooperative move if it (for example) exceeds a

particular threshold, or exceeds that of  $i$ , and if not, to play a move that assumes the other agent will not cooperate. Since such strategies are public, the other agent has an incentive to maintain her reputation at a high level, for otherwise (in this setting) her own payoffs will end up low, and she will fail to maximize utility in this game iteration, and potentially also over the course of further repetitions of the game.

This style of analysis has been picked up in the linguistics and philosophy literature for the particular case of Gricean Quality. The kinds of reputations needed to make judgements about truth-conducive communicative reliability have been modeled, again in a repeated game setting, by McCready (2015): on her view, reputations are derived in part from *histories* of repeated games. Game histories are  $n$ -tuples of sequences of records representing the history of the agent's actions at each decision point. For communication: these are choices about what to say together with truth evaluation (drawn from the set  $\{T, F, ?\}$ , where '?' means that the statement can't be verified as true or false, or is otherwise defective, for example because of presupposition failure).

In McCready's (2015) theory, a player's reputation in a game is derived from the history of their actions in that game.<sup>17</sup> A player's reputation with respect to some choice is their propensity, based on past performance, to make a particular move at that point in the game, defined as the proportion of the total number of game repetitions that the player chose that action at that choice point. An agent's propensity to play a strategy is a real number in  $[0, 1]$ , and an agent has a propensity for using strategy iff their propensity to play exceeds a contextually set degree in that range, i.e. the contextual standard for having that propensity (Kennedy, 2007). Note the similarity to the definition of significance for aesthetic texts in the previous section.

This model is defined specifically for truth-conducive communication. What other kinds of sincerity do we find in language?

One such place is social meaning, a domain recently getting substantial atten-

---

<sup>17</sup>Since in initial game iterations there is no game history, more must be going on; McCready takes initial judgements of reliability to be based on properties of the speaker and how those specific properties play into the judgements of the agent making the decision about trust. In some cases, such properties will lead to rational choices about reliability, for instance information about the speaker's profession that's derivable from their self-presentation: it's reasonable to assume someone in the airport with a nametag on might be trustworthy on the question of how to get to passport control. In many cases, though, such judgements will be based on factors like race or gender, which are not relevant to many judgements about reliability in which they nonetheless seem to play a role (Fricker, 2009; McCready and Winterstein, 2019). In McCready's theory, these factors and interactional history together are the basis of judgements about agent reliability.



tion in formal semantics and pragmatics. This term refers to meanings that relate to the social positioning of speakers, in terms of membership in social groups, self-presentations, or ways in which language use can work (or be leveraged) to alter or reinforce structures of power in particular discourse contexts. Social meaning comes in many varieties, including indices of social identity or persona (Eckert, 1989), honorifics (McCready, 2019), gender marking and the use of gendered pronouns, and slurs, on which there is an extensive literature in philosophy of language. Here, I will focus on aspects of social meaning which involve a notion of persona: a social role, stance, or group which the speaker signals via their linguistic choices.

The signaling of personas differs from truth-conducive communication in a way critical for my purposes here. To be truth-conducting, the content of a speech act must match the state of the world, which is known only by the speaker but not by the hearer in ordinary communicative situations. In standard signaling games, this is reflected by letting only the signaler observe the state of the world, and setting payoff structures in a way that only gives positive payoffs for sender and receiver if the correct state is recovered, as already mentioned in the last section. This model isn't appropriate for social meanings, as observed by Burnett (2018); the speaker chooses to present themselves in a certain way depending on how they want to be seen, and accordingly chooses expressions carrying social meaning which convey that impression. Burnett models this by tying payoffs to speaker preferences instead of to external states. But this makes the kind of reliability modeled for instance by McCready (2015) inappropriate for the setting, as there is no sense in which truth-tracking is a conceptually sensible tool here.

There are two possible ways to make sense of what cooperative communication amounts to in this case. Both involve moves from reliability to more general notions of sincerity. The first starts with the observation that at least some personas are, in fact, associated with truth-conditional content, and ties speaker sincerity to belief in that content. The second generalizes the first, and takes sincerity to reduce to the speaker communicating in a way that corresponds to their own preferences about social presentation. Let us begin with the first.

