

EUROPEAN JOURNAL OF ANALYTIC PHILOSOPHY

UDC 101 | ISSN 1845-8475

Vol. 20, No. 2, 2024

EDITORIAL

MARKING THE 40TH ISSUE OF THE EUROPEAN JOURNAL OF ANALYTIC PHILOSOPHY, **Marko Jurjako**

SPECIAL ISSUE: „Aesthetic Education and Screen Stories“

Introduction by the guest editors, Iris Vidmar Jovanović and Ema Luna Lalić
ARTISTIC FREEDOM REALISED, **Murray Smith** | AESTHETIC EDUCATION: A PERCEPTUAL-COGNITIVE MODEL, **Ted Nannicelli** | ETHICS EDUCATION FROM SUFFERING ON SCREEN? TRAGIC VISIONS IN ARRIVAL, **James MacAllister** | THE SUMMIT OF SAFE HORROR: DEFENDING MOST HORROR FILMS, **Cara Rei Cummings-Coughlin** | CRINGE OVERHANG: THE PERLOCUTIONARY EFFECTS OF CRINGE COMEDY, **Alexander Sparrow** | WHAT COUNTS AS (EVIDENCE OF) NARROW AESTHETIC COGNITIVISM, **Mario Slugan**

RESEARCH ARTICLE

REVISITING MCGEE'S COUNTEREXAMPLE TO MODUS PONENS, **Nicolás Lo Guercio and Mariela Rubin**

BOOK REVIEWS

Any Kind IMAGINATION AND CREATIVE THINKING, **Ksenija Savčić** | Holly Lawford-Smith GENDER-CRITICAL FEMINISM, **Eugen Golubić** | Sanja Dembić PHILOSOPHY OF MENTAL DISORDER: AN ABILITY-BASED APPROACH, **Luca Malatesti**

EUROPEAN JOURNAL OF
ANALYTIC PHILOSOPHY

UDC 101
ISSN (Print) 1845-8475
ISSN (Online) 1849-0514
<https://doi.org/10.31820/ejap>

Open access



Editor-in-Chief

Marko Jurjako
University of Rijeka, Croatia

Associate editors

Elisabetta Lalumera
University of Bologna, Italy

Lovro Savić
University of Oxford, UK

Mia Biturajac
Institute of Philosophy, Croatia

Editorial administrator and proofreading

Iva Martinić
University of Rijeka, Croatia

Editorial board

Lisa Bortolotti (University of Birmingham), Anneli Jefferson (Cardiff University), James W. Lenman (The University of Sheffield), Luca Malatesti (University of Rijeka), Alfred Mele (Florida State University), Carlo Penco (University of Genoa), Katrina Sifferd (Elmhurst College), Majda Trobok (University of Rijeka), Rafał Urbaniak (University of Gdansk)

Advisory board

Miloš Arsenijević (University of Belgrade), Elvio Baccharini (University of Rijeka), Carla Bagnoli (University of Modena and Reggio Emilia), Boran Berčić (University of Rijeka), Clotilde Calabi (University of Milan), Mario De Caro (University of Rome), Raphael Cohen-Almagor (University of Hull, UK), Jonathan Dancy (University of Reading/University of Texas, Austin), Mylan Engel (University of Northern Illinois), Katalin Farkas (Central European University), Luca Ferrero (University of California, Riverside), Paul Horwich (City University New York), Pierre Jacob (Institut Jean Nicod, Paris), Kerry McKenzie (University of California, San Diego), Kevin Mulligan (University of Geneva), Snježana Prijić-Samaržija (University of Rijeka), Michael Ridge (University of Edinburgh), Sally Sedgwick (Boston University), Mark Timmons (University of Arizona, Tucson), Nicla Vassallo (University of Genoa), Bruno Verbeek (University Leiden), Alberto Voltolini (University of Turin), Joan Weiner (Indiana University Bloomington), Timothy Williamson (University of Oxford), Jonathan Wolff (University College London)

Publisher, Editorial office

University of Rijeka, Faculty of Humanities and Social Sciences, Department of Philosophy

Address: Sveučilišna avenija 4, 51000 Rijeka, Croatia (EU)

Phone: +385 51 669 794

E-mail: ejap@ffri.uniri.hr

Web address: <https://www.ejap.uniri.hr>

Printed by Grafika Helvetica d.o.o., Rijeka (Croatia)

The publication of the journal is financially supported by the Ministry of Science, Education and Youth of the Republic of Croatia.

TABLE OF CONTENTS

EDITORIAL

Marking the 40th Issue of the European Journal of Analytic Philosophy

Marko Jurjako

Introduction to the Special Issue “Aesthetic Education and Screen Stories”

Iris Vidmar Jovanović and Ema Luna Lalić, guest editors.....255-258

Artistic Freedom Realised

Murray Smith.....259-282

Aesthetic Education: A Perceptual-Cognitive Model

Ted Nannicelli.....283-305

Ethics Education from Suffering on Screen? Tragic Visions in Arrival

James MacAllister.....307-322

The Summit of Safe Horror: Defending Most Horror Films

Cara Rei Cummings-Coughlin.....323-343

Cringe Overhang: The Perlocutionary Effects of Cringe Comedy

Alexander Sparrow.....345-362

What Counts as (Evidence of) Narrow Aesthetic Cognitivism

Mario Slogan.....363-388

REASARCH ARTICLE

Revisiting McGee’s Counterexample to Modus Ponens

Nicolás Lo Guercio and Mariela Rubin.....389-412

BOOK REVIEWS

BOOK REVIEW Amy Kind IMAGINATION AND CREATIVE THINKING

Ksenija Savčić.....413-420

BOOK REVIEW Holly Lawford-Smith GENDER-CRITICAL FEMINISM

Eugen Golubić.....421-424

BOOK REVIEW Sanja Dembić PHILOSOPHY OF MENTAL DISORDER: AN ABILITY-BASED APPROACH

Luca Malatesti.....425-431

ABSTRACTS (SAŽECI).....(AB)

AUTHOR GUIDELINES

AND MALPRACTICE STATEMENT.....(AG)

Editorial

MARKING THE 40TH ISSUE OF THE EUROPEAN JOURNAL OF ANALYTIC PHILOSOPHY

Marko Jurjako¹ 

¹University of Rijeka, Faculty of Humanities and Social Sciences, Croatia

On the occasion of the 40th issue of the *European Journal of Analytic Philosophy* (EuJAP), I am delighted to report that the journal is thriving. Since its inception, EuJAP has aimed to provide a forum for papers written in the tradition of analytic philosophy across all areas of philosophical study, with a focus on publishing high-quality contributions that address contemporary issues in analytic philosophy as well as those intersecting with neighboring disciplines. Since its inaugural issue in 2005, EuJAP has published more than 240 research articles, reviews, and discussion papers, along with 18 book reviews, contributed by over 270 authors. The journal publishes two issues per year, one of which is typically a special issue dedicated to a specific topic in contemporary analytic philosophy. Over the years, we have collaborated with more than 30 guest editors, who have played a crucial role in preparing special issues, and have relied on countless reviewers, whose efforts ensure the selection and publication of impactful papers. We extend our sincere gratitude to everyone who has contributed to the smooth operation of EuJAP and to its mission of advancing analytic philosophy.

In recent years, we have witnessed a steady but significant surge in submissions. Over the last couple of years, we have been receiving approximately 70 submissions annually, with a rejection rate of around 80%. Most of the rejected papers are desk-rejected as they are deemed to be of lower quality and unlikely to receive a positive review from our referees. Additionally, since being indexed in the Scopus and Web of Science databases, we have experienced an increase in what appears to be random submissions of papers that fall outside the scope of EuJAP. For instance, once EuJAP simultaneously received five papers from a single author, all dealing with topics related to tourism and hospitality in East Asia, accompanied by a query asking how much they needed to pay for their papers to be published in EuJAP. A clearer picture of rejection rates might emerge when focusing on papers that enter the review process. In

this case, my rough estimate is that around 50% of externally refereed papers are eventually accepted.

EuJAP remains firmly committed to diamond open access publishing, with all its issues freely available through the repository of the Portal of Croatian Scientific and Professional Journals ([HRČAK](#)). This commitment is particularly noteworthy given the increasing volume of submissions and rising publishing costs, especially since we continue to produce printed issues alongside the freely available online ones. Journals published in Croatia have so far been able to sustain this model thanks to financial support from the Croatian government and university publishers. In this regard, we extend our sincere gratitude to our publisher, the Faculty of Humanities and Social Sciences in Rijeka, as well as to the Ministry of Science, Education, and Youth of the Republic of Croatia for their invaluable financial support. We also wish to thank the National and University Library in Zagreb for providing DOI identification numbers for our research articles.

We have been regularly publishing two issues per year since the first issue in 2005. There was only a “slight” delay with the publication of Volume 10, No. 2, in 2014, which was dedicated to the work of Professor Nenad Miščević (<https://hrcak.srce.hr/en/broj/13340>).

Let me take this opportunity to honor Professor Miščević, who passed away earlier this year. He was one of the key figures in contemporary philosophy in Croatia and probably it is not too much to say that he was a person who almost single-handedly introduced and popularized analytic philosophy in the country. He was certainly one of the main figures and supporters of the establishment of the Department of Philosophy in 1998 that eventually enabled the transformation of the old Faculty of Pedagogy in Rijeka into its modern format known as the Faculty of Humanities and Social Sciences. Even though Professor Miščević was not one of the founders of EuJAP—in fact, according to some anecdotes, EuJAP was created partly as a reaction to the success of the *Croatian Journal of Philosophy* (<https://hrcak.srce.hr/en/cjp>), where Professor Miščević was a founding member and lifelong editor—there would not be EuJAP without his support, mentorship, and wisdom. Without his influence, the Department of Philosophy at the University of Rijeka, and by extension EuJAP, would not exist today. To commemorate his life and work, our publisher, the Department of Philosophy at the Faculty of Humanities and Social Sciences in Rijeka, is organizing a conference dedicated to Professor Miščević, which is planned to take place in April 2025.

We are working to maintain EuJAP's online presence and promote our papers across several social media platforms, including Facebook, Twitter/X, and Bluesky. Additionally, we use the blog capabilities of our website to share news and highlight journal publications. Occasionally, I consider expanding these activities to include features such as acknowledging the most cited papers in EuJAP or selecting an editor's choice of significant articles from the latest issue. However, such initiatives often remain unrealized due to limited resources, my own constraints, and the all too transient enthusiasm for such efforts.

Nonetheless, our efforts in running and promoting the activities of EuJAP have produced visible results. EuJAP is performing well on the scientometric front. The steady, linear progression of citations for EuJAP's articles can be tracked on its [Google Scholar web page](#). The journal is indexed by all major indexing services, including Scopus and Web of Science (WoS). Over the last two years, we have maintained our status in the first quartile (Q1) according to [Scimago](#), a database that derives publication metrics from Scopus. Furthermore, EuJAP is ranked in the first quartile for philosophy according to WoS's Journal Citation Indicator (JCI). I remain hopeful that by continuing to publish impactful papers and promoting the journal through our editorial board members, authors, reviewers, colleagues, and social networks, we will sustain our position as an important venue for analytic philosophers and further enhance our standing within the global academic community.

It is worth noting, however, that despite its publishing success and established reputation, EuJAP is not currently recognized in Croatia for career advancement purposes. Similar to the Italian ANVUR classification system, Croatia employs a centralized system for classifying journals in the humanities, which determines their validity for career progression. This system divides journals into A1 and A2 categories as defined by the relevant *matični odbor* (field-specific board) for the humanities. The A1 category is supposed to include the highest-quality journals published in Croatia, while the A2 category encompasses all other journals meeting the minimal criteria for scientific publications in Croatia. Currently EuJAP does not belong to either the A1 or A2 categories, effectively rendering it non-eligible for career advancement in Croatia.

Earlier this year, we submitted EuJAP for inclusion in the A1 category. However, given the lack of transparency regarding the criteria for A1 classification—especially when examining the current list of A1 and A2 journals—we remain uncertain about the prospects of achieving this status. That said, our primary goal is not to become a popular local

journal but to establish ourselves as a prominent venue for analytic philosophy on the European and global stage. I dare say, while currently sitting in one of the birthplaces of contemporary analytic philosophy, that we are on a promising trajectory to achieve this ambition.

EuJAP would not have achieved its current status without the diligent work of its past and current editorial boards, and especially its previous chief editors—Carla Bagnoli, Majda Trobok, and Luca Malatesti—who played pivotal roles in shaping the journal into its present form. Each of them faced and overcame distinct challenges, from founding the journal to maintaining its standards and fostering its growth. I remain confident that their foundational efforts have set the stage for EuJAP to continue thriving and establishing itself as a prominent platform for analytic philosophy.

Marko Jurjako

Editor-in-Chief, *European Journal of Analytic Philosophy*
Cambridge, UK, December 2024

INTRODUCTION TO THE SPECIAL ISSUE ON AESTHETIC EDUCATION AND SCREEN STORIES

GUEST EDITORS

Iris Vidmar Jovanović¹  and Ema Luna Lalić¹ 

¹ University of Rijeka, Croatia

ABSTRACT

This article is an introduction to the special issue of the *European Journal of Analytic Philosophy* dedicated to aesthetic education and philosophy of screen stories.

Keywords: analytic aesthetics; aesthetic education; screen stories; philosophy of film.

The present issue of the *European Journal of Analytic Philosophy* is mostly composed of papers that were presented at the conference *Aesthetic Education and Screen Stories*, held at the University of Rijeka in September 2023 and organized as part of the research project *Aesthetic Education through Narrative Art and Its Value for the Humanities*, funded by the Croatian Science Foundation (project AETNA).

The aim of the conference was to bring together scholars working in philosophy of art and aesthetics, film theory, media studies, cultural studies and other domains dedicated to the study of film and television, with the aim of exploring cognitive, ethical, educational and other aspects of the aesthetic experience of film and television series and serials—screen stories for short. We were also interested in analyzing the notion of aesthetic experience, aesthetic education and in exploring methodological advantages and disadvantages of various approaches to screen stories, in particular those related to neuro-aesthetics, naturalized aesthetics and cognitive aesthetics. In our opinion, the papers collected here present illuminating theoretical perspectives on these issues.

The papers collected here can be roughly divided into those dedicated to the problem of aesthetic education, those dealing with ethical dimension of specific artistic genres, and those concerned with particular kind of knowledge available from cinematic art. In the following, we provide an overview of the papers in this special issue.

In “Artistic Freedom Realized”, *Murray Smith* considers the notion of artistic freedom and the question of when or how it is realized, contrasting two dominant views on it: on the theory advanced by Jon Elster, the problems with which artists engage arise from a mix of chosen, imposed, and invented constraints; on the theory of David Bordwell, those problems can be solved through the replication, revision, synthesis, or rejection of existing solutions. Smith concludes that neither the folk theory, nor the alternative theory, are correct; rather, there is a “sweet zone” of artistic creativity poised between a disabling surfeit of options, and a stifling sparsity of them. The inspiration for this view comes from Stravinsky, who claims: “[t]he more constraints one imposes, the more one frees one’s self of the chains that shackle the spirit”. To corroborate his view, Smith analyzes the interplay between invention and convention in art, drawing on a variety of examples from film and rock music.

In “Aesthetic Education: A Perceptual-Cognitive Model”, *Ted Nannicelli* discusses the puzzle of aesthetic education, pointing to the fact that although we invest significant time, energy, and resources to aesthetic

education, we also act as if it is true that there is no disputing taste: while we may try to persuade students to come around to particular judgments, we do not penalize students for judging one way or another. Nannicelli explores the implications of this puzzle, emphasizing the importance of aesthetic education and suggesting a new paradigm that may better explain its value. As he argues, the primary purpose of aesthetic education is not to educate taste but to facilitate the development of certain perceptual-cognitive capacities so as to enhance aesthetic experience and improve aesthetic appreciation. To support his theory, Nannicelli relies on research in perception and shows how one can improve one's perceptual capacities.

In "Ethics Education from Suffering on Screen? Tragic scenes from *Arrival*", *James MacAllister* discusses the capacity of tragedy to enhance ethics education. He first discusses Amelie Rorty and Peter Lamarque's accounts of the relationship between tragic art, moral education, and aesthetic value, claiming that they face issues due to the manner in which they employ the notion of aesthetic value. MacAllister then presents his own account, arguing that tragic films have potential to ethically educate audiences in a way that enhances the aesthetic value of the films in at least three directions: 1) by deepening moral understanding, 2) by deepening understanding of the nature of human beings and ethical purpose, and 3) by deepening understanding of ethical theory. To exemplify his points, MacAllister analyzes Denis Villeneuve's film *Arrival*.

Horror genre is at the center of *Cara Rei Cummings-Coughlin's* paper "The Summit of Safe Horror: Defending most Horror Films", intended as a defense of horror from the charge issues by Di Muzio, who fears the morally corruptive influence of horror films on the audience who regularly engages with them. Cummings-Coughlin proposes the category of safe horror, which does not invite the effect of uncanny, but rather has features that allows films to be intense enough to cause excitement and terror yet not so intense as to cause a negative moral attitude to form in our soul, such as comic relief and foolish choices by the characters.

In "Cringe Overhang: The Perlocutionary Effects of Cringe Comedy", *Alexander Sparrow* brings the speech act theory to his analysis of cringe comedies, i.e. those that make people uncomfortable. Sparrow analyzes the mixture of feelings that comedies can have, showing how the surplus of embarrassment is due to maximization of the violation in the comedy at the expense of the benign. On Sparrow's view, cringe comedy's funniness is reliant on its lack of social psychological distancing; by leaving no room between the viewer and the character, embarrassment is

maximized, the comedy is less benign (i.e. a stronger violation) and more polarizing as a result. His account explains i) why cringe comedy produces a comedic “overhang” in some viewers, where they continue to cringe even after the comedy has stopped, and ii) why cringe comedy produces a laughter response in some audiences, and stress responses in others.

In “What Counts as (Evidence of) Narrow Aesthetic Cognitivism”, *Mario Slugan* presents a critical analysis of narrow aesthetic cognitivism, the view that narrative fictional works are a source of knowledge. Slugan does a great job in showing deficiencies of theoretical approaches to the notion of learning from art, suggesting that scholars have to a great extent neglected to consider empirical evidence on the issue. Focusing on the claim that fiction gives us true propositions, Slugan points out that arguments in favor of the possibility of works expressing truth do not discuss how truth is justified. He then concedes that justification of truths in fictional and nonfictional works rests on the experiences outside of the work and that it is an empirical matter whether belief has been acquired or not. He concludes with a discussion of what would count as evidence in favor of aesthetic cognitivism.

We would like to conclude this introduction by expressing our gratitude to the *European Journal of Analytic Philosophy* for dedicating this issue to the topics discussed at the conference *Aesthetic Education and Screen Stories*, and to its editor, Marko Jurjako, whose help, support, and assistance were fundamental in bringing this issue together. Special thanks go to Murray Smith (University of Kent), the conference keynote speaker, for agreeing to contribute a paper to this special issue. We would also like to thank all the authors who submitted their papers and the reviewers whose critical comments were invaluable in refining the final versions. This is the first time the *European Journal of Analytic Philosophy* has featured papers primarily dedicated to art and focused solely on screen story arts. We hope this collection of papers will be an enjoyable and thought-provoking read and that it will inspire further discussions!

ARTISTIC FREEDOM REALISED

Murray Smith¹ 

¹ University of Kent, United Kingdom

Original scientific paper – Received: 20/06/2024 Accepted: 20/09/2024

ABSTRACT

The fewer the constraints we encounter in an artistic endeavour, the greater our artistic freedom; as technology advances and presents us with more numerous options and ever-greater creative flexibility, so our artistic freedom burgeons. Such is the folk wisdom on the phenomenon. Is the wisdom wise? Many artists and art theorists think not; for Stravinsky, “[t]he more constraints one imposes, the more one frees one’s self of the chains that shackle the spirit”. I consider the orthodox view and this alternative view through an exploration of the constraints arising from physical, psychological, and technological factors, as well as the heterogeneous array of routine practices we label “conventions”. To deepen the analysis, I turn to the “problem-solving” perspective on artistic creativity, examining the interplay between invention and convention in art, drawing on a variety of examples from film and rock music. On the theory advanced by Jon Elster, the problems with which artists engage arise from a mix of chosen, imposed, and invented constraints; on the theory of David Bordwell, those problems can be solved through the replication, revision, synthesis, or rejection of existing solutions. I conclude that neither the folk theory, nor the alternative theory, are correct; rather, there is a “sweet zone” of artistic creativity poised between a disabling surfeit of options, and a stifling sparsity of them. I argue that Stravinsky’s counter-theory of artistic creativity, though not literally true, acts as a “felicitous falsehood”—an epistemically valuable overcorrection.

Keywords: artistic freedom; artistic creativity; shot/reverse shot; film noir; neo-noir; Peter Gabriel; Nick Cave.

1. Introduction

We said that the American cinema pleases us, and its filmmakers are slaves; what if they were freed? And from the moment that they were freed, they made shitty films.

François Truffaut¹

Consider the following item of folk wisdom on the arts: the degree of creative freedom enjoyed by an artist is a function of their freedom from constraint; an idea which might derive from the same principle applied to freedom (of thought, of expression, of action) more generally. “Prima facie it would seem that nobody could have a motivation for discarding options”, writes Jon Elster.

Most people would rather have more money than less, more occupational options rather than fewer, rewards sooner rather than later, a larger range of potential marriage partners rather than a smaller one, and so on. (Elster 2000, 1)

Advocates of free jazz, free improvisation, and free verse seem to have embraced the principle fully, freeing themselves of some of the most traditional constraints in music and poetry respectively.

Not all folk wisdom is misleading; it isn’t called “wisdom” for nothing. And indeed we will find that there are reasons why this particular folk belief is widely held. Nonetheless it is, I shall argue, at best misleading, and a very different conception of the relationship between artistic freedom and constraint is more plausible.