The first method starts with the observation that some, though not all, personas are associated with ideologies. This observation is the key to the work of Henderson and McCready (2024) on dogwhistles, which are analyzed as expressions which signal personas associated with certain kinds of political stances to those listeners whose priors make the connection between message and persona visible to them. Political personas of this type, according to Henderson and McCready, come with ideological content, which serves as a basis for trust in some hearers in

that it allows them to draw conclusions about the likely actions of political agents (Henderson and McCready, 2019). These ideologies are the kinds of thing that can be judged for sincerity or lack thereof.

On the view of Henderson and McCready, ideologies indicate approval or disapproval of various actions, people, or (abstract) objects, in very much the way discussed in the previous section with respect to ideological affect in the deriving of aesthetic content. They also communicate more global assumptions about the world. This latter aspect of ideological content is modeled as what they call epistemic bases  $\mathcal{B}$ , which are sets of propositions: background beliefs associated with the ideology in question. These can be beliefs about what kinds of things are present in the world, stereotypes about particular groups of people, beliefs about political systems or values, conditionals representing beliefs about causal structures, and so on and so forth. The point is that all of these are truth-evaluable, and, as propositions, can serve as objects of belief.

We can now address the question of what personas a particular speaker can sincerely adopt, in the sense of sincerely signaling that persona via a linguistic act. Henderson and McCready define the sincere performance of a persona by a speaker in terms of the relationship between the basis of that persona and the speaker’s beliefs. This is of course an analogue of Gricean Quality for the domain of social meaning, which Henderson and McCready call *Social Sincerity*. Formally, Social Sincerity amounts to requiring the personas compatible with the speaker’s utterance,  $\mathbf{emf}(u)$ , to associate with ideological bases which correlate with the speaker’s beliefs.

- (14) *Social Sincerity* (Henderson and McCready, 2024)  
 $\forall s, u, \pi [\text{utter}(s)(u) \wedge \pi \in \mathbf{emf}(u) \wedge \iota_\pi \rightarrow \text{MOST}(p \in \Pi_2(\iota_\pi))(\text{Bel}(s, p))]$   
 ‘If a speaker utters a sentence compatible with persona  $\pi$ , they believe a significant number of the propositions comprising the basis for  $\pi$ .’

This principle generalizes the notion of Gricean Quality in that it doesn’t limit attention anymore to the truth-conditional content of utterances, but it remains difficult to apply directly to aesthetic communication, simply because it lacks the kind of structure that results from aesthetic interpretation.

Fortunately, however, the requisite structure is readily available. Recall that the crucial mechanism of aesthetic interpretation is a function *Aes* which yields real-numbered values of various kinds for each input term in a way which may differ from agent to agent. But this is rather similar to the structure of utility functions in game theory, which also yield real numbers in a way that’s indexed

to the distinct agents whose actions and payoffs are at issue in the game. Further, for the specific case of social meaning, those payoffs depend on the preferences of the agent with respect to personal presentation rather than on ensuring that the response of the hearer matches some external criterion. In this sense, we might say that sincerity in social meaning games depends on the signaler following their own preferences over personas.

This view of sincerity disregards the role of the hearer completely. But hearer preferences also play a large role in social meaning: indeed, Henderson and McCready take dogwhistles to be useful just because of their ability to (potentially) bypass (potentially) hostile listeners, a situation reflected in the formal model by allowing receiver preferences over personas to affect signaler payoffs. Still, in the social meaning case one wouldn't want to say that hearer preferences are involved in determining whether a particular discourse move communicating social meaning was sincere. Indeed, a move designed to maximize speaker utility by considering only hearer preferences is one that we would ordinarily judge to be precisely insincere, because it ignores the genuine preferences of the speaker over self-presentations.

All this suggests a route to the analysis of sincerity in aesthetic communication. In a recent paper on sincerity in artistic production, Nguyen (2021) considers a particular album made by Kanye West, which was mostly critically panned. Nguyen acknowledges that it might be a bad record, but states a preference for a bad album that reflects the artist's vision over one that is made only with an eye toward commercial success. For Nguyen, it's more of an art crime to sell out than to make bad work; we can read this as valuing artistic sincerity over artistic competence (itself perhaps a way of understanding Gricean Manner, a point I'll return to in the conclusion). Nguyen writes the following:

We ask for sincerity [in art practice], rather than kindness, cooperativeness, or reliability, because we are trying to nurture creativity and originality, and because we are hoping to encourage a very particular form of shared experience.

It's this notion of creativity and originality that seem to be key to determining whether a piece of art is sincerely produced or not. Since of course literary work is art, the same holds for it as well, and consequently for the generalization of Gricean Quality to sincerity we are interested in here.

Nguyen offers the following definition for aesthetic sincerity.

(15)     **Trust in aesthetic sincerity** (Nguyen)

To trust in X's aesthetic sincerity is to rely on X to meet their commitment to act from aesthetic considerations (for some set of acts).

This style of definition clarifies the connection between aesthetic sincerity and the communicative (truth-conducive) and social sincerity under consideration so far. We can easily write parallel definitions in this style for the kinds of sincerity we have seen so far:

(16) **Communicative sincerity**

To trust in X's communicative sincerity is to rely on X to meet their commitment to act as a truth-cooperative speaker.

(17) **Social sincerity**

To trust in X's aesthetic sincerity is to rely on X to meet their commitment to sincerely present a social persona.

But all this leaves aside the question of what exactly it means for an agent to act from aesthetic considerations. Nguyen's Kanye West example suggests that it's a commitment of the artist to their own artistic vision and aesthetic sense. This idea can be spelled out easily in the setup of the last section: a work  $W$  by producer  $p$  is sincerely produced if, and only if,  $Aes_S^p(W) \geq t$ , for  $t$  a threshold. This is just to say that the work is aesthetically significant for the producer themselves; note the similarity to the generalized notion of sincerity for social meaning proposed above, on which a self-presentation was judged sincere if it followed the speaker's actual preferences over personas as indicated by their utility function.

Thus, Gricean Quality for social meanings and aesthetic content correlates with communication that matches the speaker's actual preferences; likely, this definition will work for any domain that exhibits the kind of real-numbered structure involved in the interpretation of these kinds of content. Further, a generalized notion of aesthetic trust can be constructed from this basic definition as well, just as with truth-conducive communication in McCready (2015); a producer who can be trusted to create sincerely is one whose history of aesthetic communications reflects a series of aesthetically sincere communicative acts. I take this parallelism to be a welcome result.

Of course, the sincerity or lack thereof of the producer might not be mirrored in the effect of the work on others. This is expected given the formalism. I might trust a producer to be aesthetically sincere because their production history reflects their aesthetic values; but if, consistently,  $Aes_S^I(W_p) \leq t$  (for  $I$  the speaker), I might not trust the producer to make work that's significant for me, because the work they

have made up to now failed to consistently move me. Here: personal taste comes to the fore; my own preferences may not match those of the producer, with the consequence that we have extreme disalignment in the aesthetic value we assign to particular communications. This is not a situation that arises with truth-conducive communication, for there disalignment simply means that someone is in error. But for the domain of aesthetic interpretation it makes perfect sense: given the subjectivity of this area (as reflected in the agent-relativity of the *Aes* function), we should expect aesthetic significance and trust as defined in the metalinguistic interpretation to share factors like faultless disagreement with object-language subjective predicates (Lasersohn, 2005).

The reader might wonder if this observation simply disqualifies what we have here as an analogue for Gricean Quality at all. One way to read Gricean Quality is as a partial description of the speaker's obligations to their listeners: they should speak the truth as best they can. But what is the obligation of an author to the reader? An author can write as sincerely as they like, but it might be that our trust in them qua author doesn't reflect this, because in some sense we might not care about whether they are sincere, but only about whether the work affects us. In this sense, the notion of sincerity defined here seems at first glance completely unlike what we find with truth. I myself don't take this as a disqualifier, but instead as a further sharpening of the nature of the domain under discussion, and how it differs from truth-conditional communication.<sup>18</sup>

## 4 Prospects

This paper has proposed a semantic framework for making sense of aesthetic communications. The theory has several key elements: affective and emotive responses derived via a function yielding the affect a particular agent derives from a text, itself derived from three functions corresponding to different sources of affect; the use of probabilities in interpretation with respect to affect; and generalized notions of significance and sincerity that can be used to begin the project of giving an aesthetic pragmatics. The framework is (I believe) successful in what it sets out to do, but there are of course many avenues for further investigation, elaboration, and extension. In this last section, I want to discuss some of these, beginning with an aspect of literary interpretation I have put aside so far, despite it being absolutely key to the interpretation of many literary texts.