2. Conventions and other varieties of constraint

What could be more obviously—perhaps even trivially—true than the idea that freedom, in the artistic as in any other domain of action, obtains in an inverse relationship with constraint? It seems almost analytically true, in the sense that our notion of freedom is defined at least in part by the absence of constraint. You’re free to express your beliefs if there’s no censor around to penalise or imprison you if you give expression to what are deemed politically unacceptable beliefs. You’re free to travel throughout a given territory which grants you such freedom of movement

¹ Bordwell, Staiger, and Thompson (1985, 5) quoting Truffaut (1963-4, 20).

(for us Brits, that now means just the UK, but it used to mean the whole of the EU), and there are no border agents around to inhibit your movements. In these cases, your freedom to express your beliefs and to travel (within the territory) are, precisely, *unconstrained*—that is what it means to have such freedoms.

Very importantly, these freedoms are in practice conditional on freedom from poverty, that is, on your economic well-being: you won't be in a good position to exercise the freedom to express your beliefs or to travel the territory if you're constrained by a lack of material means. As David Axelrod makes the point: "If you're sitting around the kitchen table talking about democracy and the rule of law, chances are you're not worried about the cost of the food on your table" (Smerconish 2024). And the more material means you have at your disposal, the greater your freedom to exercise these other freedoms. This economic caveat doesn't change the logic of our orthodox understanding of freedom, according to which freedom means freedom from constraint; it simply identifies a further constraint which can limit our freedom.

It seems odd, then, not only that this apparent truism might not be quite true, but that something like the opposite might be true—at least in certain contexts. However, that is just what a good number of artists and theorists of art tell us. "[I]n art as in everything else, one can build only upon a resisting foundation: whatever constantly gives way to pressure, constantly renders movement impossible", argues Igor Stravinsky.

[M]y freedom will be so much the greater and more meaningful the more narrowly I limit my field of action and the more I surround myself with obstacles. Whatever diminishes constraint, diminishes strength. The more constraints one imposes, the more one frees one's self of the chains that shackle the spirit. (Stravinsky 1947, 64-5)

Stravinsky's formulation courts paradox; part of our task is to figure out whether we can illuminate the source of the conundrum and restate it in less paradoxical fashion.²

² One possible solution is that we should distinguish artistic *freedom* from artistic *creativity*, thereby sidestepping any problems with our conception of artistic freedom. The difficulty here is that assumptions about freedom as lack of constraint are built into orthodox conceptions of artistic creativity, so we cannot free ourselves of these faulty assumptions simply by speaking of artistic creativity rather than freedom. By the same token, by tackling the problematic folk theory of artistic freedom, we stand to improve our conception of artistic creativity.

Writing in a less personal, more coolly theoretical mode, Rudolf Arnheim famously declared: “Art begins where mechanical reproduction leaves off, where the conditions of reproduction serve in some way to mold the object” (Arnheim 1983, 57). When an artist goes about creating a new work, they come up against the “conditions of reproduction”—by which Arnheim means the tools and materials constituting a given artform. Minimally, for a painter, that means a surface to be marked, an instrument with which to mark it, and a substance applied by the instrument to mark the surface; for a filmmaker, a camera, a set or a real location, and the capacity to edit the shots produced by the camera.³ Both must work with, and within the limits of, these conditions of reproduction. Arnheim famously advocated for filmmakers to limit themselves to the technological “conditions” of “silent cinema”, eschewing the temptations of colour cinematography and synchronous sound. But even filmmakers embracing these innovations in the technological basis of film still worked within a set of conditions of reproduction—precisely those set by the new technology. Though the conditions will vary across artforms and across time and place, no artist working with(in) a medium can avoid constraints of this type.

Beyond the limits set by the technology with which an artist works lies another set of constraints: that vast and motley array of routine practices we label *conventions*. In the context of the arts, a convention is any well-established norm of practice that is understood by both artists and appreciators (including the professional variety of the latter, known as “critics”, as well as ordinary readers, listeners, and viewers), and is understood by them mutually (that is, an aspect of an artist’s understanding of a convention is that appreciators understand it as well, and vice versa).⁴ In the context of film, to take one example, we can distinguish narrative and stylistic conventions. The narrative convention of *closure* is the practice whereby the major questions raised by a narrative film are answered, and the immediate fate of the central

³ At least, so Arnheim and many other theorists have held, but what this really describes is orthodox live action filmmaking. Animation does not require a set or location, and not all films make use of editing: consider both the earliest single-shot films, the single-shot feature film emerging in the digital era, and the many avant-garde films eschewing editing (MacDonald 2009). Nonetheless, the potency of the orthodox conception of is evident from the case of *Rope*, discussed further below, which trades on the assumption that editing is, in the terminology of Kendall Walton, a *standard* feature of mainstream live action filmmaking: an aspect of a particular practice of filmmaking (and in Walton’s terms, a particular category of art), not an invariant “medium” of film (Walton 1970). For discussion, see Carroll (2003, 2006) and Smith (2006).

⁴ On the nature and structure of mutual understanding between artist and audience, see Carroll (1996a, 218), on the case film metaphor, and Lopes (2024, 100), on fifteenth-century painting, drawing on Baxandall (1988).

characters is resolved at the end of the film. The stylistic convention of *shot/reverse-shot* is the practice of alternating shots of two or more interacting figures, founded on the pattern of conversational turn-taking, while maintaining consistent screen direction across distinct shots (if character A is shown to the left of character B in shot X, then even as the camera changes position to shot Y, character A will still be represented as situated to the left of character B). [See **Figure 1**]



Figure 1: Shot/reverse-shot in *L.A. Confidential* (1997)

Note that describing a representational or artistic practice as a convention does not necessarily mean that it is *arbitrary*—that an indefinite range of other practices would have served the functions of the conventional practice just as well. The function of the shot/reverse-shot sequence is to represent, in a streamlined and dramatically heightened form, the states of mind—beliefs, desires, intentions, emotions—of two or more characters as they interact. Such character-focussed interactions are central to Hollywood-style storytelling; the degree of intensity of these encounters among characters will vary with the rise and fall of the action, but will typically be marked by a high degree of clarity with respect to what characters want and what is at stake in the unfolding story.

The shot/reverse-shot practice does not conjure up such dramatic focus out of nothing, however. Such conventions “[cannot] grow except on the rich substratum of universal, naturally developing event comprehension” (Boyd 2009, 151-2), and more generally on the basis of the “contingent universals” of human cognition and behaviour (Bordwell 2008; Nannicelli 2017). Conversational turn-taking is one such contingent universal: across time and place, humans interact through alternating the role of speaker and listener. (Of course the shape of such conversational interaction will vary beyond this basic structure, and the dynamics of social power will undoubtedly shape their exact form: who gets to speak, when, and for how long.) Deictic gazing—the instinct to follow the gaze of an interlocutor to its target—is another contingent universal; attending to whatever our fellow humans attend to and deem significant in the immediate environment of interaction is basic to human behaviour. Shot/reverse-shot builds on both universals: the alternation of shots mimics the pattern of alternation in conversational turn-taking, and the reciprocal glances of the characters towards each other tighten our

attention on them, drawing on our hardwired, evolved and embodied understanding of the significance of such glancing.

While in the first instance filmmakers, and in the second instance critics, are more likely to possess explicit knowledge of routine practices like shot/reverse-shot and narrative closure—that is, they are more likely to be able to describe them in abstract and theoretical terms, as I have done here—what filmmakers, critics, and ordinary viewers share is familiarity with the look, feel, and function of these conventions. They know a shot/reverse-shot sequence when they see one and have a tacit understanding of its functions. Similarly, they experience the “sense of an ending” when a narrative work heads towards closure, and tacitly recognise the function and importance of such resolution in (canonical) narrative form. This tacit knowledge is comparable with linguistic *competence*: ordinary speakers of a language are skilled in both generating and understanding utterances, even though they do not possess a self-conscious, theoretical understanding of language in general or any particular language they speak (their ability to describe the grammatical rules of the language will be limited, for example).

Longstanding conventions, working in clusters, form genres, or at a higher level, traditions. Thus, we can recognise a variety of genres—romantic comedy, horror, the war film, the Western, and so on—all of which work within the broader parameters of the Hollywood tradition. The clusters of conventions which constitute genres, or the broader traditions of which they form a part, are themselves (in part) matters of convention. The Hollywood crime thrillers of the 40s and 50s that came to be labelled *films noirs* were (as the tag implies) typically shot in black and white, and—picking up on the point above—that norm of practice wasn't arbitrary; a low-key, monochrome palette was a good expressive fit for the bleak mood of the stories these films told. [See **Figure 2**]



Figure 2: Black and white cinematography in *The Big Combo* (1955)

As we can see from this example, the claim that a convention is not arbitrary does not depend on it being directly traceable to and explicable by an underpinning contingent universal (as is plausible in the case of shot/reverse-shot). In this case, what motivates and explains the routine choice of black and white for crime thrillers is that a well-established photographic and cinematographic convention (that black and white tonality conveys dark moods effectively) would fulfil the expressive function appropriate for the emerging trend in crime films that came to be called *film noir*. In this way conventions may build upon one another, and a convention may be motivated (rather than arbitrary) by drawing on the affordances and associations of other conventions, and not only by implicit, direct appeal to a contingent universal.

So the use of black and white for crime thrillers was a motivated norm in this sense; but one that could be broken, as seen in *Leave Her to Heaven* (1945): as nasty a *film noir* as one could wish for, resplendently shot in—Technicolor. [See **Figure 3**] For critic James Agee, the combination was jarring: “the story’s central idea might be plausible enough in a dramatically lighted black and white picture”, he averred, but the “rich glare of Technicolor” was inappropriate (quoted in Berliner 2017, 106; see also Keating 2010, Part III). By the time of the neo-noir trend in the 1990s, however, colour had become the norm for such films, as seen in *L.A. Confidential* [see **Figure 1**, above]. To some extent, the choice of black and white in the original noirs, and colour in the later neo-noirs, was overdetermined by the broader cinematographic norms of the period: in the 40s most films were shot in black and white, while by the 90s, most were shot in colour.



Figure 3: Technicolor cinematography in *Leave Her to Heaven* (1945)

Similarly, while the use of shot/reverse-shot is pervasive in Hollywood filmmaking, but some directors—most famously Orson Welles and William Wyler—depart from this widely-observed convention by rendering dramatic scenes in other ways and using shot/reverse-shot much more sparingly than in the typical case. Decried at the time by some, the departure from convention was also celebrated, by André Bazin and generations of subsequent critics. David Bordwell suggests that continuity editing, of which shot/reverse-shot is a central component, is so tenacious “because it has proved itself well suited to telling moderately complicated stories in ways that are comprehensible to audiences around the world” (*Film Style*, 155), and such cross-cultural accessibility is a goal of Hollywood as an institution. Once again, the norm is well-motivated; but for all that, as the cases of Welles and Wyler demonstrate, it is not inviolable.

As these examples tell us, not everything can be a matter of convention, in both a metaphysical and a historical sense. A convention is, or describes, a *type* rather than a *token*; it describes a norm of practice, not any particular instance of that practice. An individual token—a particular work—can be more or less conventional, and thus specific, highly-conventional works can serve as useful examples of conventions. But it would be a category mistake to construe an individual work *as* a convention, since the concept “convention” describes a more abstract kind of object (*viz*, a type).

Metaphysically, then, we need to acknowledge both tokens (individual works) and types (conventions, practices, and norms forming the context for the making and reception of individual works). This has at least two implications. Any act of tokening—of artistic production—will have a minimally creative dimension, by virtue of bringing a new, quantitatively-distinct work into being. Even a work judged to be utterly conventional will be “novel” in this minimal sense: something that did not exist before now does exist.⁵ Registering the distinction between tokens (works) and types (conventions) is also important historically, for it is through tokens—individual works—that *invention* is realised. Conventions don’t fall from the sky: practices have to be invented and

⁵ Relatedly, see Smith (2024a), where I argue: “While individual works may conform to established artistic norms to different degrees, no work is entirely *sui generis*; even the most avant-garde works invoke certain norms by (often ostentatiously) violating or disdaining them. On the other hand, every work is unique and distinctive to some degree, and to that extent cannot be exhaustively understood by reference to a genre or category of which the work is held to be a member”.

explored before they can become norms, even where they exploit contingent universals like conversational turn-taking and deictic gazing, or existing conventions such as the use of black and white tonality to express pessimism.

Things change; the arts have histories, evolving internally and in response to social, political, and technological developments. These observations bring new questions into view. How do filmmakers create new experiences for us, engaging us through novelty, against the background of established convention and tradition? How are we as spectators able to grasp these innovations—given that they are precisely new? And how does all of this relate to the question of artistic freedom and constraint?

3. The problem-solution dynamic

Earlier I suggested that the absence of constraint is simply part of the definition of freedom, and so the folk axiom that the fewer the constraints faced by an artist the greater the creative freedom they enjoy is straightforwardly a matter of logic. Resistance to the axiom is therefore futile! That's what makes it an axiom.

Where we humans are concerned, however, logic is never enough. In coming to understand the apparently paradoxical idea that artistic freedom increases with constraint, we enter the world of psychology. In particular, it is through the psychology of *problem-solving* that we arrive at the revised conception of artistic freedom. Problem-solving is central to human cognition, and may be defined as a “higher-order” form of cognition which recruits our basic cognitive capacities (of perception, memory, causal inference, affective appraisal, and so on) in the cause of finding solutions to both unique, and recurrent, problems faced by human agents. James Peterson puts such problem-solving at the centre of avant-garde film spectatorship:

[W]hen we encounter an avant-garde film, we struggle to integrate details into coherent wholes. We may even have trouble perceiving individual images. The process of comprehension, usually taken for granted, is suddenly laid bare. (Peterson 1994, 13)

But art appreciators more generally engage in a kind of problem-solving, when they ask themselves the fundamental questions: what kind of a work is this, and what sense do I make of it?

So we can think of viewers of paintings and films, music listeners, theatregoers and the operati—art appreciators in general—as, whatever else they are, problem-solvers. But given that the problem posed in the present essay concerns the freedom of the *artist*, our focus must fall on the activity of that figure, and the extent to which that activity can be construed as a kind of problem-solving. Drawing on the model of E. H. Gombrich in art history, Bordwell made such a perspective central to his understanding of the interplay between invention and convention, or what he termed “the dynamic of change and continuity”, in the history of filmmaking (Bordwell 1997, 155; for a wider exploration of the art-historical background, see Burnett 2008).

The most basic problem a filmmaker faces is: *how can I make this film?* or, *how do I get this project out of development hell?* Filmmakers, like all artists, as Paisley Livingston puts it, face the “very general problem of finding a way to make or do something the experience of which can be intrinsically worthwhile or rewarding for some audience” (forthcoming 2025, 2). But this most fundamental—and if a filmmaker is to sustain a career, recurrent—problem breaks down into a multitude of more modest problems (what location should we use for this scene?), which in turn break down into yet smaller problems (given that we’re using this location for a night-time scene, should we shoot at night, or shoot day-for-night?). At this level, as Bordwell suggests, “[f]ilmmaking is an avalanche of such minute choices” (Bordwell 1997, 149). In this way the act of making a film can be modelled as a temporally-extended, large-scale problem which breaks down into ever-smaller problems, nested within one another. The dissection of the overarching problem into sub-problems is not just an artefact of the analysis of problem-solving, but an important aspect of the problem-solving process itself—making that process tractable.

The problem-solving perspective is thus a pragmatic account of artistic creation, emphasizing craft skills and the agency of artists, standing in contrast to both the *Romantic* and the *culturalist* perspectives. While the Romantic perspective focuses on the individual as a source of creativity (and in the ideal case, of genius), culturalism puts the stress at the collective, social level. In the latter case, little attention is paid to the realisation of social norms—including artistic norms—by individual actors, making it difficult to see how such norms are either invented, sustained, adjusted, or challenged in the course of their enactment. In the former case, little attention is paid to the context in which the artist finds him or herself, the array of existing traditions and norms with which and within which they must work. The problem-solving perspective steers a course between such Romanticism and culturalism, insisting that, on the

one hand, social norms of practice only come into being through the actions of individual agents, working to propose, consolidate, or change these norms, while on the other hand, the artist—no matter how exceptional the individual’s creative intelligence, and no matter how strong their impulse to break with convention and forge novel interventions—can only act within a socially-established, sustained, and sanctioned normative context.

Relatedly, the problem-solving account readily recognises the agency of both individuals (producers, directors, editors, screenwriters, and so on) and groups of various types and sizes (for example, the studios, the various film unions and guilds, the Motion Picture Association and its predecessors). We can thus reframe the goal of cross-cultural accessibility pursued by Hollywood as a problem recognised, tackled, and (more or less successfully) solved by these various agents, acting in co-ordination.

Beyond their size and scale, we can distinguish the problems faced by artists along various other axes. The problems confronted by artists have something like an “arc” of development: the problem of *inception*, of giving a new idea the first outline of definition, is very different to the problem of *development*, of giving the idea substance and structure, and to the problem of *completion*, of driving a now developed idea towards a finished state. Some problems will be *foreseen*, others *unforeseen*; things change. Bordwell suggests that many problems will be *extrinsic* to a particular project, in the sense that they will arise in relation to all projects of a certain type. A narrative filmmaker, at least one working within mainstream norms, will thus always face the following questions:

How do you ensure that viewers recognize the main characters on each appearance? How do you delineate cause and effect in unambiguous ways? How do you portray psychological states that propel the action? How do you draw the viewer’s attention to the most important events in a shot or scene? (Bordwell 1997, 151)

Filmmakers and artists also face *intrinsic* problems, that is, problems of their own making—problems which are unique to each particular project. Patrick Keating analyses a montage sequence in *The Magnificent Ambersons* (1942) as a solution to a specific problem posed by the film: how do you represent the absence of the past? (Keating 2018, 01:35). Every project will give rise to such problems, but in some cases, the artist showcases a problem, setting it up as a challenge and putting it at the very centre of the work. Hitchcock liked to work in this way: how can

you make a film where the action is restricted to a few square metres? Well, you make *Lifeboat* (1944). How can you make a feature-length film which appears to do without editing, relying wholly on staging and camera movement? You make *Rope* (1948). Such experiments in form have been embraced by many contemporary filmmakers: in *Buried* (2010), Rodrigo Cortés pushed the conceit of *Lifeboat* still further by restricting the action to the space of a coffin buried underground; in *Memento* (2000), Christopher Nolan—and before him, playwright Harold Pinter, with *Betrayal* (1978)—created for themselves the challenge of narrating a story backwards.

4. Constraints again: Chosen, imposed and invented

Artists and filmmakers like Hitchcock, Pinter, and Nolan, who “put formal problems at the center of their work” (Bordwell 2006, 74), allow us to see most clearly how the problems and challenges faced by artists work as *productive* constraints. Elster (2000) provides a taxonomy and analysis of such constraints. From the moment that Damien Chazelle envisaged *La La Land* (2016) as a musical, he elected to work with(in) the conventions of the musical—a quite different set of conventions from those evident in a drama about music, like Chazelle’s breakthrough film *Whiplash* (2016). In such cases, the constraints are, in Elster’s terminology, *chosen* by the artist.

By contrast, many constraints will be *imposed* on the artist by factors over which they exercise little or no control: physical and technological (Arnheim’s “conditions of reproduction”, for example, like the limits on the length of an individual shot at the time that Hitchcock made *Rope*, or the limits to the possible height of buildings given the materials and building methods available at any given historical moment); financial (limits to the budget available); legal-regulatory (consider the constraints imposed by the Production Code on Hollywood filmmaking of the studio era); and perhaps also perceptual-psychological (on the view of some music theorists, serial composition runs up against the limits of what humans are capable of discerning through listening, and so in some sense fails as an artistic project—see Raffman 2003). Unforeseen accidents will also impose hard constraints: when Heath Ledger died during the shooting of *The Imaginarium of Dr Parnassus* (2009), the project could not be completed as planned—a situation driving a renewed burst of creativity as Terry Gilliam and his team worked to find a way of reimagining parts of the film in order to bring it to completion; that is to say, to find a solution to the problem of Ledger’s untimely and unanticipated death.

Consider also the following two examples of creativity engendered by constraint from the world of rock music. In 1980 Peter Gabriel released his third solo album—titled *Peter Gabriel*, just like his first two albums. The album marked a notable shift in the overall sonic and musical texture of Gabriel's work, arising from a number of changes in instrumentation, composition, and production style. Salient among these changes was a new approach to drumming, in which use of the hi-hat and cymbals was eschewed. Here we encounter a third type of constraint noted by Elster: those constraints *invented* by artists. Such constraints are chosen by artists, though not chosen from an array of existing conventions; rather they are brought into being—precisely, invented—through the specification of the constraint itself. The decision to dispense with hi-hat and cymbals went hand in hand with—perhaps causing, perhaps being caused by, probably both—the adoption of various African drumming techniques. The resulting album has a quite different rhythmic and textural character to anything produced by Gabriel as a solo artist previously, or by Genesis (the band through whom Gabriel had found fame), or pretty much anyone in the history of rock music up to that point.

Gabriel's fourth album, once again called *Peter Gabriel*, continued with the percussion experiment. With his fifth album the (self-imposed, invented) cymbal embargo was lifted, Manu Katché's hi-hat driven rhythms playing a key role on several tracks. As Elster notes:

The creation of a work of art can (...) be envisaged as a two-step process: *choice of constraints* followed by *choice within constraints*. The interplay and back-and-forth between these two choices is a central feature of artistic creation, in the sense that choices made within the constraints may induce the artist to go back and revise the constraints themselves. (Elster 2000, 176)

Thus, having explored the percussive space cleared by removing cymbals and hi-hat—the choices available *within* the space created by this constraint—Gabriel changes the choice *of* constraints.

And so also with the self-imposed prohibition on the provision of unique linguistic identifiers for his albums, this fifth solo album being named *So* (albeit conceived by Gabriel as a kind of anti-title, adopted as a concession to record company pressure for a unique album title—in Elster's terms, an external, imposed constraint). Having explored the space of possibilities created by forging a distinctive album title—one

effect of which is to push fans to identify the first four Peter Gabriel albums via their distinct visual designs, the albums becoming informally known as “Car”, “Scratch”, “Melt”, and “Security” [see **Figure 4**]—Gabriel relaxes this particular invented constraint from this point on in his career (his subsequent albums bearing the titles *Us*, *Up*, *Scratch My Back*, *New Blood*, and *I/O*). Thus, in respect of both Gabriel’s approach to album titles and percussion, with *So* he switches back from idiosyncratic, invented constraints to much more familiar conventional constraints.

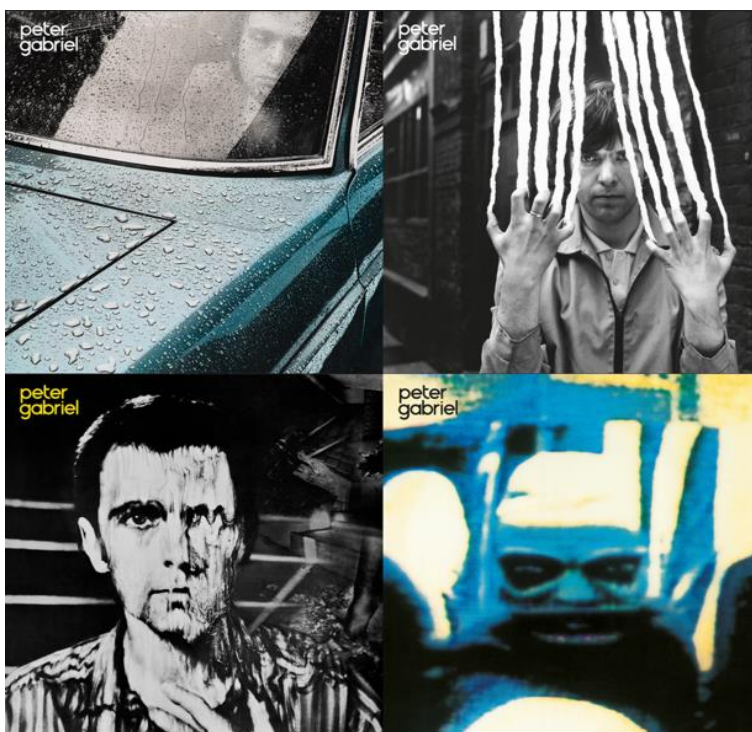


Figure 4: Cover designs for Peter Gabriel’s first four solo albums

Cut forward twenty-five years and we find Nick Cave, a musician one generation on from Gabriel, performing a similar creative manoeuvre. After almost a quarter-century of musical activity under the guise of Nick Cave and the Bad Seeds, along with four other members of the Bad Seeds, Cave formed Grinderman, a kind of allotrope of the Bad Seeds. While in the Bad Seeds Cave sings and plays piano and keyboards, in Grinderman he sings and plays guitar. As with Gabriel’s games with percussion, and notwithstanding the many continuities between the Bad Seeds and Grinderman, Cave’s switch from piano to guitar makes a significant difference. By adopting a new set of constraints and working

within a different set of parameters—those of the guitar, which Cave played with less conventional facility and expertise than the piano, though with plenty of gusto—Cave jolted himself out of certain routines associated with and reinforced by his role as a pianist. Cave describes the strategy in relation to the band as a whole:

When I went in to do [*Grinderman* (2007)], I deliberately made it difficult for the band, in that I took away the instruments that they normally play, and they played other instruments. Mick Harvey didn't play electric guitar, but played acoustic guitar; Warren [Ellis], who's traditionally played violin, didn't bring his violin to the studio; there was no piano, which is a huge change for the Bad Seeds, I mean there was very little piano (...) and that is my instrument (...). No matter what we did it had to sort of sound different because of that (...) the guitar is something you kind of embrace, and the piano you kinda—when you play it, you sort of push it away. It feels very different (...). And you kinda got the history of rock 'n' roll in your hands, with a guitar. Suddenly it's like, "oh—that's why rock 'n' roll is the way it is". (Gross 2008)

The new parameters of the guitar come with different affordances, of both a negative and a positive variety: a guitar, like a piano and any given instrument, enables some things and disables others.

Earlier I noted Arnheim's stress on the importance of the technological constraints of a medium in spurring creativity: at the time that Arnheim first composed his treatise on the art of film, it was necessary for a filmmaker to render a world of colour and sound in the form of black and white moving images unaccompanied by recorded synchronous sounds. In artforms dependent on advanced technology, like film and recorded music, the shaping force of technology is often evident: the unique sound of Gabriel's third and fourth albums is a case in point, as is the impact of digital technologies on contemporary filmmaking in general. Very commonly our folk axiom is applied in relation to technological developments: the more options a technology affords us, the more it enhances our creative freedom, so the story goes. But again, this is at best just half the story. In a discussion with Jarvis Cocker, Paul McCartney stressed the creative and aesthetic benefits of the very tight recording schedule The Beatles had to work within during their early years (Cocker and McCartney 2018). Cocker jests that in the time it took The Beatles to complete the recording of a song, nowadays the sound of the snare drum might be set; commenting on the discussion in same spirit, Dan Barratt

(2024) has suggested that, if the band, and their producer George Martin, had had at their disposal the options and flexibility of contemporary recording technology, The Beatles might still be working on *A Hard Day's Night* (1964).

McCartney's point concerns the benefits of temporal constraints in relation to the production of art: the right amount of temporal pressure during the making of a work can be creatively and aesthetically fertile. But note that the temporal constraints imposed by the type of work being produced may be similarly beneficial. Geoffrey O'Brien remarks that the tv sitcom "[*Seinfeld*'s] best episodes [twenty-two minutes long, excluding adverts] feel like feature films, and indeed have busier narratives than most features (...). Comic opportunities that most shows would milk are tossed off in a line or two. The tension and density of working against the time constraint is a reminder of how fruitful such constraints can be. If Count Basie had not been limited to the duration of a 78 rpm record, would we have the astonishing compression of "'Every Tub' (three minutes, fourteen seconds) or 'Jumpin' at the Woodside' (three minutes, eight seconds)?" (1997, 13-14; cited by Elster 2000, 211 n92). In case of 78 rpm records, the tight temporal constraint arises from the limits of recording technology at the time; in the case of *Seinfeld* (1989-98) the relative brevity of the episodes arises from institutional norms relating to television broadcast windows and advertising practices in the US. So while the character and origin of the durational limits of the recordings and the show differ, what they share (on O'Brien's argument) is the positive potential of temporal brevity built into the form of the work.

Some trends in contemporary music are precisely a reaction against the omniflexible character of digital recording: "When computers made home recording mainstream at the turn of the century, musicians could experiment with the cut-and-mix aesthetic to a vastly greater degree. Before then, the cost of recording and producing professional-sounding music had been exorbitant for most people; afterwards, a diligent one-person band could produce a lot of music for almost no expense. This helps explain why dub became a subcultural phenomenon in fin-de-siècle Germany. When almost anything seems feasible, self-imposed limits have a powerful allure" (Bertsch 2024). In Elster's terms, dub musicians invent their own constraints, carving out a much more delimited space of sonic possibility than the underlying technology allows.

5. Four ways to solve an artistic problem

Elaborating a schema devised by Noël Carroll—in turn suggested by the work of Arthur Danto—Bordwell suggests that there are broadly four ways in which artists can solve the problems they face, constituting a spectrum of options: through the *replication, amplification and revision, synthesis, or rejection* of the artistic norms that they inherit.⁶ An artist can play it safe by working to observe the norms of an existing framework; a love song, a horror movie, a heist film. Or they can take those norms and intensify or extend them, without fundamentally changing the underlying form or challenging an appreciator's ability to latch onto that form; in modest scope, such revision is probably the most common type of artistic production. Moving further along the spectrum, artists may synthesize existing forms, as with genre hybrids: consider the blending of gangster and domestic drama in *The Sopranos* (1999-2007) or the fusion of new wave and dance elements in the music of Talking Heads. At the opposite pole of the spectrum we reach the outright rejection of a set of norms, a strategy requiring either that the artist appeals to some other set of norms, or in the very act of rejection, instantiates a new possibility; while revision involves a difference in degree, rejection involves a difference in kind. And if the new forms generated through either synthesis or rejection persist, they themselves become targets for replication, revision, synthesis, or rejection, and so the process iterates.

Bordwell's schema can be understood, then, as a way of concretizing the dialectic between convention and invention. Artists always create against a background of norms and established conventions, but always have the scope for invention. As Brian Boyd notes, “we *need* to imitate in order to innovate” (2009, 122). Note, however, that although there is clearly a sense in which, as we move from replication towards rejection, the artist exercises their capacity for invention to a greater and greater degree, works in all parts of the spectrum can be realised with more or less skill and care. That is why it is possible to find and praise excellence in an artwork which works from and stays within—in Bordwell's terms, one which replicates—an established set of norms. Many well-crafted genre works fit into this category: a thriller or horror film might engineer its thrills and frights exceedingly well while working entirely within the relevant genre conventions.

⁶ Carroll's is a three-part schema inspired by Danto's philosophical account of the “artworld” (Danto 1964): “something is a work of art in a classificatory sense only if (...) it can be appreciated as a repetition, amplification or repudiation of prior traditions of the artworld” (1996b, 382).

For this reason, as James Ackerman notes, it is important to distinguish between *novelty* (in my terms, inventiveness) and *quality* (the realisation of value) (Ackerman 1991, 17). Something can be novel without being of any great value: Kant held that nonsense expressions exemplify this possibility (Kant 1952, §46, 168). Boyd concurs: “Most novelty matters little. Every move we make is in some sense new, yet unlikely to be of lasting moment” (2009, 120). And as we’ve just seen, something can be excellent in a given domain without being significantly novel or inventive: the beautifully wrought Adirondak chair that graces the porch of my rented summer house is a fine example of a now-generic design dating back to the early 1900s, conforming entirely to our expectations of its form. With this point in mind, many theorists of creativity define the latter as the combination of novelty with quality; they argue, that is, that we should only bestow the honorific “creative” on an action or object which embodies both originality and value (“quality”, in Ackerman’s terms). Berys Gaut argues that the originality of a creative act or product is quite distinct from its value, on the basis of the Kantian insight noted above: that there can be original (and in that sense “creative”) nonsense—and nonsense, by definition, can’t be valuable. “So we think of Picasso and Braque as exhibiting creativity”, Gaut concludes, “partly because of the originality of their Cubist paintings, but mainly because that originality was exhibited in paintings which, *considered apart from their originality*, have considerable artistic merit” (2003, 150, my emphasis).

But this analysis may not be strict enough, to the extent that we think of the novelty and the value of a creative item as bound up with one another. If the original properties of a creative entity are allowed to float free of its valuable properties, then the Kantian worry looms again: a spoonful of original nonsense allows us to regard an act or object as creative, but because these properties are separate from those that make it valuable, our sense that it is authentically creative is diminished. Putting the point in positive terms, we may say that i) a work exhibits creativity through its realization of original, (artistically) valuable properties, and ii) that the more the originality and the value we attribute to a creative work inhere in the same properties of that work, the more we will be inclined to judge it creatively excellent. The most obvious case where this holds is when we claim of an artwork that it is *insightful* (or any cognate claim that an artwork is epistemically valuable), for the claim entails that the work is not only cognitively substantial, but in some way novel (otherwise it could not be said to offer an insight).⁷

⁷ At least this is the case for strong versions of the “insight” claim, where an artwork is held to make an original cognitive contribution, in part by virtue of its unique form and medium-specific

For a further example, we can return to Peter Gabriel's early solo career. We might say that the artistic excellence of Gabriel's third and fourth albums lies (among other things) in their atmospheric, cymbal-less and tom-tom heavy percussive soundscape—in those very features of the work that are most original and arise from the creative agency of Gabriel and his collaborators. We would then want to say that the reason we attribute such a high level of creativity to these recordings, or to the Cubist paintings of Picasso and Braque, is that their originality, so to speak, saturates them; we cannot easily disentangle their great originality from whatever other artistic virtues they may possess. So the *mark of creativity* is the convergence and integration, and not merely the co-presence, of originality and other forms of value (Smith 2024b).

6. Conclusion: The sweet zone of creativity

With the various examples sketched above in mind, let's return to the underlying idea: that far from diminishing the artist's freedom, constraints function to increase and realise that freedom. But exactly how and in what sense do constraints work to increase the freedom of the artist? Elster and Bordwell contend that the constraints function to *focus* the attention and creativity of the artist. By providing focus, the artist is liberated from the oppression of the blank page and the empty canvas, enabled to move more readily from nothing to something, addressing the problem of inception in particular. Perhaps it isn't, after all, exactly freedom that the constraints create, so much as encouraging and facilitating action by limiting choice. As Livingston puts it, precommitment to a set of constraints “forestalls aesthetic indifference [and] directionless scrutiny” (Livingston 2009a, 140). The constraints help to free us from creative inertia, from the indecision of Buridan's ass, paralysed by too many equally tempting options. But the constraints must not be “too tight”, lest the artist be left with too little room for creative manoeuvre. Elster (2000, 210, 234-40) suggests that Stendhal backed himself into a creative corner, in just this way, with his unfinished novel *Lucien Leuwen*—unfinished, on Elster's account, because Stendhal had no viable narrative moves (given his rigid views concerning romantic conduct) available to him.

properties. For a particular case, consider the “bold thesis” in the “film-as-philosophy” debate, discussed by Livingston (2009b), Smith (2010), and Smuts (2009).

Constraints help the artist to focus, and in that way address the problem of inception. But they also assist with the problems of development and completion. Once the artist settles on a set of constraints, the psychological reward of realising a successful work within these constraints beckons. “[T]he overcoming of obstacles is in itself a liberating activity”, writes Marx (quoted by Elster 2000, 178). It’s instructive to return to Stravinsky (1947) at this point in the proceedings: “My freedom will be so much the greater and more meaningful the more narrowly I limit my field of action and the more I surround myself with obstacles”. Like Marx, Stravinsky recognises the importance of obstacles to push against and overcome in realising his freedom—his scope for artistic agency—and making it *meaningful*. All the theoretical artistic freedom in the world will be meaningless—without value—unless it can be cashed in for purposeful action in relation to the realisation of an artistic project. The constraints provide the framework for such meaningful, focussed action, placing the artist in *the space of creativity*, neither stifled by too slender a set of options, nor incapacitated by a surfeit of them.

If this is correct, it is not the case that artistic freedom continuously increases with increasing constraint, as Stravinsky claims, but rather there is something like a sweet spot—or a sweet zone—of creativity poised between the poles of procrastination (wrought by too many options) and suffocation (produced by too few). The relationship between creativity and constraint is thus not linear, but rather takes the form of a bell curve. This is the grain of truth in the folk axiom: our artistic freedom does increase with fewer constraints, *up to a point*—that point being the sweet spot, or zone, of artistic agency; in other words, the space of creativity. To return to the example of The Beatles: while McCartney notes that too many options can hamper creativity, it is important to remember the creative ferment engendered by advances in recording technology in their later albums, particularly from *Revolver* (1966) onwards. The new technical options and greater flexibility—the diminished constraints—were creatively enabling.

Stravinsky’s claim is, then, a corrective overstatement, in the sense that the *inverse* of the axiom is not true: it is not the case that artistic creativity simply increases with more constraints. Stravinsky’s overcorrection functions as an instance of a “felicitous falsehood”—a claim which is literally false, but takes us closer to the truth. In Catherine Elgin’s words, such a claim is “an inaccurate representation whose inaccuracy does not undermine its epistemic function” (2017, 3).

Although this analysis of the nature of creative freedom has particular purchase in relation to the arts, its relevance extends well beyond this context.⁸ Consider the case of sports and games. Games by definition are structured by invented rules, which create a space within which players seek excellence (of strategy, physical or cognitive pre-eminence, and so on—paradigmatically, by winning). The constraints provided by a set of rules and the “game space of possibility” are necessary for players to realise creativity and excellence. These rules can be and often are tweaked in order to optimize the dynamics of games, and increase the affordances for excellence—to strike the right balance between constraint and freedom, necessity and flexibility. As Elster notes: “there is a ceaseless adjustment of rules so as to make athletes and players bound neither too tightly nor too loosely” (2000, 281). Or as Ken Bruce, co-creator of the BBC radio quiz game *PopMaster*, puts it:

[N]obody wants to listen to a quiz, where people are [always] getting the wrong answer and scoring zero points. You want people to score well, but you don't want them to know everything. (O'Connell, 2024)

Our item of folk wisdom is, then, a half-truth: artists need a space of possibility within which they can exercise their agency. If that space is too narrowly circumscribed or entirely closed down, the artist will be paralysed; that hazard is recognised by our folk understanding of artistic freedom and creativity. What I hope to have made salient here is the other side of the story, which that same folk theory obscures (as Stravinsky saw so clearly): working within constraints—chosen and imposed, conventional and invented—provides an essential complementary ingredient in creative agency, facilitating and making meaningful the artist's “theoretic freedom” (Stravinsky 1947, 64).

Acknowledgments

My thanks to Dan Barratt (in particular for suggesting the connection between the “sweet zone” of creativity and bell curve distribution, and for pointing me to the dialogue between Cocker and McCartney), and Ted Nannicelli (especially for directing me to the Terry Gross interview with Nick Cave). My gratitude also to the two anonymous journal reviewers for helpful feedback, and to Iris Vidmar Jovanović for her keen insights and endless patience.

⁸ Elster's (2000) book-length treatment ranges across psychology, politics, and aesthetics.

REFERENCES

- Ackerman, James S. 1991. "Style." In *Distance Points: Essays in Theory and Renaissance Art and Architecture*, 3-22. Cambridge, MA: MIT Press.
- Arnheim, Rudolf. 1983. *Film as Art*. Reissue with new preface. London: Faber and Faber.
- Barratt, Dan. 2024. "Film Theory and the Philosophy of Digitalization." Paper presented at the Society for Cognitive Studies of the Moving Image annual conference, Budapest, June 7, 2024.
- Baxandall, Michael. 1988. *Painting and Experience in Fifteenth-Century Italy*. 2nd Edition. Oxford: Oxford University Press.
- Berliner, Todd. 2017. *Hollywood Aesthetic: Pleasure in American Cinema*. New York: Oxford University Press.
- Bertsch, Charlie. 2024. "Inner City Sounds: Damaged and Damaged Versions, by Ghost Dubs." *The Battleground*. Accessed September 17, 2024.
https://open.substack.com/pub/thebattleground/p/inner-city-sounds?utm_campaign=post&utm_medium=web
- Bordwell, David. 1997. *On the History of Film Style*. Cambridge, MA: Harvard University Press.
- . 2005. *Figures Traced in Light: On Cinematic Staging*. Berkeley: University of California Press.
- . 2006. *The Way Hollywood Tells It: Story and Style in Modern Movies*. Berkeley: University of California Press.
- . 2008. "Convention, Construction, and Cinematic Vision (with Afterword)." In *Poetics of Cinema*, 57-82. New York: Routledge.
- Bordwell, David, Janet Staiger, and Kristin Thompson. 1985. *The Classical Hollywood Cinema: Film Style and Mode of Production to 1960*. London: Routledge.
- Boyd, Brian. 2009. *On the Origin of Stories: Evolution, Cognition, and Fiction*. Cambridge, MA: Harvard University Press.
- Burnett, Colin. 2008. "A New Look at the Concept of Style in Film: The Origins and Development of the Problem-Solution Model." *New Review of Film and Television Studies* 6 (2): 127-149.
- Carroll, Noël. 1996a. "A Note on Film Metaphor." In *Theorizing the Moving Image*, 212-23. Cambridge: Cambridge University Press.
- . 1996b. "Film History and Film Theory: An Outline for an Institutional Theory of Film." In *Theorizing the Moving Image*, 375-91. Cambridge: Cambridge University Press.
- . 2003. "Forget the Medium!". In *Engaging the Moving Image*, 1-9. New Haven: Yale University Press.
- . 2006. "Engaging Critics." *Film Studies: An International Review* 8: 161-169. Special issue on Film, Cognition and Emotion.

- Cocker, Jarvis and Paul McCartney. 2018. "In Conversation at The Liverpool Institute for the Performing Arts." Accessed on September 18, 2024. *Paul McCartney*. August 15, 2018. <https://www.youtube.com/watch?v=-gxdRjRqcZQ>
- Danto, Arthur. 1964. "The Artworld." *Journal of Philosophy* 61 (19): 571-84.
- Elgin, Catherine Z. 2017. *True Enough*. Cambridge, MA: MIT Press.
- Elster, Jon. 2000. *Ulysses Unbound: Studies in Rationality, Precommitment, and Constraints*. Cambridge: Cambridge University Press.
- Gaut, Berys. 2003. "Creativity and Imagination." In *The Creation of Art: New Essays in Philosophical Aesthetics*, edited by B. Gaut and P. Livingston, 148-174. Cambridge: Cambridge University Press.
- Gross, Terry. 2008. "Interview with Nick Cave." *Fresh Air, WHYY*. April 28, 2008. <https://freshairarchive.org/segments/nick-cave-digs-himself-singular-niche>
- Kant, Immanuel. 1952 (1790). *The Critique of Judgement*. Trans. James Creed Meredith. Oxford: Clarendon Press.
- Keating, Patrick. 2010. *Hollywood Lighting from the Silent Era to Film Noir*. New York: Columbia University Press.
- . 2018. "The Strange Streets of a Strange City: The Ambersons Montage." *NECSUS: European Journal of Media Studies*. July 11, 2008. <https://necsus-ejms.org/the-strange-streets-of-a-strange-city-the-ambersons-montage/>
- Livingston, Paisley. 2009a. "Poincaré's 'Delicate Sieve': On Creativity and Constraints in the Arts." In *The Idea of Creativity*, edited by Michael Krausz, Denis Dutton, and Karen Bardsley, 127-146. Leiden: Brill.
- . 2009b. *Cinema, Philosophy, Bergman: On Film as Philosophy*. Oxford: Oxford University Press.
- . Forthcoming 2025. "The Problem/Solution Model." In *Observing Film Art: On the Theory, History, and Criticism of David Bordwell*, eds. Charlie Keil and Murray Smith. New Brunswick, NJ: Rutgers University Press.
- Lopes, Dominic McIver. 2024. "Cultures and Values." In *The Geography of Taste*, edited by Dominic McIver Lopes, Samantha Matherne, Mohan Matthen, and Bence Nanay, 71-111. New York: Oxford University Press.
- MacDonald, Scott. 2009. "Putting All Your Eggs in One Basket: The Single-Shot Film." In *Adventures of Perception: Cinema as Exploration—Essays/Interviews*. Berkeley: University of California Press. 240-54.