---

<sup>18</sup>Thanks to Stefan Himmerwinter for pushing this point in conversation.

When reading a literary text, the question ‘what does this mean?’ is perhaps more usually understood as asking a question about the text’s deeper significance than about its literal interpretation, whether truth-conditional or not. What happens if a text is read nonliterally? What kinds of symbolism are at work in the text? Is the text a metaphor for something? Is there a way to read the text as holding lessons for the reader, whether moral or otherwise? In short, what does it *mean*? These are questions of theme (John, 2015), and they are often central to making sense of a literary work, whether in the attempt to recover the intent of the author or to extract significance for the reader.

One possible way to understand this sort of question is as what we might call *translations*: ways of taking the content of the text and mapping it to some related but different content which has symbolic import. Suppose we think of this in terms of a *translation function*, which I will write  $\mathcal{T}$ . The input to such a function is the content of the text, broadly conceived as truth-conditional content plus pragmatic content (presupposition, expressive content, social meaning, and so forth); the output is distinct but related. The basic intuition is that when interpreting a text for aesthetic response, we look for ways to map the content of the text onto other meaningful contents, while preserving aspects of the structure of the original. In this sense, translations are isomorphic to the source text. I am not the first, of course, to have this insight; many in literary criticism have said similar things, and indeed it turns out that Bauer and Beck (2022) have written precisely that isomorphic structures are involved in interpretation, though in a way somewhat different from what I have in mind here (in part because they assume that all aspects of a text can or should be subject to the mapping, presumably because the texts they consider as examples are relatively brief). Still, the way in which translations interact with aesthetic interpretation makes available new insights.

One involves the criteria we use for selecting an isomorphism to use as translation. This is a difficult and complex question and I have nothing like a full answer to it, but we can at least identify some key points. Firstly, the simplest case of a translation is, of course, the one in which  $\mathcal{T}$  is the identity function, ie. where the text read literally produces an affective response. Secondly, it seems a requirement that when choosing a translation that assigns symbolism to texts is that the output of  $\mathcal{T}$  should carry substantive affective value. This principle can be codified in various ways: one could say, for instance, that the result must be significant for the agent doing the translation, as in (18):

$$(18) \quad Aes_S^a(T([[Text]])) \geq t$$

Alternatively, one could require the result of the translation to be more significant than the original text, as stated in (19):

$$(19) \quad Aes_S^a(\mathcal{T}([Text])) \geq Aes_A^a([Text])$$

A second point is that  $\mathcal{T}$  should apply to as much of the text as possible. It seems clear that very few texts can be mapped in their entirety to symbolic structures: whatever it might be that the white whale symbolizes in *Moby Dick*, it seems very implausible that each of the three holes left as punctures in his starboard fluke can be mapped to any specific symbolic element; the same holds for many or most descriptive elements in texts. Nonetheless, the more textual elements that fall naturally into a translation, the better the translation is for capturing a reading of the text. We can formalize this as the ‘totality principle’ in (20), which says that a translation function that operates on more of the text than a second is to be preferred to it.

$$(20) \quad (TOT) \quad dom(\mathcal{T}_i([Text])) \sqsubseteq dom(\mathcal{T}_j([Text])) \rightarrow \mathcal{T}_i \leq \mathcal{T}_j.$$

These are only initial and programmatic remarks about the nature of translation functions. There is of course a great deal more to say here, but I am not in position to do it currently and so will leave doing so for future research.

There are very many more other interesting questions and issues to address, both in the semantics of aesthetic interpretation and in pragmatics. One would like to know more about exactly how affect is assigned by *Aes*, or rather by the functions  $\mathcal{S}, \mathcal{I}, \mathcal{F}$  that underpin it. This paper is already complex enough without turning to the psychology literature to spell out exactly how affect works and what the structure of its domains looks like. I have here pursued a very thin notion of affect on which it simply correlates to a domain of real numbers, but there are many other theories available, where affect is multidimensional or lives in much more complex domains (compare e.g. Silva 2023 on affective quality spaces for emotive content, or Carranza Pinedo 2022, who uses a complex affective domain for the analysis of slurs, among other things).