- Nannicelli, Ted. 2017. "Contingent Universals." *Literary Universals Project*. Edited by Patrick Colm Hogan, Vito Evola, and Nigel Fabb.
<https://literary-universals.uconn.edu/2017/08/19/contingent-universals/>. Accessed September 16, 2024.
- O'Brien, Geoffrey. 1997. "Sein of the Times." *New York Review of Books*. August 14: 12-14.
- O'Connell, Paddy. 2024. "Interview with Ken Bruce". *Broadcasting House*. September 15, 2024.
- Peterson, James. 1994. *Dreams of Chaos, Visions of Order: Understanding the American Avant-garde Cinema*. Detroit: Wayne State University Press.
- Raffman, Diana. 2003. "Is Twelve-Tone Music Artistically Defective?". *Midwest Studies in Philosophy* 27: 69-87.
- Smerconish, Michael. 2024. "Interview with David Axelrod." Smerconish, CNN, 1 June 2024.
- Smith, Murray. 2006. "My Dinner with Noël; Or, Can We Forget the Medium?" *Film Studies: An International Review* 8: 140-8.
- . 2010. "Review of Livingston, Cinema, Bergman, Philosophy." *Notre Dame Philosophical Review*. May 19, 2010. Accessed September 18, 2024. <https://ndpr.nd.edu/reviews/cinema-philosophy-bergman-on-film-as-philosophy/>
- . Forthcoming 2024a. "Reflective Equilibrium, Aesthetic Experience, and Aesthetic Judgement." *Synthese*.
- . Forthcoming 2024b. "Imagination and Creativity in Film." In *The Oxford Handbook of Imagination and Creativity*, edited by Amy Kind and Julia Langkau. New York: Oxford University Press.
- Smuts, Aaron. 2009. "Film as Philosophy: In Defence of a Bold Thesis." *Journal of Aesthetics and Art Criticism* 67 (3): 409-20.
- Stravinsky, Igor. 1947. *Poetics of Music in the Form of Six Lessons*. Trans. Arthur Knodel and Ingold Dahl. Cambridge, MA: Harvard University Press.
- Truffaut, François. 1963-4. "Sept Hommes à Débattre." *Cahiers du Cinéma* 150-1: 20.
- Walton, Kendall. 1970. "Categories of Art." *The Philosophical Review* 79 (3): 334-367.

AESTHETIC EDUCATION: A PERCEPTUAL- COGNITIVE MODEL

Ted Nannicelli¹ 

¹The University of Queensland, Australia

Original scientific paper – Received: 12/04/2024 Accepted: 10/05/2024

ABSTRACT

Here is a puzzle about aesthetic education. In a variety of contexts, we commit significant time, energy, and resources to aesthetic education. We teach (and in many cases publicly subsidize) university courses and degrees that have aesthetic education as their primary aim; we also invest public resources into museums, including enrichment programs that are also designed to afford aesthetic education. It would seem that if our commitment to aesthetic education is rational, then aesthetic appreciation is something that can be done better or worse. However, we also, in a variety of contexts (oddly enough, some of them being the same sorts of contexts that are designed to abet aesthetic education), act as if it is true that there is no disputing taste. We may try to persuade students to come around to particular judgments, but we do not penalize students for judging one way or another.

The aim of this paper is to dissolve the apparent puzzle of aesthetic education by clarifying its aims and advancing a conception of it that deemphasizes the role of taste. I claim that, pace “the default view of aesthetic education” (as I shall call it), the primary purpose of aesthetic education is not to educate taste. It is, rather, to facilitate the development of certain perceptual-cognitive capacities so as to enhance aesthetic experience and improve aesthetic appreciation. Thus, I call the view of aesthetic education advanced here “a perceptual-cognitive model”.

Keywords: aesthetic education; aesthetic cognition; aesthetic normativity; aesthetic appreciation; taste.

1. Introduction

Here is a puzzle about aesthetic education. In a variety of contexts, we commit significant time, energy, and resources to aesthetic education. We teach (and in many cases publicly subsidize) university courses and degrees that have aesthetic education as their primary aim; we also invest public resources into museums, including enrichment programs that are also designed to afford aesthetic education. It would seem that if our commitment to aesthetic education is rational, then aesthetic appreciation is something that can be done better or worse. However, we also, in a variety of contexts (oddly enough, some of them being the *same* sorts of contexts that are designed to abet aesthetic education), act as if it is true that there is no disputing taste. We do not, as educators, tell students they are wrong or give them low marks if they remain unimpressed by Woolf's novels, Ozu's films, Monet's paintings, Beethoven's music, and so forth. We may try to persuade students to come around to particular judgments, but we do not penalize students for judging one way or another.

On the face of it, this puzzle about aesthetic education appears to be a manifestation of the general puzzle about aesthetic normativity in a specific context. That is, you might think that the puzzle of aesthetic education boils down to a question about whether aesthetic judgments and aesthetic value are objective or subjective. Thus, Nick McAdoo (1987) explicitly identifies the puzzle of aesthetic education as an instance of Kant's "antinomy of taste", according to which pure judgments of taste are both subjective *and* universal. According to McAdoo, "The main impediment to [aesthetic education] is [that] 'appreciation' is characteristically understood not only in an objective, but also in a subjective sense" (1987, 307). A slightly different way of putting the dilemma is this: On the one hand, it's easy to make sense of aesthetic education if aesthetic judgments are objective; the question simply becomes how they could be objective—what the source of normativity is. On the other hand, if aesthetic judgments are subjective, the question of normativity dissolves but then it is hard to understand aesthetic education as having a rational basis. For this reason, it may seem, as Alan H. Goldman puts it, "The issue of requiring or even encouraging aesthetic education is tied also to the question of whether certain sorts of taste in art are objectively better than others" (1990, 105).

This common characterization of the puzzle of aesthetic education has affected different disciplines in different ways, but it is fair to say that one upshot is that some humanistic disciplines regard aesthetic education as incoherent at best (because the antinomy cannot be resolved) or even pernicious (because the antinomy is resolved by acknowledging that the apparent objectivity of aesthetic judgments is illusory). Scholars in media studies (broadly conceived

to include film and television studies), to take one example, often define their pedagogical project in opposition to aesthetic education. For example, in the 5th edition of a textbook that bills itself as “the foremost guide to television studies (...) for over two decades”, the author tells readers, “*Television [Visual Storytelling and Screen Culture]* does not attempt to teach taste or aesthetics. It is less concerned with evaluation than interpretation. It resists asking, ‘Is *The Bachelor* great art?’ Instead, it poses the question, ‘What meanings does *The Bachelor* signify and how does it do so?’” (Butler 2018, ix). So, the puzzle of aesthetic education, characterized as an instance of the antinomy of taste, raises both a seemingly intractable philosophical question and has practical implications for educators in the humanities.

The aim of this paper is to dissolve the apparent puzzle of aesthetic education by clarifying its aims and advancing a conception of it that deemphasizes the role of taste. I claim that, *pace* “the default view of aesthetic education” (as I shall call it), the primary purpose of aesthetic education is *not* to educate taste. It is, rather, to facilitate the development of certain perceptual-cognitive capacities so as to enhance aesthetic experience and improve aesthetic appreciation. Thus, I call the view of aesthetic education advanced here “a perceptual-cognitive model”.

Let me briefly say what I mean by “enhance aesthetic experience” and “improve aesthetic appreciation”. For reasons that should become clear as I defend the perceptual-cognitive model of aesthetic education against the default model, I take a deflationary view of both aesthetic experience and aesthetic appreciation. By “enhance aesthetic experience” and “improve aesthetic appreciation”, I mean, basically, helping students to more closely attend to and accurately identify the artistically relevant properties (and their relations) of artworks. The account of aesthetic experience underlying this idea is “the content account” developed by Noël Carroll (2010, 2023). According to his most recent statement of the content account, one

is having an aesthetic experience if they are attending with understanding to the formal properties and/or the expressive properties and/or the aesthetic properties of the artwork and/or the interactions thereof and/or to the reflexive relations between said properties and the viewer, listener, or reader. (Carroll 2023, 9)

Likewise, I take aesthetic appreciation not to be a matter of an emotive response or a valuing, but, rather, a matter of attending to and recognizing what the artist(s) have done in the work and how they have done it. On a capacious view of aesthetic experience, aesthetic appreciation might be one aspect of aesthetic experience. I am agnostic about whether we should regard

aesthetic appreciation as an aspect of aesthetic experience more broadly, but will use both terms in similar ways.

So, to summarise: On my view, aesthetic education is best understood on a perceptual-cognitive model, according to which its aim is to enhance aesthetic experience and improve aesthetic appreciation in deflationary, content-oriented conceptions of those terms. I hasten to add that this account of aesthetic education retains a normative dimension insofar as it (and the conceptions of aesthetic experience and aesthetic appreciation to which it is tied) hold that one can more or less accurately experience and/or appreciate artworks. For reasons I shall explain presently, I think this perceptual-cognitive model of aesthetic education has greater explanatory power than “the education of taste”, better meshes with what those in the business of aesthetic education actually do, and can help to avoid the sorts of commonly raised objections to aesthetic education qua the development of taste. In the spirit of Noël Carroll’s (2022) recent suggestion that we “forget taste” when it comes to the evaluation of art, my proposal is that we jettison the default view of aesthetic education as the development of taste in favour of an understanding of aesthetic education as aimed at developing perceptual-cognitive capacities that will improve aesthetic experience and aesthetic appreciation.

2. The default view of aesthetic education

It’s somewhat surprising that in contemporary, analytic philosophical aesthetics, relatively little has been written about aesthetic education (as a philosophical question rather than a practical matter)—particularly if we set aside work that is specifically focused on Friedrich Schiller’s *On the Aesthetic Education of Man* (e.g. Grossman 1968; Tauber 2006; Matherne and Riggle 2020, 2021; Leontiev 2023). In the work that has been done on aesthetic education, aestheticians have largely focused on one particular matter—the education of taste (McAdoo 1987; Goldman 1990; Fenner 2020). Moreover, outside of philosophical aesthetics, it is also common to conceive of aesthetic education as the education of taste. So, I will use the term “the default view of aesthetic education” to refer to the idea that the primary aim of aesthetic education is the education of taste.

Indeed, it is a tacit acceptance of the default view of aesthetic education as the education of taste that, I conjecture, underlies media studies’ skepticism (if not outright hostility) towards aesthetic education. Jeremy Butler’s textbook, cited above, is representative of the view that aesthetic education simply is not part of media studies’ overall pedagogical project. Elana Levine and Michael Z. Newman’s *Legitimizing Television* draws heavily upon the work of Pierre Bourdieu as well as John Fiske, the latter of whom wrote,

So, those of who are in media education have a responsibility to our students and to our society to first of all try and destroy this hierarchy of legitimation [of the media]. (Levine and Newman 2011, 153)

Levine and Newman describe their own project as “the work of analyzing patterns of taste judgment and classification [in order to] unmask misrecognitions of authentic and autonomous value, bringing to light their political and social functions” (2011, 7). For Levine and Newman, as for Bourdieu, Kant’s antinomy is dissolved by recognizing that the universality of aesthetic judgments is a façade that hides the process by which the taste of the dominant classes is naturalised and reified.

A similar position is expressed, albeit more opaquely, by television scholar Helen Piper (2016), who reads Kant via Terry Eagleton. According to Piper,

[A]esthetic judgment may be problematic not simply because professional criticism is an act of cultural power, but because any judgement (by whomsoever it is made) will lack ethical authority unless underpinned by consensual ideals. (Piper 2016, 167)

Piper characterizes Kant’s subjective universality (which she refers to as ‘universal subjectivity’) as “the old idea of aesthetic value as something that transcends space and time”, (2016, 180-181) and claims that it “makes it difficult, if not impossible, to reconcile the aesthetic (certainly the pure aesthetic) with the culturally specific” (170). In view of these considerations, Piper concludes, “such observations raise further questions about the pedagogic responsibilities for scholars of a medium that has such enormous national reach, community potential and (possibly unmet) duties of social recognition” (181).

In short, the dominant view in the field is, roughly, that whatever else media studies pedagogy might involve, it ought not involve aesthetic education. This is because aesthetic education, in turn, involves reifying hierarchies of taste—hierarchies that are not reflective of objectively better or worse taste, but merely of how cultural power gets exercised.

3. Taste

Given that the conception of aesthetic education as the education of taste is the default view of aesthetic education, I want to first try to dispel the concerns raised by media studies scholars and like-minded critics of taste. Despite my

advocacy for a perceptual-cognitive model, I realize that many people do think the primary aim of aesthetic education is the development of taste. However, when aesthetic educators concern themselves with development of taste, conceived as the exercise of aesthetic judgment, it is necessarily a secondary concern. It is necessarily a secondary concern because improving one's aesthetic judgments cannot happen in advance of improving the perceptual-cognitive skills that are central to aesthetic experience and aesthetic appreciation.

So, although my primary aim is to develop a perceptual-cognitive model of aesthetic education, let me begin by trying to assuage some concerns that media studies scholars have raised about taste.¹ The first thing to say is that media studies' reading Kant via Bourdieu and Eagleton results in a number of distortions of Kant's views. Most significantly, in this context, Levine and Newman and Piper follow Bourdieu and Eagleton, respectively, in conflating aesthetic judgments (as a class) with a more specific sort of aesthetic judgment—namely, the pure judgment of taste (see, e.g., Zangwill 2001). As readers of this journal are likely to recall, Kant proposes three kinds of aesthetic judgement—judgments of the agreeable, judgments of taste [or judgments of beauty], and judgments of the sublime. Furthermore, Kant claims that judgments of taste may be “pure” or “impure.” Pure judgments of taste are subjectively universal; they are judgements with which everyone *ought* to agree, involving disinterested satisfaction in an object that is not subsumed under a concept of an end or purpose. A judgment of taste is impure if it involves “charm and emotion”, or if it takes as its object “dependent” (also called “adherent”) beauty. For unlike in making a judgment of “free beauty”, in making a judgment of dependent beauty, one subsumes the object under a concept of an end or purpose. As Paul Guyer has put it, Kant holds that “in ‘pure’ judgements of taste our pleasure in beauty is a response only to the perceptible *form* of an object, not to any matter or content it may have” (2014, 435). Thus, Kant's examples of “free beauty” are natural objects (e.g. flowers and birds), as well as artifacts such as “designs *a la grecque*, foliage for borders, wallpapers (...) and all music without words” (ibid).

The features of aesthetic judgments that media studies scholars worry about—disinterestedness, universality, autonomy—feature in Kant's account of pure judgments of taste. But, somewhat remarkably, these aesthetics skeptics (as Sarah Cardwell (2013) has called them) overlook the fact that, on a plausible reading of Kant's own account, the representational arts possess dependent beauty and are thus not the objects of pure judgements of taste. But if

¹ I will not attempt to assuage Carroll's concerns about taste because I think they are all on the mark if we understand *taste* as he does—i.e. “a hedonic concept linked to pleasure” (2002, 4).

representational artworks, like films and television programs, are instances of dependent beauty rather than free beauty, then there is no obstacle to making (impure) aesthetic judgments of them that take into account their socio-historical contexts, cultural or political functions, and so forth (Zangwill 2001).

Moreover, and aside the question of how to interpret Kant, it is not clear why media studies scholars assume that aesthetic judgments take the form of Kantian pure judgments of taste. One is hard pressed to identify a single scholar or critic who, regardless of Kant's actual view, advocates the aesthetic appreciation of art on Kant's model of pure judgments of taste. Certainly, there are Kantian ideas that have influenced various norms of aesthetic criticism. For example, a commonplace of aesthetic criticism holds that one's *aesthetic* appreciation of an artwork requires one to bracket off various sorts of instrumental value the work might have (e.g. as a financial investment, as a status symbol, etc.) But this is hardly tantamount to making a pure judgment of taste and does not, in any case, enjoy universal acceptance amongst philosophers of art (see, e.g., Carroll 2010, 2023; Wolterstorff 2015). Consider another example: as mentioned earlier, there is a puzzling question about the normativity of aesthetic judgments. Kant characterizes this in terms of a claim to universality. But this is hardly the only way one might think about the normativity of aesthetic judgments, and it is far from clear that this conception of aesthetic normativity attends most, or even much, aesthetic criticism. And it seems even more implausible that this conception of aesthetic normativity underpins the efforts of those who do engage in aesthetic education precisely because it is not as if professors frequently insist that students ought to assent to their aesthetic judgments. In short, there are good reasons to think that if taste *is* involved in aesthetic education, then it is rather different from that which is involved in Kantian pure judgments of taste.

If the Kantian sense of taste is not what's at stake in aesthetic education, could the development of a more capacious sense of *taste* still be aesthetic education's primary aim? I take it this is a central part of Alan Goldman's proposal:

The main and final point of aesthetic education, then, is not knowledge in itself or knowledge required to make correct evaluations, but the preparation of the faculties for intensely meaningful and enjoyable experiences of artworks that can afford them to those trained to appreciate them. (Goldman 1990, 116)

Despite Goldman's attempt to shift focus from aesthetic "evaluation" to aesthetic "experience", his invocation of "enjoyable experiences" nevertheless recalls Kant's emphasis on the satisfaction or pleasure attending the judgment of taste. This is not an idiosyncratic feature of Goldman's proposal; even shorn

of any Kantian baggage—say, for example, on a Humean conception—the fundamental difficulty with *taste* in the context of aesthetic education is its essential emotive, subjective dimension.

As I hope to show in what follows, we would do better to conceive of aesthetic education as the development of perceptual-cognitive skills *without* invoking the experience of pleasure (or another emotive experience). I have not presented a knock-down argument to show that taste is never an aim of aesthetic education, let alone that it cannot be developed via aesthetic education. Rather, I have tried to highlight the fact that it is hard to make sense of aesthetic education as the development of taste in the absence of a resolution of Kant’s antinomy. While there is no shortage of proposed solutions, there is no widespread acceptance of any of them.

Despite this, we *do* invest significant time and money in aesthetic education, which ought to make us wonder if, in fact, there isn’t an entirely different account of aesthetic education that has just as much, if not more, explanatory power, but is more parsimonious and avoids the sorts of intractable philosophical debates that attend *taste*. I now turn to developing such an account.

4. Perceptual learning

A *prima facie* plausible conception of aesthetic education emphasizes the process of honing students’ perceptual capacities. From secondary school art courses to postgraduate art history seminars, students are encouraged to attend to particular qualities (and their relations) of artworks: Notice the tonal contrast between the foreground and background of the photograph; see how the use of empty space balances the frame; look at the pattern of brushstrokes; listen for the resolution of the suspension; and so forth. In these instances, educators seek to help students learn what to attend to and how. Moreover, in his discussion of aesthetic education, Goldman highlights the significance of being able to attend to certain relations amongst a work’s properties—“formal, referential (representational and expressive), and historical” (1990, 113). Noting that perceptual and cognitive processes sometimes overlap and intermingle (including in one way I am about to describe), I want to focus, for the moment, on perception, saving a discussion of the involvement of more thoroughly cognitive capacities in the apprehension of referential and (art-) historical properties for the next section.

Learning to direct one’s attention in particular ways is one of several ways in which perceptual learning can occur. Following Eleanor Gibson (1963) and Kevin Connolly (2019), whose recent account builds upon Gibson’s, I

understand *perceptual learning* to involve “long-term changes in perception that are the result of practice or experience” (Connolly 2019, 7). My claim is that a central component of aesthetic education is perceptual learning. In order to pump our intuitions about the plausibility of perceptual learning, philosophers have invoked contrasts between how a spoken language sounds to a native speaker versus a non-speaker, how the relationship between pieces on a chess board appears to a grandmaster versus an amateur, how wine tastes for a connoisseur versus a novice taster, and how Beethoven’s 9th Symphony sounds to a conductor versus a non-musician (see Connolly 2019 for discussion). Importantly, though, there is also a substantial body of evidence for perceptual learning from research in psychology and neuroscience (see e.g., Goldstone and Byrge 2015; Prettyman 2018), including some studies suggesting that artistic expertise affects the perception of artworks (e.g. Vogt and Magnussen 2007; for a review of the literature as it pertains to aesthetics, see Ransom 2022).

In order to see more clearly how perceptual learning might occur as a result of experience or training in artistic contexts, let me briefly outline three of the main mechanisms by which perceptual learning is thought to occur—attribute differentiation, unitization, and attentional tuning (Goldstone and Byrge 2015; Connolly 2019). In instances of attribute differentiation, one becomes able to discriminate amongst multiple properties or features that, prior to experience or training, one perceived as a single property or feature. Some well-known studies of attribute differentiation involve the perception of color. Most people experience a single, overall perception of color. However, evidence suggests that, with training, people can learn to discriminate between brightness (the amount of black or white added to a color) and saturation (the luminance of the color), and selectively attend to these individual features of colors (Goldstone and Byrge 2015, 823). Although there are no studies that have replicated these findings in the context of art appreciation, specifically, it is certainly plausible that the ability to differentiate between these two features of color is relevant for art appreciation and that developing this capacity is one of the aims towards which an educator might strive.