Another question involves the use of the formalism in other analytical and classificatory projects, such as the differentiation of genres and styles of writing. Do genres correspond to affective structures with peaks in different sources of affect, as I suggested above for propaganda? How finely can genres can be distinguished? Are there constraints one can introduce on agents, affective structures, or translations that correspond to (more) traditional modes of criticism or interpretation? How does all this relate to applications of affect assignment like sentiment analysis

or even the algorithms that suggest content for users of social media or shopping sites based on their past revealed preferences?<sup>19</sup> There are many questions on the semantic side.

The situation is no different for pragmatics. In section 3 I considered mostly Gricean cooperation, with some comments on other domains; there I gave an analogue to Gricean Quality in terms of aesthetic trust and sincerity. I argued that Quality for truth-conditional content differs in that it's a constraint on what's given by the speaker to the hearer, but that aesthetic sincerity and the trust that arises from it is derived from author alone. This is an interesting disanalogy; one wonders how it might surface in the other aspects of Gricean cooperation, which I have not explicitly addressed here. However, relevance for this domain is already present in the theory via the notion of aesthetic significance. A text is relevant in an aesthetic sense just in case its aesthetic value  $Aes^a_s(\langle E, [[E]] \rangle) \geq t$ ; here the main difference from (some formal implementations of) Gricean Relevance is that it involves affect rather than informational value to the question under discussion. But the other two maxims are left undiscussed. I have little to say about Quantity here, as I don't see how to make sense of it on the current theory, which lacks the requisite notion of entailment due to being based in a domain for which the idea of logical consequence makes little sense, and because maximization of affect is the goal, so there doesn't seem to be such a thing as 'too much information.' But I think there is more to say about Manner.

What does Manner amount to? Can we think of it as aesthetic competence in the sense of Nguyen (2021)? This seems a possibility, but I suspect that Manner involves a kind of accessibility of proper translation that mere competence doesn't capture, in the sense of the degree of difficulty faced by the interpreter of deriving significance from the work, or finding an appropriate translation. We might then think of Manner as requiring the author to produce work that has a significance that is accessible to the reader; an author then is obeying the Maxim of Manner for the aesthetic domain if they project that the reader can arrive at a significant, for them, interpretation of the text, which will also be a significant one for the author, given that the author is also following Quality. This is the kind of interaction between maxims that one hopes for. It also goes some way to alleviate concerns about the speaker-centeredness of Quality discussed at the end of the last section. Still, this view seems to imply that much experimental and conceptual literature involves failure or flouting of Manner, because of its relative inaccessibility; this implication needs a closer look.

---

<sup>19</sup>Thanks to Daniel Gutzmann (p.c.) for suggesting this application.



A final general question asks where the limits of this style of interpretation might lie. Can it be applied to other domains of artistic practice: dance and choreography, music, visual art, and so on? The domain of application is I think larger than just art created using language. Still, the way the mechanism works requires that the object of analysis have some aspects of the structure of language to be used successfully. Specifically, it must consist of discrete parts with discrete meanings which are combined into larger units. In aesthetic interpretation, affect is assigned to the individual parts and then to each unit they are part of. For some domains, this seems possible: perhaps, for example, it's possible to separate the performance of a dance into discrete movements, each of which has an individual meaning, and where each sequence of movements also has a (at least partly) derivable meaning, though it's not clear that it actually is in practice; but I don't think that attempting to do this makes sense for music, because it seems to me impossible to assign meanings to individual musical notes independently of the larger melodic contours in which they appear. To use the system for visual art would require extracting discrete parts of the object in ways that might be hard to warrant, but are likely possible in certain cases.

The theory does open one possibility that I think is highly welcome. Because affect assignment is entirely interpreter-dependent, the interpreter need not look to the author to justify the significance they find in a text. A consequence of this is that there need be no author of a text at all for it to be significant, and the text need not be intended as literature to be read for aesthetics. This is just as much so for texts authored by AI systems as it is for texts arising by random chance, for instance as rain flows over a paragraph painted on a wall over the course of years, and some portions of the text fade into illegibility. If we choose to interpret aesthetically, any text can be read as a work of art, and we can derive significance from it, just as any object in the world can be seen as an artwork or an installation in a way that gives the viewer aesthetic pleasure (or doesn't). Here, again, maximization of affect is the goal, meaning that we should do the best we can to extract meaning from natural configurations of linguistic or nonlinguistic objects we choose to view as art. Opening the possibility of this kind of interaction with linguistic objects via a formal treatment is a major positive contribution of this work, and opens the further possibility of other modes of interpretation and interaction with both textual and nontextual objects.

## References

- Luvell Anderson and Ernie Lepore. What did you call me? Slurs as prohibited words. *Analytic Philosophy*, 54(3):350–363, 2013.
- Nicholas Asher and Alex Lascarides. *Logics of Conversation*. Cambridge University Press, Cambridge, 2003.
- Yehoshua Bar-Hillel. Out of the pragmatic wastebasket. *Linguistic inquiry*, 2(3): 401–407, 1971.
- Roland Barthes. The death of the author. In *Image, Music, Text*, pages 142–148. Hill and Wang, 1977.
- Matthias Bauer and Sigrid Beck. Isomorphic mapping in fictional interpretation. In Emar Maier and Andreas Stokke, editors, *The Language of Fiction*, pages 277–296. Oxford University Press, 2022.
- Maria Bittner. *Temporality: Universals and Variation*. Wiley-Blackwell, 2014.
- Nicolas Bourriaud. *Relational Aesthetics*. Les Presse Du Reel, 1998.
- Heather Burnett. Signalling games, sociolinguistic variation and the construction of style. *Linguistics and Philosophy*, 2018.
- Heather Burnett. A persona-based semantics for slurs. *Grazer Philosophische Studien*, 2020.
- Elisabeth Camp. Slurring perspectives. *Analytic Philosophy*, 54(3):330–349, 2013.
- Elisabeth Camp. Insinuation, common ground, and the conversational record. In Daniel Fogal, Daniel Harris, and Matt Moss, editors, *New Work on Speech Acts*. Oxford University Press, Oxford, 2018.
- Victor Carranza Pinedo. *The Landscape of Affective Meaning*. PhD thesis, University of Milan and ENS, 2022.
- Wendy Hui Kyong Chun. *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition*. MIT Press, 2021.

- Stephen Davies. Authors' intentions, literary interpretation, and literary value. *The British Journal of Aesthetics*, 46(3):223–247, 2006.
- Christopher Davis. *Constraining Interpretation: Sentence Final Particles in Japanese*. PhD thesis, University of Massachusetts at Amherst, 2011.
- Christopher Davis and Elin McCready. The instability of slurs. *Grazer Philosophische Studien*, 97:63–85, 2020.
- Quentin Dénigot. *Blowing the Whistle: Formal Approaches to the Communication of Identities in Discourse*. PhD thesis, Université de Paris, 2022.
- Penelope Eckert. Variation and the indexical field. *Journal of sociolinguistics*, 12(4):453–476, 2008.
- Penelope Eckert. Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual review of Anthropology*, 41: 87–100, 2012.
- Penny Eckert. *Jocks and Burnouts: Social Identity in the High School*. Teachers College Press: New York, 1989.
- Donka F. Farkas and Kim B. Bruce. On Reacting to Assertions and Polar Questions. *Journal of Semantics*, 27(1):81–118, 2010.
- J. Farrell. Meaning and credibility in cheap-talk games. *Games and Economic Behaviour*, 5:514–531, 1993.
- Harry G Frankfurt. *On bullshit*. Princeton University Press, 2005.
- Miranda Fricker. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press, 2009.
- Matthew Fuller and Eyal Weizman. *Investigative Aesthetics: Conflicts and Commons in the Politics of Truth*. Verso, 2021.
- Peter Gärdenfors. *Knowledge in Flux*. MIT Press, 1988.
- Cheryll Glotfelty. Introduction: Literary studies in an age of environmental crisis. In *The Ecocriticism Reader: Landmarks in Literary Ecology*, pages xv–xxxvi. University of Georgia Press, 1996.