Consider another example of how attribute differentiation might work in an artistic context: In cinema, a “cut” is the joining together of two distinct pieces of film footage (or “shots”). The number of shots and cuts in a film can vary considerably, but a contemporary Hollywood feature film typically has well over 1000 shots.² Because the goal of the classical Hollywood style is to

² In a 2010 survey of 10 Hollywood films released every 5th year between 1935 and 2005, inclusive, James Cutting and colleagues (2010) found that the median number of shots was 1,132. Generally speaking, however, shots (and cuts) have become more plentiful as average shot-length has decreased over time. So, for example, the film in Cutting’s sample with the most shots was a 2005 release, which had 3,099 shots.

immerse the viewer in the world of the story, many cuts are designed to be imperceptible such that viewers perceive a continuous flow of story events across space and time. For this and other reasons having to do with the neurobiology of our visual system, most people tend to notice many fewer cuts than a film actually has (e.g. Magliano and Zacks 2011; Smith 2012; Heimann et al. 2017). But, if we want to appreciate how narrative events have been sequenced, or how a filmmaker cues particular responses by patterning cuts in specific ways, we need to be able to notice them—to differentiate between one shot and the next. And this is precisely one of the skills that film teachers aim to develop in their students.

The next mechanism of perceptual learning is unitization; it is the inverse of attribute differentiation. In cases of unitization, experience or training allows one to perceive multiple objects or features as a single percept. In describing unitization, several commentators note empirical evidence that suggests it occurs in medical settings where a professional with sufficient training learns to see a particular array of figures under a microscope or on a radiograph as, say, a malignant growth (e.g. Kranse et al. 2013; Seitz 2017). In an artistic context, we might think of the ability to hear distinct notes as comprising a particular interval or several distinct chords as a progression. Most of us have probably, at one time or another, had the experience of listening to a relatively unfamiliar type of music with someone who has had significant exposure to it or has expertise. Whereas I might hear “Coltrane changes” (chord substitutions for standard jazz chord progressions), my wife hears only a string of random chords. Or, perhaps when your Javanese friend hears a harmonic progression in a Javanese Gamelan piece, you only hear random notes. Fittingly, the process of learning to perceive intervals, chords, melody, rhythm, and other features of music is referred to as “ear training”. For the trained musician need not stop and think about whether the notes they hear constitute a tritone; the learning they have undergone is *perceptual* insofar as they simply hear a tritone (it is cognitive as well, insofar as they know it is a *tritone*, but what’s important here is that they perceive it as such).

Unitization is, thus, plausibly a common mechanism by which perceptual learning occurs as part of ear training and music education more broadly. Moreover, there is some empirical evidence supporting the intuition such examples are designed to pump—i.e., that musical training frequently results in the ability to actually hear sounds differently (see, e.g., Fujioka et al. 2004; Pantev and Herholz 2011).³ The question of direct evidence for musical

³ However, one commentator cautions, “the information this research has provided on auditory perceptual learning *per se* and its mechanisms is qualified by the fact that most training is multimodal and sensorimotor in nature, and by the relative paucity of experimental studies allowing the control of confounding variables” (Irvine 2018, 11).

training on perceptual learning notwithstanding, it seems undeniable that a central aim of musical education is, precisely, to facilitate perceptual learning.

The third mechanism by which perceptual learning is thought to occur is *attentional tuning* (sometimes called *attentional weighting*). In cases of attentional tuning, “perception becomes adapted to tasks and environments (...) by increasing the attention paid to perceptual features that are important, and/or by decreasing attention to irrelevant dimensions and features” (Goldstone and Byrge 2015, 819). A number of studies of visual perception in athletes have shown that highly experienced athletes, including soccer players and fencers, focus their visual attention on features of their opponents’ bodies in ways that allow them to perceive indications of where or how their opponents will move next. For example, one study has shown that, when defending, expert soccer players focus their attention on the hips of their opponents longer than non-experts (Williams and Davids 1998; for a review, see Connolly 2019).

Why think that such cases of attentional tuning are instances of perceptual learning? Several philosophers have marshalled evidence from empirical studies to argue that, as Ned Block puts it, “the phenomenal appearance of a thing depends on how much attention is allocated to it” (quoted in Connolly 2019, 89). Kevin Connolly points out that *where* one attends, as well as how much attention is allocated, also affects one’s perception of “low-level” features (2019, 89). Of course, one might still wonder whether such perceptual effects involve long-term changes in perception rather than the ability to exercise one’s attention in a specific way (like deploy a particular skill) when wanted.

Although there may not be a knock-down argument for the perceptual learning view, its plausibility can be bolstered by marshalling a few hypothetical examples which seem to be best explained by the perceptual learning thesis. Remaining with sports for a moment: The baseball great Ted Williams, who had 20/10 vision, reported that, in the batter’s box, he could see the individual stitches on a baseball and, thus, the part of the ball he wanted to strike with the bat. Given that a fastball travels from a pitcher’s hand to home plate in approximately half a second, it’s hard to conceive of Williams’s attentional weighting as deliberately or consciously engaged; more plausibly, he came to automatically see the ball in this particular way as a long-term effect of his (storied) training regimen.⁴ The rest of us mere mortals probably make use of attentional tuning in more mundane ways. After years of experience, we may be able to *just hear* that a song is in $\frac{3}{4}$ time or has a I-IV-V chord progression,

⁴ On automaticity and perceptual expertise, see Stokes 2021.

or *just see* checkmate in two moves, that a child is lying, that we need to apply the break to slow down in time to exit, or that a card player is bluffing.

Empirical evidence of attentional tuning in the context of artistic appreciation is somewhat limited, but several studies have shown that experts and non-experts attend to different features of artworks and for different durations (e.g. Nodine, Locher, and Krupinski 1993; Vogt 1999; Kapoula and Lestocart 2006; Vogt and Magnussen 2007; Pihko et al. 2011). In the context of artistic appreciation, we can, as above, enumerate a variety of cases in which attentional tuning is plausibly the best explanation of the ability of trained artists and critics to perceive features of artworks that novices are unlikely to notice or unable to perceive at all.

Indeed, part of the reason we read art criticism is that we are interested what experts perceive *in* a work—especially when we have the sense that there are salient perceptual features of a work that we have missed or sensed but couldn't quite identify. That is, part of the value of good criticism is the critic's ability to lucidly report what they perceived in a work in a way that helps us novices attend to those features. Consider this blog entry from the singular David Bordwell:

Some viewers and critics think the jarring quality of [*The Bourne Ultimatum* (2007)] proceeds from rapid editing. The cutting is indeed very fast (...). But there are other fast-cut films that don't yield the same dizzy effects (...). Something else is up. Partly, it's not the pace of the editing but the spasmodic quality of it. Cuts here seem abrasive because they interrupt actions and camera movements. Pans, zooms, and movements of the actors are seldom allowed to come to rest before the shot changes. This creates a strong sense of jerkiness and visual imbalance. (Bordwell and Thompson 2011, 167-168)

You can probably think of your own examples of turning to a trusted critic in the knowledge that, thanks to their expertise or training, are able to attend to artworks in ways that allow them to perceive salient features that you cannot. Although I have little musical training (and none on piano), I recognize that Yuja Wang's playing is astonishing. So, after watching a performance or listening to a recording of hers, I will turn to the review of an expert critic knowing that they have perceived qualities of the performance that I have not. If you are at all like me, you may regularly read the very *best* critics of an artform in which you *do* have some training because their expertise still regularly allows them to perceive artistic features of which you perhaps have only a fleeting recognition.

Or perhaps, like the Salieri of Peter Schaffer's *Amadeus* (1984), you have enough training in an artform to appreciate the astonishing ability of singular experts like Mozart to *just hear* that a particular interval or phrasing will work for a piece. Needless to say, not all perceptual learning leads to perceptual expertise—whether in art, sport, or some other context. And no doubt that some individuals, like Ted Williams, David Bordwell, Yuja Wang, and W.A. Mozart, are gifted with immense innate capacities that allow them to reach a level of expertise that most of us will never approach no matter how much training we have. But expertise is a scalar concept, and the empirical studies suggest there are substantive differences in how attention is tuned by those with extensive training and experience in a given domain in comparison to laypersons. As Dustin Stokes puts it in a recent, empirically informed philosophical account worth quoting at length:

For many kinds of expertise, the exceptional performance of the expert involves sensory perception in some important way. The expert knows where to look, how to listen, what she is tasting, and so on. She sees things more quickly, distinguishes patterns that others cannot detect at all, and rapidly makes comparisons between perceptible features that others can scarcely understand. This kind of achievement clearly involves experience and learning, and often requires explicit, time-consuming training specific to the relevant domain. It is also intuitive that this kind of expertise is, in a rich sense, genuinely perceptual. Put simply, it is plausible that many experts are better perceivers in the domain(s) of their expertise. (Stokes 2021, 242)

In the context of aesthetic education, educators are, in Stokes's sense, perceptual experts, and one of their central aims is to abet perceptual learning that will help students attend more closely to relevant artistic features. Thus, although I have identified three plausible mechanisms of perceptual learning—attribute differentiation, unitization, and attentional tuning—the cultivation of particular sorts of attention, broadly speaking, is the upshot of successful cases of perceptual learning.

At this point, an interesting question naturally arises. If the goal of facilitating perceptual learning in sport is to become more competitive and, in the context of artistic creation, to become a better musician, composer, painter, or whatever, what could be the aim of fostering perceptual learning in the context of artistic appreciation? Here we need to return to the term “aesthetic education”. What is distinctive about aesthetic education, I think, is that unlike in any other context, the ideal upshot of perceptual learning is enhanced aesthetic experience and fuller aesthetic appreciation. That is, the rationale for facilitating perceptual learning in the classroom is that it will afford students

more complete aesthetic experiences of artworks and help them more fully appreciate them.

Recently, Bence Nanay (2016) has developed a sustained account of the role of distributed attention in many paradigm cases of aesthetic experience. In part, Nanay's aim is to rehabilitate the idea of "aesthetic attention". However, Nanay distinguishes his use of this term from the sense in which Jerome Stolnitz (1960) used it (and which George Dickie (1964) famously critiqued)—i.e. as *disinterested attention* or attention to an object for its own sake. Rather, the sort of attention that Nanay characterizes as *aesthetic* is "focused with regards to the perceptual object [and] distributed with regards to the properties of the perceptual object" (2016, 29). Nanay is careful not to claim that such attention is either necessary or sufficient for aesthetic experience. However, he argues that, in some cases, such attention is needed "to appreciate the unity and integration of what we experience aesthetically" (2016, 29). Of course, artworks are particular *sorts* of perceptual objects (though not all artworks are perceptual objects); they are designed and shaped with particular intentions and they aim at particular effects. It is for these reasons that William P. Seeley characterizes artworks as "attentional engines": "They are artifacts designed to direct our attention to those features sufficient to categorize them so that we can recover their artistically salient content" (2020, 50). So, attention is also important to aesthetic education in another way: it is not only central to perceptual learning *in general*, but, furthermore, to our aesthetic engagement with artworks.

Two potential objections should be addressed. First, if attention really is centrally involved in many paradigm cases of aesthetic experience, then, surely, no cultivation of it is necessary since evidently people with no aesthetic education still have aesthetic lives full of rich aesthetic experiences. Second, if artworks are, in fact, typically designed to direct attention in particular ways, then, once again, the development of attention seems unnecessary. There is some truth in both of these objections. It is true that people with standard, functional perceptual capacities already have the ability to direct their attention in whatever ways are minimally necessary to have aesthetic experiences. It is also true that a great number of artworks, especially what Noël Carroll (1998) terms *mass artworks*, are designed to direct our attention in ways that make them legible. Just as standard, functional perceptual capacities are sufficient for aesthetic experience, they are also sufficient for engagement with at least some forms of art. Notwithstanding the elements of truth in these objections, both neglect the fact that aesthetic appreciation can be more or less thoroughgoing and aesthetic experience can be more or less rich.

In view of the ease of legibility of mass art, as described by Carroll, let us take movies as an example of art that is immediately accessible in some ways, but

which can be more fully appreciated if one's attentional skills are refined. In the discipline of film studies, the textbook *Film Art*—first published by David Bordwell and Kristin Thompson (1979), and in later additions also published with Jeff Smith (2024)—stands out as a paradigm case of an effort to tune attention with the aim of enhancing students' aesthetic appreciation and aesthetic experiences of cinema. Consider this passage:

In watching a narrative film, we usually don't notice style; we're too busy following the story. Suppose, though, we want to notice stylistic patterning—to enhance our appreciation, or to understand how we might also create films. How can we study style? One suggestion is apparent: *Look and listen carefully*. (Bordwell, Thompson, and Smith 2020, 307)

And, in fact, much of the 300 pages that precede this passage go into exacting detail, replete with specific examples and still images, about how one might attend to cinema in a way that brings these patterns into sharp relief.

Film Art's emphasis on teaching students how to attend to films is complemented by explanations of the various ways films themselves are designed to guide attention in the sorts of ways Carroll (1998) and Seeley (2020) describe. In one interesting passage, Bordwell, Thompson, and Smith note that Tim J. Smith's eye tracking research provides some empirical evidence for the hypothesis that films regularly direct viewers' gazes by manipulating various aspects of mise-en-scene, including lighting and staging. But of course, lighting and staging are just two of many aspects of film form that filmmakers use to guide attention. William P. Seeley and Noël Carroll (2013; 2014) argue that another key cinematic technique here is “variable framing”—the way in which a film changes our vantage point on the objects and events of the story world. For many years, film scholars and educators gave less attention to sound. But recently that has changed thanks to recent scholarship on the topic, as well as the relative affordability of multi-channel surround sound systems in university screening rooms. With a decent sound system, I can now ask my class to try to attend to the 360-degree panning of the sound of helicopter in the opening sequence of *Apocalypse Now* (1979).

Thus far, my focus has been upon the development of perceptual skills and the engagement with the formal properties of artworks in aesthetic education. But, of course, we also bring our cognitive capacities to our experience of art, and, as I indicated at the outset, it is plausible that the development of cognitive skills and knowledge is another central aim of aesthetic education. As Seeley puts it,

Consumers need a lot of background knowledge to *see-with* an artwork. At a minimum they need an understanding of the productive and evaluative norms and conventions that define different categories of art. Artistic norms and conventions are tools that guide how we *see what is shown* with an artwork. (Seeley 2020, 31)

5. Developing art-historical knowledge

Recall Alan Goldman's claim, which served as a jumping off point for this discussion, about aesthetic education as learning to direct one's attention to three sorts of relations within artworks: formal, referential (representational and expressive), and historical (1990, 113). So far, my focus has largely been on the perception of formal features and relations. But the apprehension of many other properties, including representational, expressive, and historical properties requires us to deploy our cognitive resources (primarily, in this context, knowledge)—and, as Seeley suggests, sometimes those cognitive resources in fact bear upon how we actually perceive a work's properties.

Let me briefly sketch some of the sorts of knowledge that aesthetic education aims to develop and how such knowledge directs our attention to (and abets our apprehension of) various features and their relations. First, consider the material basis of an artform. In degrees devoted to arts ranging from cinema to music to painting, it is not uncommon for students to be required to spend some amount of time in practice-based courses—even if their focus is, say, musicology, art history, dramaturgy, film studies, and so forth. One of the reasons for this is, to paraphrase David Bordwell, such courses help them “think like artists” by illuminating the various affordances and constraints offered by the material basis of the artform in which they are working (on the relevance of constraints for artistic creation, see Smith this issue). It's hard to appreciate the bravura of the long take(s) (of scenes involving hundreds of extras) of *Russian Ark* (2002) or *Roma* (2018) if one doesn't have a sense of what's physically and practically involved in choreographing an extended take for a mobile camera. A similar point holds for the performance of musical and theatrical pieces. Wang's performances, for example, are bold and dazzling in part because of the speed and clarity with which she is able to play intricate phrasings across the length of the piano. As Kendall Walton puts it,

The energy and brilliance of a fast violin or piano passage derives not merely from the absolute speed of the music (together with accents, rhythmic characteristics, and so forth), but from the fact that it is fast *for that particular medium*. (Walton 2008 [1970], 207)

In some cases, the nature of a work's expressive properties has a direct connection to the material base of the artform. The sombre, ominous, stark qualities of *films noir* are partly engendered by the use of black and white film stocks and low-key lighting setups. The warmth and resonance of an electric guitar solo might derive from choices about string gauge, pickups, and amplifier. The richness of color in Venetian paintings of the Cinquecento partly owes to a particular method of applying paint.

Of course, artists don't ply their materials in vacuums; rather, they work within particular art-historical contexts in which the use of materials is governed by widely shared norms and conventions. So, the artistically salient features of any work (including representational and expressive properties) are also partly determined by the artistic norms and conventions operative in the context of creation. Thus, another aim of aesthetic education is to familiarize students with those norms and conventions. As Seeley and others have pointed out, oftentimes particular groups of norms and conventions comprise particular categories of art—genres, styles, and so forth. In film studies, we teach students about the norms of spatio-temporal continuity in the classical Hollywood style, and the narrative conventions of the horror genre. In English literature courses, we detail the stylistic conventions of post-modernist literature. And in art history classes, we explain the norms of depicting depth in painting across various cultures and eras.

Importantly, the knowledge of such categories of art, and, more specifically, the norms and conventions constituting them, as well as the historical relations between various categories sometimes have top-down effects on how we perceive particular works (and their features). As E.H. Gombrich put it, “A style, like a culture or climate of opinion, sets up a horizon of expectation, a mental set, which registers deviations and modifications with exaggerated sensitivity” (2000 [1960], 60.) This is part of the point of Kendall Walton's (2008 [1970]) famous thought experiment involving an alien society that has an artform known as ‘guernicas’” (2008 [1970], 204). Instances of this artform have the “colors and shapes of Picasso's *Guernica*, but the surfaces are molded to protrude from the wall like relief maps of different kinds of terrain” (2008 [1970], 204). Walton suggests that, for a member of this society, Picasso's work would count as a “guernica”, yet

Its flatness, which its standard for us, would be variable for members of the other society, and the figures on the surface, which are variable for us, would be standard for them. (Walton 2008 [1970], 205)

Moreover, he concludes,

This would make for a profound difference between our aesthetic reaction to *Guernica* and theirs. It seems violent, dynamic, vital, and disturbing to us. But I imagine it would strike them as cold, stark, lifeless, or serene and restful, or perhaps bland, dull, boring—but in any case *not* violent, dynamic and vital. (Walton 2008 [1970], 205)

Those of us who have grown up on movies that are shot and cut with the freneticism of *The Bourne Ultimatum* (2007) will need to place a film like *His Girl Friday* (1940) in a related, but slightly different category. Compared to *The Bourne Ultimatum*, *His Girl Friday* may strike one as neither fast nor energetic. But *His Girl Friday* is fast and energetic for a Hollywood screwball comedy, which is the relevant comparison category.

This latter point takes us beyond a descriptive, psychological question about top-down effects on our perception of artworks to normative questions about the categories in which artworks *should* be placed and, in relation, what aesthetically relevant features they actually have. As Walton notes, “One way of approaching this problem is to deny that apparently conflicting aesthetic judgments of people who perceive a work in different categories actually do conflict” (2008 [1970], 210). But here we see another common aim of aesthetic education—namely, the cultivation of art-historical knowledge that facilitates the *correct* categorization of artworks and, thus, better apprehension of the aesthetically relevant features of those works. Walton argues that normativity about artistic categories tacitly underlies our ordinary, appreciative practices:

[O]ne who asserts that a good performance of the *Adagio Cantabile* of Beethoven’s *Pathétique* is percussive, or that a Roman bust looks like a unicolored, immobile man severed at the chest and depicts one as such, is simply wrong, even if his judgment is a result of his perceiving the work in different categories from those in which we perceive it. (Walton 2008 [1970], 211)

Interestingly, Walton goes on to assert that rather than privileging our own aesthetic judgments, we also admit when we have judged incorrectly if we subsequently recognize that we perceived a work in the wrong category. According to Walton,

We are likely to regard cubist paintings, or Japanese *gagaku* music, as formless, incoherent, or disturbing on our first contact with these forms largely because, I suggest, we would not be

perceiving the works as cubist paintings, or as *gagaku* music. But after becoming familiar with these kinds of art, we would probably *retract* our previous judgments, admit they were mistaken. (Walton 2008 [1970], 211)

I think Walton is right about this if the ‘we’ in the sentence refers to people who already have a fair amount of aesthetic education and/or experience engaging with unfamiliar artforms. But, in fact, I submit, part of the point of aesthetic education is precisely to help students recognize that artworks can be categorized and perceived more or less correctly, and, furthermore, that we can and *should* make more accurate judgments about works partly by learning about the categories in which they are correctly perceived.

Consider, for example, the familiar experience of showing experimental film to students whose familiarity with cinema does not extend far beyond the multiplex. Common responses include comments like: “That’s not a movie”, “That’s the worst movie I’ve ever seen”, and “That was the most boring movie I’ve ever seen”. It hardly matters if one shows them work by Hollis Frampton, Maya Deren, Gunvor Nelson, Michael Snow or Su Friedrich; the responses follow a similar pattern. The reason is that relatively few of the features that students assume to be “standard” (to use Walton’s terminology) for cinema are present in experimental film (character, narrative structure, narrative causality, etc.). And, in fact, sometimes students are vocally frustrated that experimental films so radically defy their expectations.

In such contexts, the aim of aesthetic education is to provide students with the relevant art-historical knowledge that allows them to reconceptualize the categories in which they perceive films. That is, we try to get them to see the films with which they are familiar as belonging to a more specific category than “movies”—namely, the classical Hollywood style. And (again, staying with Walton’s terminology) we teach them what features are standard, contra-standard, and variable for that category. We also teach them about the art-historical context in which experimental film arose and help them become more familiar with instances of that category, so they can also start to recognize *its* standard, contra-standard, and variable features. The aim is, I submit, not to get students to *like* experimental film (though we may hope that some of them will come to like it); rather, it is to provide them with the knowledge that allows them to see it in the correct context and, as a result, attend to its artistically relevant features. The hope is that students will be in a better position to aesthetically appreciate such works.