- H. Paul Grice. Logic and conversation. In Peter Cole and Jerry Morgan, editors, *Syntax and Semantics III: Speech Acts*, pages 41–58. Academic Press, New York, 1975.
- Jeroen Groenendijk and Martin Stokhof. Dynamic predicate logic. *Linguistics and Philosophy*, 14:39–100, 1991.
- Irene Heim and Angelika Kratzer. *Semantics in Generative Grammar*. Number 13 in Blackwell Textbooks in Linguistics. Blackwell, Oxford, England, 1998.
- Robert Henderson and Elin McCready. Dogwhistles, trust and ideology. In *Proceedings of the 22nd Amsterdam Colloquium*, pages 152–160, 2019.
- Robert Henderson and Elin McCready. *Signaling without Saying: The Semantics and Pragmatics of Dogwhistles*. Oxford University Press, 2024.
- Eileen John. Theme. In *The Routledge Companion to Philosophy of Literature*, pages 205–216. Routledge, 2015.
- Chris Kennedy. Vagueness and gradability: The semantics of relative and absolute gradable predicates. *Linguistics and Philosophy*, 30(1):1–45, 2007.
- Peter Lasersohn. Context dependence, disagreement, and predicates of personal taste. *Linguistics and Philosophy*, 28:643–686, 2005. doi: 10.1007/s10988-005-0596-x.
- Ernie Lepore. The heresy of paraphrase: when the medium really is the message. *Midwest Studies in Philosophy*, 33(1):177–197, 2009.
- David Lewis. *Convention*. Harvard, 1969.
- David Lewis. Truth in fiction. *American Philosophical Quarterly*, 15:37–46, 1978.
- Emar Maier and Andreas Stokke, editors. *The Language of Fiction*. Oxford University Press, 2022.
- Elin McCready. What man does. *Linguistics and Philosophy*, 31:671–724, 2008. doi: 10.1007/s10988-009-9052-7.
- Elin McCready. Emotive equilibria. *Linguistics and Philosophy*, 35:243–283, 2012.

- Elin McCready. *Reliability in Pragmatics*. Oxford University Press, 2015.
- Elin McCready. *The Semantics and Pragmatics of Honorification: Register and Social Meaning*. Oxford University Press, 2019.
- Elin McCready and Grégoire Winterstein. Testing epistemic injustice. *Investigationes Linguisticae*, 41:86–104, 2019.
- Maurice Merleau-Ponty. *Phenomenology of perception*, volume 26. Routledge London, 1962.
- Sarah Murray. Varieties of update. *Semantics and Pragmatics*, 7(2):1–53, 2015.
- Reinhard Muskens, Johan van Benthem, and Albert Visser. Dynamics. In Johan van Benthem and Alice ter Meulen, editors, *Handbook of Logic and Language*, pages 587–648. Elsevier, Amsterdam, 1997.
- Stephen Neale. Coloring and composition. In Robert Stainton, editor, *Philosophy and Linguistics*, pages 35–82. Westview Press, Boulder, CO, 1999.
- Alex Neill. "An unaccountable pleasure": Hume on tragedy and the passions. *Hume studies*, 24(2):335–354, 1998.
- C. Thi Nguyen. Trust and sincerity in art. *Ergo*, 8, 2021.
- Martin Nowak and Karl Sigmund. The dynamics of indirect reciprocity. *Journal of Theoretical Biology*, 194:561–574, 1998a.
- Martin Nowak and Karl Sigmund. Evolution of indirect reciprocity by image scoring. *Nature*, 393:573–577, 1998b.
- L.A. Paul. *Transformative Experience*. Oxford University Press, 2014.
- Christopher Potts. *The Logic of Conventional Implicatures*. Oxford University Press, 2005. Revised version of 2003 UCSC dissertation.
- Laura Silva. Towards an affective quality space. *Journal of Consciousness Studies*, 30(7-8):164–195, 2023.
- Robert Stalnaker. Assertion. In Peter Cole, editor, *Syntax and Semantics 9*. Academic Press, New York, 1979.

- Johan van Benthem and Fenrong Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logic*, 17(2):157–182, 2007.
- Robert van Rooij. Game theory and quantity implicatures. *Journal of Economic Methodology*, pages 261–274, 2008.
- Frank Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25: 221–261, 1996.
- Brian Weatherson and Andy Egan. Epistemic modals and epistemic modality. In Brian Weatherson and Andy Egan, editors, *Epistemic Modality*. Oxford University Press, 2011.