6. Conclusion

In the preceding discussion, I have suggested that there are a variety of benefits to conceiving aesthetic education as primarily involving the development of perceptual-cognitive skills rather than the cultivation of taste. One advantage of the perceptual-cognitive model I have sketched is that it dissolves an apparent puzzle about aesthetic education that looks like a version of Kant's antinomy of taste. A second merit is that the perceptual-cognitive model more accurately describes what aesthetic educators actually do. A third asset is that, even when aesthetic educators *do* aim at developing taste (*qua* critical judgment), they necessarily do this *by* attempting to hone students perceptual-cognitive skills.⁵ I concluded the previous section by suggesting that one underlying purpose of the perceptual-cognitive model of aesthetic education is to improve aesthetic appreciation. In this sense, the perceptual-cognitive model still has a normative dimension. Aesthetic appreciation (of which critical evaluation is a part) can be done better or worse, more or less accurately. But the perceptual-cognitive model avoids the challenges of Kant's antinomy of taste (and media studies scholars' objections to *taste*) by grounding normativity in facts about features an artwork actually has rather than perceivers' emotive responses to it. Only by conceiving of the normativity of aesthetic education in this way (as grounded in facts about what artworks are actually like) can we dissolve the apparent puzzle of aesthetic education, let alone make a practical case to those who hold the purse-strings about its value.

Acknowledgments

I am grateful to Noël Carroll, Murray Smith, Iris Vidmar Jovanović, and two anonymous referees for helpful feedback on an earlier draft of this article. Whatever infelicities or mistakes remain are my sole responsibility. I would like to dedicate this paper to the memory of David Bordwell—aesthetic educator extraordinaire.

⁵ Moreover, this view of aesthetic education meshes with an intuitively plausible view of education in general, according to which educators aim not to tell students what to think, but to teach them *how* to think so that they might arrive at more informed, accurate conclusions. Thanks to Murray Smith for mentioning this to me in personal correspondence.

REFERENCES

- Bordwell, David and Kristin Thompson. 2011. *Minding Movies: Observations on the Art, Craft, and Business of Filmmaking*. Chicago: The University of Chicago Press.
- Bordwell, David, Kristin Thompson, and Jeff Smith. 2020 [1979]. *Film Art: An Introduction*. 12th ed. New York: McGraw Hill Education.
- Butler, Jeremy. 2018. *Television: Visual Storytelling and Screen Culture*, 5th ed. New York: Routledge.
- Cardwell, Sarah. 2013. "Television Aesthetics: Stylistic Analysis and Beyond." In *Television Aesthetics and Style*, edited by Jason Jacobs and Steven Peacock, 23-44. London: Bloomsbury.
- Carroll, Noël. 1998. *A Philosophy of Mass Art*. Oxford: Clarendon Press.
- . 2010. "Aesthetic Experience: A Question of Content." In his *Art in Three Dimensions*, 77-108. Oxford: Oxford University Press.
- . 2022. "Forget Taste." *Journal of Aesthetic Education* 56 (1): 1-27.
- . 2023. "Two Approaches to Aesthetic Experience." *Journal of Aesthetic Education* 57 (4): 1-17.
- Carroll, Noël, and William P. Seeley. 2013. "Cognitivism, Psychology, and Neuroscience: Movies as Attentional Engines." In *Psychocinematics: Exploring Cognition at the Movies*, edited by Arthur P. Shimamura, 53-75. Oxford: Oxford University Press.
- Connolly, Kevin. 2019. *Perceptual Learning: The Flexibility of the Senses*. Oxford: Oxford University Press.
- Cutting, James E., Jordan E. DeLong, and Christine E. Nothelfer. 2010. "Attention and the Evolution of Hollywood Film." *Psychological Science* 21 (3): 432-439.
- Davids, K., Williams, A.M., & Williams, J.G. (eds.). 1998. *Visual Perception and Action in Sport*. First edition. New York: Routledge.
<https://doi.org/10.4324/9780203979952>
- Dickie, George. 1964. "The Myth of the Aesthetic Attitude." *American Philosophical Quarterly* 1: 56-65.
- Fenner, David (2020). "Developing Aesthetic Taste." *Journal of Aesthetic Education* 54 (2):113-122.
- Fujioka, Takako et al. 2004. "Musical Training Enhances Automatic Encoding of Melodic Contour and Interval Structure." *Journal of Cognitive Neuroscience* 16 (6): 1010-1021.
- Gibson, Eleanor J. 1963. "Perceptual Learning." *Annual Review of Psychology* 14: 29-59.
- Goldman, Alan H. 1990. "The Education of Taste." *The British Journal of Aesthetics* 30 (2): 105-116.
- Goldstone, Robert L., and Lisa A. Byrge. 2015. "Perceptual Learning." In *The Oxford Handbook of Philosophy of Perception*, edited by Mohan Matthen, 812-832. Oxford: Oxford University Press.

- Gombrich, E.H. 2000 [1960]. *Art and Illusion: A Study in the Psychology of Pictorial Representation*. Princeton: Princeton University Press.
- Grossman, Walter. 1968. "Schiller's Aesthetic Education." *The Journal of Aesthetic Education* 2 (1): 31-41.
- Guyer, Paul. *A History of Modern Aesthetics*. Cambridge: Cambridge University Press.
- Heimann, Katrin S., et al. 2017. "'Cuts in Action': A High-Density EEG Study Investigating the Neural Correlates of Different Editing Techniques in Film." *Cognitive Science* 41: 1555-1588.
- Irvine, Dexter R.F. 2018. "Auditory Perceptual Learning and Changes in the Conceptualization of the Auditory Cortex." *Hearing Research* 366: 3-16.
- Kapoula, Zoï, and Louis-José Lestocart. 2006. "Perception de l'espace et du mouvement dans Study of a Dog (Étude de chien)." *Intellectica* 44: 215-226.
- Kranse, Sally, et al. 2013. "Applying Perceptual and Adaptive Learning Techniques for Teaching Introductory Histopathology." *Journal of Pathology Informatics* 4 (1). 34.
<https://doi.org/10.4103/2153-3539.123991>
- Leontiev, Kim. 2023. "Schiller on Aesthetic Education as Radical Ethical-Political Remedy." *The British Journal of Aesthetics* 63 (4): 553-578.
- Levine, E. and Michael Z. Newman. *Legitimizing Television: Media Convergence and Cultural Status*. New York: Routledge.
- Magliano, Joseph P., and Jeffrey M. Zacks. 2011. "The Impact of Continuity Editing in Narrative Film on Event Segmentation." *Cognitive Science* 35: 1489-1517.
- Matherne, Samantha, and Nick Riggle. 2020. "Schiller on Freedom and Aesthetic Value: Part 1." *The British Journal of Aesthetics* 60 (4): 375-402.
- . 2021. "Schiller on Freedom and Aesthetic Value: Part II." *The British Journal of Aesthetics* 61 (1): 17-40.
- McAdoo, Nick. 1987. "Aesthetic Education and the 'Antinomy of Taste.'" *The British Journal of Aesthetics* 27 (4): 307-318.
- Nanay, Bence. 2016. *Aesthetics as Philosophy of Perception*. Oxford: Oxford University Press.
- Nodine, C.F., P.J. Locher, and E.A. Krupinski. 1993. "The Role of Formal Art Training on Perception and Aesthetic Judgment of Art Compositions." *Leonardo* 26 (3): 219-227.
- Pantev, Christo, and Sibylle C. Herholz. 2011. "Plasticity of the Human Auditory Cortex Related to Musical Training." *Neuroscience and Biobehavioral Reviews* 35 (10): 2140-2154.
- Pihko, Elina, et al. 2011. "Experiencing Art: The Influence of Expertise and Painting Abstraction Level." *Frontiers in Human Neuroscience* 5, article 94. doi: 10.3389/fnhum.2011.00094

- Piper, H. (2016). "Broadcast Drama and the Problem of Television Aesthetics: Home, Nation, Universe." *Screen*, 57(2), 163-183.
hjjw021. <https://doi.org/10.1093/screen/hjjw021>
- Prettyman, Adrienne. 2018. "Perceptual Learning." *Wiley Interdisciplinary Reviews: Cognitive Science* 10 (3). <https://doi.org/10.1002/wcs.1489>
- Ransom, Madeleine. 2022. "Aesthetic Perception and the Puzzle of Training." *Synthese* 200, 127: <https://doi.org/10.1007/s11229-022-03555-8>
- Seeley, William P. 2020. *Attentional Engines: A Perceptual Theory of the Arts*. Oxford: Oxford University Press.
- Seeley, William, and Noël Carroll. 2014. "Cognitive Theory and the Individual Film: The Case of Rear Window." In *Cognitive Media Theory*, edited by Ted Nannicelli and Paul Taberham, 235-252.
- Seitz, Aaron R. 2018. "Perceptual Learning." *Current Biology* 27 (13): R623-R641.
- Smith, Murray. 2024. "Artistic freedom realised." *European Journal of Analytic Philosophy*. 20 (2): 259-282.
<https://doi.org/10.31820/ejap.20.2.1>
- Smith, Tim J. 2012. "The Attentional Theory of Cinematic Continuity." *Projections: The Journal for Movies and Mind* 6 (1): 1-27.
- Stokes, Dustin. 2021. "On Perceptual Expertise." *Mind & Language* 36: 241-263.
- Stolnitz, Jerome. 1960. *Aesthetics and Philosophy of Art Criticism*. New York: Houghton Mifflin.
- Tauber, Zvi. 2006. "Aesthetic Education for Morality: Schiller and Kant." *The Journal of Aesthetic Education* 40 (3): 22-47.
- Vogt, Stine. 1999. "Looking at Paintings: Patterns of Eye Movements in Artistically Naïve and Sophisticated Subjects." *Leonardo* 32 (4): 325.
- Vogt, Stine, and Svein Magnussen. 2007. "Expertise in Pictorial Perception: Eye Movement Patterns and Visual Memory in Artists and Laymen." *Perception* 36: 91-100.
- Walton, Kendall L. 2008 [1970]. "Categories of Art." In his *Marvelous Images*. Oxford: Oxford University Press.
- Wolterstorff, Nicholas. 2015. *Art Rethought: The Social Practices of Art*. Oxford: Oxford University Press.
- Zangwill, Nick. 2001. *The Metaphysics of Beauty*. Ithaca: Cornell University Press.

ETHICS EDUCATION FROM SUFFERING ON SCREEN? TRAGIC VISIONS IN *ARRIVAL*

James MacAllister¹ 

¹The University of Edinburgh, United Kingdom

Original scientific paper – Received: 09/02/2024 Accepted: 13/07/2024

ABSTRACT

In this paper I argue that tragic films can have significant potential for ethics education when they prompt audiences to sympathise with suffering on screen. I first summarise two accounts of the relationship between tragic art, moral education and aesthetic value (those provided by Rorty and Lamarque). I then discuss problems with these accounts and explain how a new criterion of aesthetic value might help to resolve them. I thereafter argue that tragic films have potential to ethically educate audiences in a way that enhances the aesthetic value of the films in at least three directions: by deepening moral understanding, by deepening understanding of the nature of human being and ethical purpose and by deepening understanding of ethical theory. I conclude by showing how Denis Villeneuve's film, *Arrival*, screens a tragic story with ethics education potential in each of the aforementioned senses.

Keywords: screen suffering; tragic film; ethics education through film; educational ethicism; ethics in Villeneuve's *Arrival*.

1. Tragic film

In this paper I explore the possibility that audiences might be ethically educated by some tragic screen stories and, more particularly, by good tragic films. In my view a good tragic film will very often possess at least two of the following three dimensions. The plot will involve suffering (usually familial in nature), that the audience will be prompted to sympathise with, and where the suffering on screen carries potential for ethics education.¹ The idea that a good tragic film will involve suffering of ethical import, usually within a family, is inspired by Aristotle's *Poetics*.² However, the two main arguments in this paper, the ones put forward at the start of the next paragraph, are not, strictly speaking, Aristotelian. Aristotle did not speak about tragic film having potential for ethics education in the senses I do. His treatise on tragedy was fairly obviously not about film but ancient Greek theatre. In speaking of a good tragic film having tragic dimensions, I therefore mean to suggest that there are both continuities and discontinuities between contemporary tragic films and ancient tragic art. While the three dimensions I highlight pick up on points of similarity between good tragedy, old and new alike, I also recognise that important differences exist between the experiences of viewers of contemporary screen tragedies and audiences of ancient Greek tragic performances on stage. I speak of films with tragic dimensions as opposed to films that are tragic through and through in an attempt to recognise and acknowledge these similarities and differences.

I seek to defend two related claims in this paper about good tragic films. First, that some tragic films may prompt an actual or possible audience to a gain in ethical understanding. Second, that a tragic film will be enhanced as an artwork if it contains an aesthetic property capable of prompting gain in ethical understanding. My understanding of the possible ethical, aesthetic and educational value of tragic film has been informed by the work of Amelie Rorty and Peter Lamarque. In this paper I will therefore first unpack their accounts of the relationship between tragic art, moral education and aesthetic value. I then discuss problems with these accounts and explain how a new criterion of aesthetic value might help to resolve them. I thereafter explain how tragic films have potential to ethically educate audiences in a way that enhances the

¹ My argument, that good tragic films have potential for ethics education based on an experience of shared suffering, has been influenced by the work of Stacie Friend on documentary tragedy. Friend (2007) claims that documentary tragedies deal with a non-comedic subject in a way that affords insight into experience, often moral experience.

² He stipulated that the best tragedies are concerned with "situations in which sufferings arise within close relationships", most often within families (*Poetics*, 53b18-22).

aesthetic value of the films in at least three directions: by deepening moral understanding, by deepening understanding of the nature of human being and ethical purpose and by deepening understanding of ethical theory. I conclude the paper by showing how *Arrival* (2016) screens a tragic story with ethics education potential in each of the senses mentioned just now.

2. Aristotle, wayward action and ethical purpose

Amelie Rorty (1991) argues Aristotle provides the best account of our experience of tragedy. Aristotle famously asserted that tragedy is “an imitation of an action that is admirable (...) effecting through pity and fear the catharsis of such emotions” (*Poetics*, 1149b24-28). According to Aristotle tragedy re-presents (mimesis) an action (spoudais) in a plot (mythos) that is complete and possesses magnitude (*Poetics*, 1149b10-14). He specifies that tragedy imitates not persons, but actions and life (*Poetics*, 1150a15-16). The plot organises action (*Poetics*, 1150a3-4), and “wellbeing and illbeing reside in action” (*Poetics*, 1150a16-17), so the plot is the “source and soul of tragedy; character is second” (*Poetics*, 1150b1-2). Aristotle adds that tragedy must depict characters of a certain kind in regard to action and reasoning (*Poetics*, 1149b38-1150). More specifically, Aristotle claims the best tragedies should depict someone of moderately good character but with a flaw (*Poetics* 53a). Rorty provides an illuminating account of this. She argues that translating flaw as “error” or “mistake” fails to do justice to the “dispositional nature” of the hero’s hamartia (Rorty 1991, 61).

Rorty argues that in the best tragedies the main character possesses an “erring waywardness” that renders them susceptible to a misfortunate change in the “projected arc” of their lives (Rorty, 1991, 54). This waywardness in the tragic protagonist often entails initial ignorance about who one really is (Rorty 1991). To unpack her point she takes the case of Oedipus. Initially Oedipus was acting under the understanding that he was marrying the Queen of Thebes. While true, he was nonetheless ignorant of the crucial fact that should have determined the course of his practical deliberations—he was also marrying his mother. Oedipus, she claims, depicts “the story of an action that undoes a person of high energetic intelligence” (Rorty 1991, 58). Rorty develops her argument out from Aristotle’s point that it is more than anything actions that determine the extent to which people experience “wellbeing” or “illbeing” in their lives. She claims that the best tragedies attain unity by showing how the serious actions (spoudaios) of the hero connect into a coherent whole.

She says that serious actions are those actions that define a person's life and "make a difference to how a person lives" (Rorty 1991, 57). Tragedy depicts serious actions that go wayward and deliver disaster to the main protagonist (Rorty 1991). She argues that sometimes the very actions that aim at flourishing, end up being the root of a person's undoing. Sometimes the vigour by which a person acts for the sake of their happiness ends up "blinding" them to important facts at the periphery of their awareness (Rorty 1991, 59). In tragedy more than anything it is wayward, serious actions that undermine the prosperity of the hero. She claims that Oedipus was in many respects the quintessential tragic Aristotelian hero.³ His "bold", "quick" and "intelligent" character is "essential" to his action (Rorty 1991). His bold quickness is both his greatest strength and the source of his demise. In his haste to leave Corinth he fails to interrogate rumours about who his parents really are. On his attack by a stranger, he retaliates too quickly and fails to ask who the stranger really is.

A "more measured man" perhaps would have been more cautious and not suffered the fate of Oedipus (Rorty 1991, 69), but Oedipus would not have been the bold man he was without his impetuosity. For Rorty, the downfall of Oedipus was a "by product" of his excellence. Oedipus strove to avoid his fate, but in this very striving, he brought about the conditions of his downfall. She concludes that

[I]t is no accident that excellence sometimes undoes itself, one of the dark lessons of tragedy is that sometimes there are no lessons to be learnt, in order to avoid tragedy. (Rorty 1991, 68)

In Rorty's view, tragedy does not provide audiences with "moralized warnings" about "actions to avoid" (Rorty 1991, 68). Nor does it affirm the view that chance dictates all in life. Instead, tragedy can deepen audience understanding about a neglected aspect of the relationship between human nature and ethical purpose. While it is "in our nature to strive for what is best in us" (Rorty 1991, 68) sometimes this striving can become the source of great suffering.

³ This seems fair as Aristotle says that the best tragedies contain a moment of recognition (anagnorisis), 'a change from ignorance to knowledge' that coincides with the reversal in fortune (peripetia) of the hero (Poetics, 52a11-52b) and he provides Oedipus as an example.

3. Aesthetic value, moral lessons and moral vision

I am sympathetic with the thrust of Rorty's account but it is also rather fatalistic. Her belief that tragedy does not provide moral warnings about actions to avoid rather ignores the possibility that the flaw in the main protagonist in a tragic film might be one that audiences can learn to avoid. Rorty is also not entirely clear that any moral educational benefit from tragedy can add to the value of the tragedy as an artwork. She notably asserts that the *Poetics* does not provide an "aesthetic theory". However, she does recognise that Aristotle directs his analysis onto the formal artistic properties in the best tragedies. These formal aesthetic properties interest him she claims because of the effect they can have on the understanding of the audience. Contrary to Plato, Aristotle wanted to demonstrate how tragedy can "promote instead of thwart understanding (...) attune rather than distort the emotions" (Rorty 1991, 54). While Rorty does not spell out this point exactly, the implication seems to be that tragedy has most aesthetic value when the formal structure of it delivers a particular educational effect: to help the audience see how the well-intentioned (but ill-judged) actions of an admirable character can lead to great suffering.

Peter Lamarque (1995) shares one of Rorty's core convictions about tragedy. He agrees that the flaw of the tragic hero is a "contingent by-product" (Lamarque 1995, 240) of an otherwise commendable character. However, he is more explicit about his belief that Aristotle is well placed to explain the distinctively aesthetic value of tragedy. Like Aristotle, Lamarque believes that the best tragedies have a moral content of almost universal human interest. He maintains that tragedies engage with "some of the deepest concerns of human beings in their attempts and repeated failures at living a moral life" (Lamarque 1995, 241). What most interests Lamarque is the artistic means and modes by which tragedy explores the moral concerns of human beings. Lamarque therefore unpacks the question of how the representation of suffering and disaster in tragedy can have both moral and artistic value.

He argues that tragedy can communicate moral content in at least two ways: through a *moral lesson* on the one hand and a *moral vision* on the other. In the case of the moral lesson a tragedy will expressly aim at teaching a moral principle. Lamarque does not think that any moral learning in such a mode will inevitably be superficial but he does perceive an overall deficiency in the moral lesson view of tragedy. Either the moral lesson is too *intimately connected* to the plot and characters in a specific tragedy, so that it cannot be turned into a more general moral principle, or the principle will be *so removed from* a specific tragedy that

it cannot be meaningfully connected to events and characters in it (Lamarque 1995). In both cases the moral lesson might be assimilated by the audience but it is hard to see how the moral lesson adds to the *aesthetic* value of the work.

However, the dichotomy here developed by Lamarque seems questionable. There need not be divergence between aesthetic value and a moral lesson that is stated propositionally in a tragic artwork. Indeed, some individual moral lessons in a tragedy might make significant contributions to the coherence and unity of the overall plot and pleasure in tragedy, adding to the aesthetic value of the artwork as a result, while at the same still carrying significant potential for moral education. Some moral lessons may in short carry potential for moral education when viewed in isolation but gain deeper resonance when viewed in relation to the wider plot and artistry of the artwork as a whole.⁴ A film could, for example, communicate some moral lesson in propositional form while employing artistic means to reinforce this same moral message more visually. Indeed, *Mountain* (2017) is a film that combines voice over and visual image (that is to say to moral lesson by both verbal proposition and visual artistic means) to deliver a powerful moral educational lesson.⁵

However, Lamarque does explain how the moral vision view can in isolation begin to make sense of how specifically aesthetic value can result from moral content in tragedy. In the mode of moral vision, the moral content in tragedy is *shown* rather than stated propositionally (Lamarque 1995). Here tragedy calls upon audiences to look upon a complex moral situation differently. Through this process they can “acquire” a new “vision or perspective on the world” (Lamarque 1995, 243). Lamarque insists that to fully grasp the nature of the relation between moral and artistic value in tragedy a further distinction is necessary: between internal and external audience perspectives. In the internal perspective the audience of tragedy imaginatively participates in the tragic world. When experienced in this perspective the characters in tragedy are “imagined to be fellow humans in real predicaments, objects of sympathy and concern, similar to ourselves in many respects” (Lamarque 1995, 247). Under the external perspective there is no imaginative engagement with the lives of the characters. Instead, characters are viewed as artistic constructs. In the external perspective, the focus of audience attention is on the “modes of presentation”

⁴ Friend (2007, 186) similarly claims that artistry in the “dramatic storyline” can help documentary tragedies possess an overall unity and coherence.

⁵ For discussion of how moral proposition and cinematic image combine in *Mountain* with significant potential for ethics education, see MacAllister (2024).

(Lamarque 1995, 247) and the extent to which the overall plot and structure help to bring about the desired cathartic effect.

Lamarque seems to be implying something like the following. The internal perspective encourages audiences to *experience* pity and fear toward tragic characters. The external perspective encourages audiences to reflect on whether or not their sympathy is *morally warranted* on the basis of how the suffering of the tragic hero has been *artistically shown*. Lamarque may not quote directly from the *Poetics* but he takes his suffering focussed, moral-vision view of tragedy to be broadly Aristotelian. He is right to think this. Aristotle after all defines suffering as “an action that involves destruction or pain; deaths in full view, extreme agony, woundings and so on”, and he maintains it is a central element in the plot of a good tragedy (*Poetics*, 1052b11). Aristotle also maintains that the most complete tragic plots should wherever possible “visualise what is happening. By envisaging things very vividly in this way, as if one were present at the actual event, one can find out what is appropriate” (*Poetics*, 1055a 17-21). Lamarque and Aristotle both regard the envisioning of suffering to be of central importance in tragedy. When a plot is unified and suffering is vividly visualised, audiences become best placed to recognise how the story of the hero hangs together.

4. Gaining ethical understanding from tragic film

I have so far discussed the views of Rorty and Lamarque on the moral educational powers of tragedy and I have drawn attention to weaknesses in both of these views. Rorty seems to suggest that audiences to tragedy may well be doomed to repeat the mistakes of tragic characters in art. The mistakes of tragic characters that contribute to ill-fate will be imitated in life rather than learned from and avoided. She is less than clear on whether or not any moral education from tragedy is related to the aesthetic value of tragedy too. Meanwhile, Lamarque’s account contains a questionable dichotomy and is weaker for not acknowledging that it is possible for individual propositions to morally educate in a way that adds to the artwork as artwork by either enhancing the overall unity of the plot or by combining with the visual images on show to deliver the same moral message but with deeper resonance. When viewed in isolation from each other, the accounts of Rorty and Lamarque cannot do justice to the various means by which tragic film might assist audiences to accrue gain in ethical understanding either.

However, in spite of these weaknesses, I believe both perspectives can be combined together, to complement each other and begin to more fully

account for the various means by which tragic film might be capable of ethically educating.⁶ Rorty and Lamarque do after all point to different ways that tragic film might deepen moral and ethical understanding. Tragedy can teach a dark ethical lesson about how the very character traits that make a person admirable can also bring about their downfall (Rorty's view). Lamarque highlights a further important element in any moral learning from tragic art. Tragedy can teach moral lessons via propositions but moral education will accrue most aesthetic value when the moral content is *shown* in a very specific way. Tragic art should foster sympathy toward the main protagonist and it should encourage audiences to reflect on whether or not this sympathy is morally warranted on the basis of what is artistically shown.

In order to begin my explanation about how these problems may be overcome and integrated into a wider framework about the ethics education potential of tragic film, I firstly want to note that I think it is preferable to speak of ethics education from tragic film (rather than moral education). One of the reasons I think it is preferable to employ the term ethics education is because I think this term better captures the particular focus on suffering that most tragic films have. The category of the ethical is I believe broader than the moral—the ethical includes the moral but the moral does not include the ethical. Morality, as understand it, is primarily concerned with the obligations we have to ourselves and to each other whereas ethics focusses on living beings pursuing their flourishing.⁷ Sometimes lives go well but sometimes people make mistakes or suffer ill-fortune at great cost to their prosperity. The lives of characters in films can similarly go well or be full of great suffering. Tragic films depict life stories of a special sort—stories of living beings suffering on screen because of chance events, bad luck, errors of judgment or moral mistakes.⁸ Conceived thusly, ethical concerns are at the heart of most films with tragic dimensions and these ethical concerns are I believe rich in potential for ethics education.

However, to fully overcome the problems associated with the theories of Rorty and Lamarque I think a new criterion of aesthetic value is needed. One that can explain how the variety of different moral and ethical concerns depicted in tragic films can add to the aesthetic value of tragic

⁶ Both of the perspectives contribute to part of my overall account. When put together the perspectives can more fully account (than they could if they remained as individual accounts on this topic) for how tragedy can have potential for ethics education that adds aesthetic value in a variety of ways.

⁷ Bernard Williams (2011) has influenced my thinking on the difference between ethics and morality.

⁸ Ridley (2009) makes a similar point about tragic art more broadly.

films. It is my argument that a tragic film can be enhanced as an artwork when it contains an identifiable artistic property that is capable of ethically educating an audience.⁹ It is worth noting that I am not here claiming that potential for ethics education is the *only* possible criterion of aesthetic value for tragic films. While it is my argument that potential for ethics education is one of the main ways that tragic films can be enhanced as artworks, I recognise that other aspects of a tragic film might add to their aesthetic value too. It is also worth noting that not all audiences of a film need to actually accrue any gain in ethical understanding from the film for it to accumulate additional aesthetic value in virtue of this potential. However, some actual or possible audience needs to be capable of being ethically educated by the film at some point in time.

On this account, aesthetic value is thus partly response dependent in that additional aesthetic value is dependent on the response (being ethically educated) of some actual or possible audience.¹⁰ However, aesthetic value must also be rooted in some identifiable artistic property of the artwork. There must be some specific artistic feature of a tragic film that can legitimately be said to be capable of prompting gain in ethical understanding too. In the case of tragic films, I think this artistic property may often be a coherent and unified plot that is able to prompt audience sympathy with suffering on screen and brings about gain in ethical understanding. Importantly, I think this new account of aesthetic value can help overcome the two main problems thus far identified in the theories of Rorty and Lamarque. It seems to me mistaken to think (as Rorty appears to) that audiences cannot learn to avoid repeating the mistakes of tragic characters. Some tragic films (*Force Majeure* (2014) is a good example) depict characters that make moral mistakes that lead to great familial suffering. Audiences of tragic films like this might learn not to repeat the moral mistakes of on-screen characters because of the pain and shame they can bring to a family. Such learning to avoid moral mistakes of tragic characters would constitute a gain in ethical understanding derived from an identifiable artistic property of the film—

⁹ Here it is not enough that a film just shows or says something with potential for ethics education. **How** it is shown or said matters. There should be an aspect of artistry in the property in the film that has ethics education potential. For further discussion of how film can accrue additional aesthetic value in virtue of ethics education potential, see MacAllister (2023a).

¹⁰ Lamarque (2020) also speaks about aesthetic value being response dependent. Although my analysis has largely drawn on Lamarque's earlier thoughts on how moral lessons and moral visions in tragedy can have aesthetic value, in his later work Lamarque stresses that aesthetic value more generally lies in human beings valuing how an object appears to them. Artworks are valuable as artworks when they are good works of their kind (see Lamarque, 2020).

the plot of the film after all revolves around the moral failure of Tomas and the consequences it has for him and his family.¹¹

Contrary to Rorty, it is my view that the moral flaw of the protagonist in a tragic story may sometimes be one that audiences can learn to avoid, and if aesthetic value is understood in the way proposed here, explanation can be given as to how this learning would add to the value of a tragic film as an artwork. My understanding of aesthetic value can also overcome the main problem identified in Lamarque's view on tragedy. Lamarque seems to hold the view that a moral lesson stated in propositional form in a tragedy would be irrelevant to the value of the tragedy as an artwork. It seems to me, however, that some moral lessons stated in propositional form in tragic art might be able to both deepen moral understanding and be important to the unity and coherence of the plot as a whole. A proposition so connected to the overall plot and pleasurable effect of a tragedy can in short be an artistic property of the tragedy with potential for ethics education. The account of aesthetic value developed here can help overcome the problems identified thus far in Rorty and Lamarque's views on moral education from tragic art.

This is not quite yet the end of this story though as I think there is a further gap in their thinking on tragedy. Rorty and Lamarque (and for that matter Aristotle) did not explore how tragic art might be capable of prompting deeper understanding of ethical *theory* but I think some tragic art can do this. My view is that good tragic films can have an "afterlife" where the cinematic images and ethical content live on in the experience of the audience after the final credits have rolled.¹² It is my argument that the audience might deepen their understanding of ethical theory in instances where the afterlife of a tragic film prompts them to engage with or revisit ethical theory and think more deeply about it. In sum, it is my argument that there are at least three possible ways that tragic films can have potential for ethics education that can add to the value of these films as artworks: by deepening moral understanding through moral lessons or a moral vision; by deepening ethical understanding of the nature of human being and purpose; and by deepening understanding of ethical theory.

¹¹ For further discussion of how audiences might be morally educated by *Force Majeure*, see MacAllister (2023b)

¹² Peter Kivy's (1997) discussion of the afterlife of literary artworks has informed my thinking here.

5. A story of suffering on screen

I want to pull my paper together by discussing how a good tragic film, *Arrival* (2016), might contain potential for ethics education in each of the senses outlined above. Among other things, *Arrival* screens a life story of suffering. Indeed, the film is based on a short story by Ted Chiang, actually called “Story of your life” (2002). Familial suffering is on display—the sort most commended by Aristotle. Suffering within a family is visually depicted on screen in the very first moments of *Arrival*. In a voice over Dr. Louise Banks (Amy Adams) says “memory is a strange thing. It doesn’t work like I thought it did. We are so bound by time, by its order”. While the full meaning of her utterance only becomes clear toward the end of the movie these words do provide a clue to the audience that all may not be exactly as it seems in this film.

Louise is then depicted engaging lovingly with a newly born baby, her baby. The images abruptly jump forward in time to show Louise playing a cowboy game with her daughter who now appears to be 5 or 6. A doctor is quickly thereafter shown examining Louise’s daughter, now a young adult, and we see Louise alone with the doctor in a hospital corridor. Louise is crying and in a voice over she says “and this was the end”. An image of Louise’s now bald (presumably from cancer treatment) and dying daughter follows. Louise’s child, whom we later learn is called Hannah, has died. All of this unfolds to the sound of Max Richter’s emotive piece *On the Nature of Daylight*. From the very start of *Arrival*, the images and music prime the audience to feel sympathetic pity and fear toward Louise. These first images, sounds and words artistically gesture toward why audience sympathy is warranted from an ethical point of view. The further visions that Louise goes on to experience in the movie make explicit why such sympathy is ethically justified.

Soon after, Louise remarks that “there are days that define your story beyond your life, like the day they arrived”. “They” are heptapod aliens in twelve space ships. Louise is a linguist and together with a physicist, Ian Donnelly (Jeremy Renner), she has to establish who the heptapod aliens are and why they have come. To do this they need to first decipher the heptapod language. Louise meets this challenge with intelligence and bravery. They meet the aliens in a divided room—heptapods on one side of a glass screen, humans the other. When attempts at verbal communication with the heptapods reach an impasse, Louise has the wit to try written language, a move that leads to a breakthrough. The heptapods squirt out a sign and then another in cloudy ink. These are signs in the heptapod language, a language that Louise comes to understand.

The misty air on the aliens' side of the room comes to take on the function of a "writing surface" that enables conversation between human and heptapod (Zavota 2020). Later, Louise bravely takes her helmet off and approaches the heptapods. While this may at first sight seem impetuous and risky, it is a very calculated act. She wants the heptapods to be able to clearly see her, the unique human, Louise. Until she is sure the heptapods can grasp the distinction between one specific member of a species and the whole species it is pointless to ask them "why are *you* here?" Louise's introduction precipitates a moment of interspecies connection. The heptapods not only share their own names with Louise in return, one of them also connects their digits to Louise's through the glass screen they communicate through. A handshake of sorts, an act of friendship making and commonality between beings from different planets. Thereafter Louise begins to experience more powerful visions of her daughter. Initially the audience is encouraged to think these are memories from the past. However, in a crucial moment of re-cognition within the final ten minutes of the movie it becomes clear that Louise has not yet had her daughter. Indeed, it is only in this instant that she realises the girl in her visions is her daughter and that these are visions of the future.

All along viewers have been seeing images of Louise's future, a future she too can now see and clearly, even though she has yet to actually live it. Acquisition of the heptapod language has enabled her to experience time in a non-linear way. She can now see into the future. In this moment the tragic heart of the plot comes together. Aristotle claims the best tragic plots involve moments of "astonishment" where the "events that evoke fear and pity (...) occur above all when things come about contrary to expectation but because of one another" (*Poetics*, 1052a1-3). For viewers who realise what they thought were memories from the past are actually visions from the future, this is an astonishing moment. The moment defies audience expectation but the film also come to make sense to the audience as a result of this moment. Up until then the narrative is ambiguous but after this moment it comes to possess a definite unity and coherence.¹³ The visions that Louise experiences are utterly tragic. She can see she will have a daughter called Hannah with Ian. She can see that Hannah will die young and she can see there is nothing she can do to stop it.¹⁴

¹³ Carruthers agrees that while viewers of the film know "from the beginning" that Hannah dies it is only after the "narrative turning point" near the end that viewers come to see she has "not yet been born" (2018, p 337).

¹⁴ There is I believe a further tragic dimension in the movie that is worthy of comment. Louise employs her new powers to see into the future in order to prevent humans from going to war with the

6. Ethics education form tragic visions of the future

Arrival is not just a tragic film then but a very good tragic film because it has potential for ethics education in a manner that augments its aesthetic value. *Arrival* tells a story of suffering and visions and visual and other artistic means are central to the telling of this story. The film is full of artistic flair. The first shot of the alien space craft diverting the clouds, for example, invokes a sense of the sublime and experience of this feeling is possible again later when Louise enters the mist and stands vulnerably side to side with the heptapods—she is utterly dwarfed by them. The tragic visions of the future that she experiences are often wordless yet they are integral to the plot. The film can ethically educate then in the sense suggested by Lamarque and Aristotle. Viewers of the film are artistically shown why Louise is a proper object of tragic sympathy. Louise is an admirable hero yet in a moment of “recognition” it becomes clear that devastating loss will befall her.

But if Louise is a tragic hero, as I am suggesting, how or where does she go wayward? How is the very strength of her character also the source of her suffering? The only possibility suggested in the film relates to what she tells Ian, or rather does not tell him, about their daughter. Hannah wants to know why her dad, Ian, left. Louise says “it’s my fault” to Hannah. She explains he left because she told him something about the future that “he was not ready to hear”, about an “unstoppable” illness. Ian thought Louise made the “wrong choice”. While Louise appears to think Ian left because she told him too *soon* about Hannah’s fate, I am inclined to think he left because she told him too *late*. Perhaps Louise’s “fault”, such as it is, relates to what she chose not to disclose to Ian when he asks her if she wants to make a baby at the end of the film.

Would it not have been better to have informed Ian there and then that Hannah would die very young? Louise knew this was to be Hannah’s fate but Ian did not. Perhaps Ian left because he felt he had a right to know about the suffering and loss that would follow from this moment—that Louise possessed foreknowledge of death she could and should have shared. Viewers are not shown enough to be sure. While the film does not provide answers, it does open up some interesting questions about the ethics of consent, reproduction and pregnancy (Carruthers 2018). If my interpretation of Louise’s “wrong choice” is right, there is a sense in

heptapods. She relays the dying words of a Chinese military leader’s General Shang (Tzi Ma) wife to him—“in war there are no winners, only widows”. I take this to be a tragic moment as women in tragic art are often depicted opposing war or lamenting the fallen in war.

which Louise is doubly punished for her foresight. From the moment of conception, she knows her daughter will die before her life has really begun. This must be horribly painful knowledge to live with.¹⁵ She knows she will go on to lose the father of her child too. A further painful loss, again, because of what she knows.

She has to first live *knowing she will* lose her daughter and husband and then she has to *live with the loss* of her daughter and husband. She has a “full view” of her suffering to come and the audience has “full view” of this too. To my mind Louise’s flaw (and flaw is no doubt too harsh) is the polar opposite of Oedipus’s. She has in common with him an energetic intelligence and bravery but whereas Oedipus sees too little of the future clearly, Louise sees too much of it clearly. A core strength of her character, her linguistic brilliance, is also the source of her suffering. Learning the heptapod language required a certain genius and striving on her part but the effort came at the cost of foreknowledge of great losses to come. *Arrival* thus has potential to ethically educate audiences in something like the dark sense outlined by Rorty. Perhaps the film can support some viewers to learn that great suffering can flow from how and what we strive to know. Even when the pursuit of knowledge is very well intended.

Arrival has potential for ethics education in other directions too. The film could for one help some viewers better understand an aspect of ethical theory—the virtue concept of courage. Upon first contact with the heptapods it is clear from the images that Louise is petrified but she endures the legitimate worry that she might die and acts to help the human species understand the purpose of the strangers. Louise also demonstrates courage in the way she lives with her knowledge about Hannah’s fate. What matters to Louise is to make the most of the little time she has with her daughter. As she puts it “despite knowing the journey and where it leads, I embrace it and I welcome every moment of it”. Louise is truly courageous in the Aristotelian sense here.¹⁶ She endures the prospect of pain, faces it, indeed embraces it, out of love for her daughter and to help her own species not come to war with the

¹⁵ Zavota (2020) also mentions in passing how Louise’s knowledge brings immense tragedy. Her Derridean reading of *Arrival* however focusses on the primacy of writing and of the significance of the heptapod gift in the film.

¹⁶ Aristotle held that it is apt for the virtuous to be pained by the prospect of wounds and death. Courage consists in knowingly enduring such pains (*Nicomachean Ethics*: 1117b). The fact that Louise exhibits courage in this way rather makes a mockery of Aristotle’s suggestion elsewhere in *The Politics* (1260a22-23) that the courage of women is different from men. Louise shows audiences the obvious flaws in Aristotle’s belief system here—women clearly can possess courage in the same ways that men do.

heptapods. Images first establish Louise's bravery but audience perception of this is reinforced by verbal utterance. This is a good example of how words and images can combine into a moral lesson or principle capable of deepening the ethical understanding of the audience. The principle here being that it is ethically desirable to embrace life even when you know such an approach will bring pain.

The film could open up reflection or conversation about other aspects of ethical theory too. Louise's embrace-the-journey approach to life arguably embodies a rather Nietzschean ethic. She knows her life will involve great suffering but she affirms her fate anyway. When she asks Ian, "If you could see your whole life from start to finish would you change things?" this question is also put to the audience. Louise has already declared her intention to embrace the journey, but would viewers? This moment in the film, this possibility, brings to mind Nietzsche's famous doctrine of eternal recurrence.¹⁷ Maudemarie Clark (1990) suggests that eternal recurrence can be understood as a moral thought experiment in which one is invited to think about whether or not they would be willing to live their life over and over again, in exactly the same way, for eternity. One's reaction to this possibility can provide insight into the extent to which one values their life. While it is not my argument that an eternal recurrence thought experiment is intentionally screened in this film, I do think *Arrival* can provide viewers with an experience akin to this. In so doing it can help viewers interested in this doctrine, deepen their understanding of it. First and foremost, though, *Arrival* screens a tragic story artistically and this screen story has significant potential for ethics education. In sum, I believe *Arrival* is a very good tragic film as the plot involves familial suffering that the audience are encouraged to share in and ethically learn from.

Acknowledgments

I would like to thank the two reviewers of EuJAP for their helpful comments on an earlier version of this paper.


¹⁷ Nietzsche's (2001) most detailed thoughts on eternal recurrence are in *The Gay Science*.

REFERENCES

- Aristotle. 1990. *The Politics*. London: Penguin.
- . 1996. *Poetics*. London: Penguin.
- . 2004. *The Nicomachean Ethics*. London: Penguin.
- Carruthers, Anne. 2018. “Temporality, Reproduction and the Not-Yet in Denis Villeneuve’s *Arrival*.” *Film-Philosophy* 22 (3): 321-339.
<https://www.eupublishing.com/doi/full/10.3366/film.2018.0083>.
- Chiang, Ted. 2002. *Stories of Your Life and Other Stories*. London: Picador.
- Clark, Maudemarie. 1990. *Nietzsche on Truth and Philosophy*. Cambridge: Cambridge University Press.
- Friend, Stacie. 2007. “The Pleasures of Documentary Tragedy.” *British Journal of Aesthetics* 47 (2): 184-198.
- Kivy, Peter. 1997. “On the Banality of Literary Truths.” *Philosophic Exchange* 28 (1): 17-27
<https://core.ac.uk/download/pdf/233573912.pdf>
- Lamarque, Peter. 1995. “Tragedy and Moral Value.” *Australian Journal of Philosophy* 73 (2): 239-249.
<https://www.tandfonline.com/doi/abs/10.1080/00048409512346571>
- . 2020. “On Not Being Too Formalistic About Aesthetic Value.” *Journal of Comparative Literature and Aesthetics* 43 (1): 1-12.
https://jcla.in/wp-content/uploads/2020/11/JCLA-Vol.-43.1-Spring-2020_Peter-Lamarque.pdf
- MacAllister, James. 2023a. “On the Potential in Film for Ethics Education: In Defence of Educational Ethicism.” *Journal of Philosophy of Education*, 57 (1): 257-275.
- . 2023b. “Moral Learning Through Tragedy in Aristotle and *Force Majeure*.” *Journal of Aesthetic Education* 57 (1): 1-18.
- . 2024. “The Ethics of Mountain Adventure: Education from the Tragic and Sublime on Screen.” *Journal of Adventure Education and Outdoor Learning* 24 (1): 54-64.
- Nietzsche, Friedrich. 2001. *The Gay Science*. Oxford: Oxford University Press.
- Ruben Östlund, dir. 2014. *Force Majeure*. Sweden: Coproduction Office.
- Peedom, Jennifer, dir. 2017. *Mountain*. Sydney, Australia: Stranger Than Fiction.
- Ridley, Aaron. 2009. “Tragedy.” In *The Oxford Handbook of Aesthetics*, edited by Jerrold Levinson, 408-420. Oxford: University of Oxford.
- Rorty, Amelie. 1991. “The Psychology of Aristotelian Tragedy.” *Midwest Studies in Philosophy* 16 (1): 53-72.

- Villeneuve, Denis, dir. 2016. *Arrival*. Hollywood, US: Paramount Pictures.
- Williams, Bernard. 2011. *Ethics and the Limits of Philosophy*. London: Routledge.
- Zavota, Gina. 2020. "Given (No) Time: A Derridean Reading of Denis Villeneuve's *Arrival*." *Film-Philosophy* 24 (2): 185-203.

THE SUMMIT OF SAFE HORROR: DEFENDING MOST HORROR FILMS

Cara Rei Cummings-Coughlin¹ 

¹Morgan State University, USA

Original scientific paper – Received: 22/12/2023 Accepted: 20/08/2024

ABSTRACT

Many people regularly watch horror films. While it seems clear that sporadically watching horror films will not make us bad people, if it is the main type of media that we consume, then are we still safe? I will defend most horror films from Di Muzio (2006), who worries that we are harming our moral character by watching them. Most horror films (e.g., *Candyman*, *Get Out*, and *Scream*) fall into what I call the summit of safe horror (SoSH), the inverse of the uncanny valley effect, wherein almost-but-not-quite-human robots elicit discomfort from viewers rather than empathy. In the SoSH, violence elicits excitement rather than pity for the victims because the violence is mitigated by, among other things, comic relief and foolish choices by the characters. These narrative features allow most horror films to be intense enough to cause excitement and terror yet not so intense as to cause a negative moral attitude to form in our soul, because we feel what Aristotle would consider the appropriate amount of fear. Torture porn, a subset of horror films lacking plot and focusing solely on gore (e.g., the *Saw* sequels), falls outside of the SoSH because it lacks these narrative features, making the violence depicted too intense to be entertaining. These films outside the SoSH will not necessarily cause an inappropriate amount of fear but are simply the only ones that could possibly do so. Caution: spoilers ahead!

Keywords: comic relief; courage; fear; horror film; torture porn; uncanny valley.

Introduction

Many people regularly watch horror films. While it seems clear that sporadically watching horror films will not make us bad people, if it is the main type of media that we consume, then are we still safe? I will defend most horror films from Di Muzio (2006), who worries that we are harming our moral character by watching them. Most horror films (e.g., *Candyman* (1992, 2021), *Get Out* (2017), and *Scream* (1996)) fall into what I call the summit of safe horror (SoSH), the inverse of the uncanny valley effect, wherein almost-but-not-quite-human robots elicit discomfort from viewers rather than empathy. In the SoSH, violence elicits excitement rather than pity for the victims because the violence is mitigated by, among other things, comic relief and foolish choices by the characters. These narrative features allow most horror films to be intense enough to cause excitement and terror yet not so intense as to cause a negative moral attitude to form in our soul, because we feel what Aristotle would consider the appropriate amount of fear. Torture porn (e.g., the *Saw* sequels) falls outside of the SoSH because it lacks these narrative features, making the violence depicted too intense to be entertaining.

In section 1, I will define “horror film”, contesting Carroll’s definition that requires a monster. In section 2, I will present Gianluca Di Muzio’s critique of horror films. Among other things, he claims that horror films desensitize us to actual violence. In section 3, I will present my rebuttal: only a select few horror films are even capable of negatively affecting our moral character, and they are not guaranteed to have this result. In section 4, I will introduce the uncanny valley effect in robotics and its inverse, the SoSH. In section 5, I will show that most horror films contain narrative features that mitigate the violence. Torture porn, a subset of horror films lacking plot and focusing solely on gore, lies outside the SoSH, and, therefore, enjoying that violence *could* be problematic. In section 6, I will argue that repeatedly viewing films in the SoSH is safe to do without harming our moral character. Following Aristotle, emotions are not problematic, but feeling excesses or deficiencies of emotions is. The comic relief and illogical character actions prevent us from feeling an excess or deficiency of fear. The films outside the SoSH will not necessarily cause an inappropriate amount of fear but are simply the only ones that could possibly do so. Caution: spoilers ahead!

1. What qualifies a film to be horror?

I am using “horror films” to refer to a great deal of films. There are often distinctions made between thrillers, psychological thrillers, horror films, body horror, slasher films, found footage films, and more. I’m wary of such fine-grained distinctions because films often fall into more than one sub-category. Neil Martin concludes the same thing when he discusses the fuzzy boundaries between psychological thrillers and horror films (Martin 2019, 2). Moreover, he points out that these classifications are *post hoc* and I agree (Martin 2019, 4).

One classification of horror I disagree with is Noël Carroll’s. Carroll’s goal in *The Philosophy of Horror* (1990) is to examine art-horror, a distinct genre of media beginning around the time of the publishing of *Frankenstein* (1818). Films, books, plays, comics, magazines, and any other medium depicting horror imagery is art-horror. By contrast, war, Nazi atrocities, and what we are doing to our planet is natural horror (Carroll 1990, 12-13). This distinction is useful. However, Carroll claims that art-horror must include a monster, “a being in violation of the natural order, where the perimeter of the natural order is determined by contemporary science” (Carroll 1990, 40). Carroll admits that this excludes Hitchcock’s *Psycho* (1960). However, it excludes far more. If Norman Bates is not a monster, then neither are Leatherface from *The Texas Chainsaw Massacre* (1974), Pamela Voorhees from *Friday the 13th* (1980), Hannibal Lecter from *The Silence of the Lambs* (1991), or Ghostface from *Scream* (1996), to name a few. None of these characters are supernatural. Leatherface is merely physically imposing, Pamela Voorhees is driven mad by grief, Hannibal Lecter is highly intelligent, and Ghostface is simply a crafty duo. However, each of their respective films *are* horror films.

Some might argue that *Silence of the Lambs* is not a horror film. Martin considers it and films like *Henry: Portrait of a Serial Killer* (1986), *Rosemary’s Baby* (1968), *Misery* (1990), and *Black Swan* (2010) to be horror films as well (Martin 2019, 2-3). It isn’t just that they contain horror elements, but those elements are the primary focus. Martin contrasts this with films where horror elements are incidental, e.g., the ear-cutting scene in *Reservoir Dogs* (1992) (Martin 2019, 2). Others agree, calling *Silence of the Lambs* a paradigm example of modern horror because slasher films have psychopathic serial killers in place of monsters (Gaut 2002, 296).¹ While I agree with Berys Gaut, horror films

¹ See also Schneider (2003) for why *Silence of the Lambs* is a horror film.

do not need an on-screen, corporeal being as the antagonist because sometimes characters are fighting nature or something unseen.

I want “horror films” to align with common usage and capture as many films as possible. My positive view is that any film where (i) the supernatural, suspenseful, or violent scenarios are repeatedly inciting a fearful response in viewers as a central part of the plot of the film and (ii) the plot does not closely model a real-life scenario is a horror film. Many films have horror elements. However, the first half of the definition excludes children’s films where there is a frightful antagonist, and the second half excludes dramatizations of war and true crime documentaries.

The first conjunct of my definition echos Martin in that it requires the horror elements to be central to the plot. Horror elements include not only supernatural scenarios involving things like ghosts and demons, but also tension and violence caused by human beings. *Urban Legend* (1998) seems supernatural, but it is revealed to be a quest for revenge by a young woman whose boyfriend was killed. Unlike Carroll, I do not think that we need a monster. In fact, we need never see the antagonist. *The Blair Witch Project* (1999) is a horror film even though we never see the alleged Blair Witch. The characters are terrorized and scared constantly, making the audience feel fearful. Much the same is true of the *Paranormal Activity* (2007-2021) series. While many horror films have bleak endings, there can even be horror films with happy endings. Consider *The Conjuring* (2013-2021) series, where Ed and Lorraine Warren defeat the witch or demon and save the family. There are still several tense situations causing the audience to worry about who will survive, making the horror elements central to the plot.

The second conjunct rules out violence that would be too intense to be entertaining, e.g., war dramas and true crime documentaries. Films such as *Dunkirk* (2017) and *1917* (2019) depict death, but it is not to incite the audience to fear the enemy soldiers. Instead, we focus on the bravery and sacrifice of those involved. When I was younger, I was scared of Hannibal Lecter after watching *Silence of the Lambs* (1991), but not the Germans after watching *Schindler’s List* (1993). By “true crime documentaries” I mean to capture three things. First, actual true crime documentaries such as the *Conversations with a Killer* (2019) series. Second, serialized shows such as *Forensic Files* (1996-2011). Third, biographical crime dramas such as *Extremely Wicked, Shockingly Evil and Vile* (2019). One can be entertained by Leatherface chasing Sally, because there were no such people. Leatherface, like Hannibal Lecter, is loosely based on corpse defiler Ed Gein, but has been altered drastically.

We should not be entertained by Zac Efron (as Ted Bundy) lying to Lily Collins (as Elizabeth) about murdering women because Ted Bundy actually had a girlfriend named Elizabeth from whom he kept the murders a secret. Simulated violence has a different effect on us if we know that it is based on a true story.

My definition is going to include many films that could also be categorized as sci-fi and action, which is an upshot. Thrillers by Tarantino will be included, as will many others. I'm intentionally including a lot of films because I want to end up with as many completely defensible horror films as possible. So, I'm starting with a very inclusive definition.

2. Slashing compassion?

In this section I will present Di Muzio's (2006) critique of horror films and why he finds them to be morally problematic. Basically, he argues that viewing violent media will desensitize us to real-life violence and make us less compassionate.

Di Muzio begins his article by wondering why the moral status of pornography has been discussed at length, yet the moral status of horror films has not. If pornography is morally problematic, he says, then horror films are too. I disagree. The kind of pornography that is morally problematic is not similar to most horror films. Child pornography is problematic. However, problematic pornography bears no significant similarity to the horror films I will defend.

Specifically, Di Muzio likens Tobe Hooper's 1974 film *The Texas Chainsaw Massacre* to films about Nazi experiments and child torture. If we are unnerved by the latter, then we should be unnerved by the former and films like it that he calls "slasher" or "gorefest", where

[T]he narration turns on a series of murders, with special emphasis on the chase or struggle that precedes them, on the victims' wounds and loss of blood, and on their fear and despair. (Di Muzio 2006, 278)

The fact that the atrocities committed by the Nazis were real or that the victims in horror films are often teenagers or adults, rather than children, makes no difference to Di Muzio. I agree that films accurately depicting what Nazis did or torturing a child could be problematic. However, that has no bearing on the moral status of horror films like *The Texas*

Chainsaw Massacre. Even though Leatherface is loosely based on real-life murderer and corpse defiler Ed Gein, the story is changed so significantly that the analogy between *The Texas Chainsaw Massacre* and a film about Nazis falls apart.

Di Muzio's main concern is that people who derive joy from watching horror films lack the compassion necessary to make them moral agents. He believes that horror films show nothing but violence, leaving the audience unable to conceptualize what happened, because

[T]he moment the filmmakers attempt to provide some context for the violence, or begin to introduce a 'moral of the story', a slasher film starts losing its grip on the audience's emotions. (Di Muzio 2006, 290)

I disagree. The films that have the greatest effect on audiences introduce the characters thoroughly so that we care about losing them when they eventually die. A film that opens with the death of someone unknown to the audience is not going to be as effective or popular as one that has character development. Being familiar with the characters is one reason why sequels, remakes, reboots, and cinematic universes have become so popular. Di Muzio condemns even films with character development because the viewer

[M]ust at some point detach herself from the film's content and intentionally limit the degree to which she is affected by the stimulations of extreme suffering she is being exposed to. (Di Muzio 2006, 287)

I will argue that we need not detach ourselves from films in the summit of safe horror because the violence is mitigated by narrative features that allow us to know that what we are seeing is fabricated. As such, these depictions of violence will not desensitize us to actual violence.

3. Refuting Di Muzio

There have been other responses to Di Muzio that defend horror films. I will briefly discuss Marius Pascale's (2019) for two reasons. First, he offers responses to Di Muzio that are partially correct and, therefore, do not need to be recreated in my argument. Second, his defense is too permissive, meaning that he finds all horror films defensible.

Pascale focuses on reactive attitudes in his response to Di Muzio. First, he notes that Di Muzio oversimplifies by focusing on our need for sympathy and empathy without considering the “moral and psychological downsides of unmitigated sympathy or empathy” (Pascale 2019, 144). Indeed, we can be too sympathetic and fail to have appropriate boundaries with loved ones who seek to take advantage of us. Therefore, Di Muzio’s worry about becoming less sympathetic misunderstands that we should, as Aristotle says, have an appropriate amount of each emotion. Pascale also argues that Di Muzio “provides no evidence to support a link between slasher consumption and reactive attitude decay” (Pascale 2019, 144). Again, he is correct that a psychological study to back up Di Muzio’s claims would help. However, correlation does not imply causation, and even several studies would not convince me that horror films alone are to blame and not the state of the world in which we live. My only issue with Pascale is that he seems to defend horror films that I will not when he characterizes their consumers as having “morbid aesthetic indulgences” and “macabre fascination[s]” (Pascale 2019, 145). As I will argue, consumers who view horror films so that they receive all the benefits from engaging in such actions while suffering none of the consequences *might* be harming their moral character.

While Pascale makes some excellent points, his view is too permissive, because torture porn like *Hostel* (2005) and the *Saw* sequels would be perfectly fine to view repeatedly. Obviously, I want to be more permissive than Di Muzio, but not to this extent. The difference between Pascale’s view and my own is that I am wary of defending torture porn outright. I will not condemn these films, but neither will I defend them. My argument is as follows:

1. If a horror film contains narrative features that mitigate the violence, then it falls inside the summit of safe horror.
2. Most horror films contain narrative features that mitigate the violence.
3. Most horror films fall inside the summit of safe horror. [1, 2]
4. If a film falls inside the summit of safe horror, then it is safe to view it repeatedly without harming our moral character.
5. Most horror films are safe to view repeatedly without harming our moral character. [3, 4]

In the following sections, I defend premises 1, 2, and 4.

4. The uncanny valley of robotics and the summit of safe horror

The uncanny valley effect is the phenomenon wherein almost-but-not-quite-human robots elicit discomfort from viewers rather than empathy. I argue that there is an inverse of the uncanny valley effect where most horror films lie that I call the summit of safe horror (SoSH). We actually enjoy that violence because we know very well that it is fabricated.

Masahiro Mori first proposed the uncanny valley effect in 1970. He discovered that the more lifelike a robot gets the more positive emotions it elicits from us, except for those that are almost-but-not-quite-human. Robots with less than a 75% resemblance to humans received more and more favorable responses from the viewers the more they resembled a human. However, Mori found that robots made to look almost-but-not-quite-human elicited a negative response from viewers. Once they were indistinguishable from humans, the robots received positive responses again. Mori coined the term ‘uncanny valley’ to describe the sharp decline in an otherwise upward trend (Mori, MacDorman, and Kageki 2012, 99).

I propose that violence presented on screen elicits a more and more negative response, except in what I call the SoSH, where most horror films exist. See **Figure 1**. The x-axis represents how intense the on-screen violence is, and the y-axis represents the level of enjoyment that the audience feels upon viewing said media. The violence of films in the SoSH is intense enough to incite the excitement of the audience but not so intense as to incite the pity for the victims that we feel when we watch torture porn. Hence, the surge in enjoyment for media that is violent yet not overly violent, creating the summit.

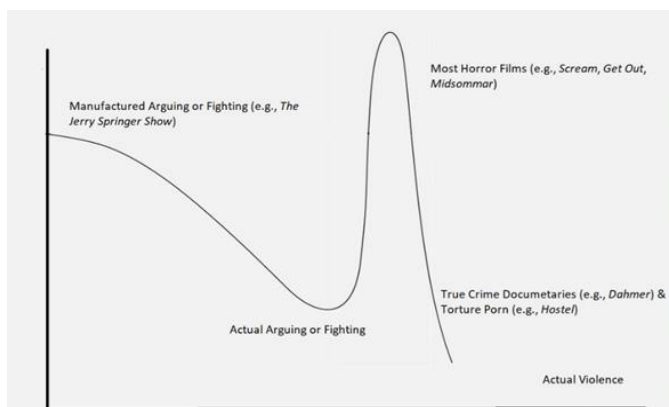


Figure 1

Figure 1 is just the uncanny valley graph rotated one hundred and eighty degrees with some media examples added. On the far left is manufactured fighting or arguing typical of reality television. Shows like *Jerry Springer* (1991-2018) and *Real Housewives* (2006-) are very enjoyable because the violence is very slight, mainly consisting of shouting and tossing drinks in someone else's face. Actual arguing or fighting shown on the news is less enjoyable. Consider watching police brutality that stops short of resulting in the victim's death. The sharp incline in enjoyment occurs because we particularly enjoy manufactured violence of the kind found in most horror films. The violence is mitigated by comic relief and illogical character actions, which signals to us that the violence is fabricated. So, it is fun to watch such scenes play out. The steep decline in enjoyment occurs for films outside of the SoSH, for they are too similar to actual violence.

Moreover, the inverse of the uncanny valley effect can be applied to all fiction, not just horror. There are summits of safe action, crime, and science fiction as well. For example, we enjoy watching characters like Mickey and Mallory in *Natural Born Killers* (1994), Cersei in *Game of Thrones* (2011-2019), Darth Vader in *Star Wars* (1977-), and Thanos in *Avengers: End Game* (2019). This too is perfectly natural and will not harm our moral character.

5. Films in and out of the summit of safe horror

In this section I will defend premise 2: most horror films contain narrative features that mitigate the violence. In 5.1, I will focus on two narrative features that mitigate on-screen violence: comic relief and illogical actions by characters. In 5.2, I will discuss the horror films that fall outside the summit of safe horror (SoSH): torture porn, which are films that lack plot and focus solely on gore.

5.1 Mitigated violence

I will now explain why most horror films fall into the SoSH. The violence is mitigated by narrative features which signal to us that it is fabricated, which is why we feel comfortable viewing it. While there are several such features, I will focus on two: comic relief and foolish actions on the part of the characters that the audience can critique.

Different things can mitigate on-screen violence for different viewers. Sanitizing the violence works for some people. For example, heroes might not bleed on screen, survive fatal wounds, or die off screen. Not

seeing the consequences of violence can make it easily digestible for some. As I like gore, I will not dwell on this narrative feature. Instead, I will discuss two features that are present in even the more graphic films.

Comic relief is not only present in comedy films, e.g., *Shaun of the Dead* (2004) or parodies of horror films, e.g., *Scary Movie* (2000). Horror films often contain moments of comic relief. Even very heavy films involving child abuse like *The Black Phone* (2021) have moments of comic relief. A Denver child abductor and killer known as the Grabber has taken Finney Blake. Meanwhile, his younger sister, Gwen, has been having premonitory dreams about the killer. She prays to Jesus for more clues in her dreams so that she can tell the police where to find Finney. After her dreams result in dead ends, Gwen gets down on her knees in front of her makeshift shrine/dollhouse and says “Jesus, what the fuck?” This line received a lot of laughs. The Grabber’s brother is also funny. He has been following the abductions and has deduced that the killer lives nearby his home. The police do not listen because he is overly enthusiastic and has been snorting cocaine. He seems like a conspiracy theorist until it is revealed that Finney is locked in his basement.

Another example of a horror film with comic relief is *Get Out* (2017). Rose and Chris are a young interracial couple. Rose lures Chris to her parents’ secluded home in upstate New York in order for them to suppress his consciousness and sell his body to the highest white bidder. Chris is generally nervous, adding to the audience’s tension. However, Chris’s friend Rod is a sassy TSA agent who is funny in all his scenes. He tells Chris and Rose jokes on their way to her parents’ house, urges the uncaring police to investigate the missing Black men, and he arrives at the end to save Chris. His scene at the police station is particularly funny. The detective does not believe his story about Rose’s family kidnapping, brainwashing, and turning young Black men into sex slaves. She even invites two coworkers to come listen to the allegedly ridiculous story and the three share a laugh. The audience laughs because such a scheme would be difficult to pull off. On the other hand, I can definitely see old white men bidding for which young Black body they would like to inhabit.

Comic relief is not new. Shakespeare employed comic relief in his tragedies in order to avoid overwhelming the audience. Humans have a complex range of emotions that we need to explore and indeed enjoy exploring. However, in order to make it through a sad movie, we need some moments of levity peppered throughout. The same is true of violent media. Comic relief gives us a break from all the death, but it is also good storytelling. A non-stop bloodbath is not as entertaining as a film where

the characters have comic, sad, and romantic moments. We learn about the characters in these breaks and begin to empathize with them and root for their survival, which makes losing them more difficult and their victories more meaningful.

In addition to comic relief, illogical character actions also mitigate the on-screen violence. Characters in horror films also do counterintuitive things all the time, especially when people around them are dying at an alarming rate. The *Candyman* and *Evil Dead* series are great examples of films where characters foolishly tempt fate. Candyman is summoned by saying his name five times in the mirror. In the original installment, Helen, being fond of urban legends, jokingly summons him. Candyman frames her for murder, and she is committed. Helen again summons him in a feeble attempt to prove her innocence, unsurprisingly causing more deaths, including her own. Much the same happens with characters in the *Evil Dead* series reading the Necronomicon. This book, written in blood and bound in human flesh, is always uncovered from a hidden place. Obviously, the book is dangerous. However, in each film, someone reads aloud from the book, which causes numerous possessions and deaths. While we do not expect that reading a book aloud will summon evil, we know that some things should be left alone. If I found a ring in a cemetery, I would not put it on my finger. Likewise, if people around me were dying, then I would not attempt to summon Candyman, even to demonstrate my alleged courage.

Also consider my favorite horror film franchise: *Scream*. In the original installment, Sidney Prescott receives a phone call from the killer who claims to be on her front porch. She calls his bluff by actually opening her front door and going outside. Later, when the town issues a curfew, the high school students decide to throw a party at Stu Macher's house. This is not mere teenage bravado. Instead, this stupidity is to advance the plot. We need the characters to be in one central location for the denouement to occur and Stu, being one of the killers, knows this. However, everyone except Billy and Stu should be afraid of Ghostface. I know that they are young, but this is reckless, even for teenagers.

The audience can be critical of characters in horror films because our world is not the same as theirs. As Pete Falconer argues, "a reasonably informed audience will have knowledge and expectations of a fictitious world that the characters inhabiting that world cannot access" (Falconer 2023, 2). They do not (usually) know that they are in a horror film. So, it is no surprise that they act foolishly. Falconer (2023, 3) would not agree with me calling them foolish, because their situations are so implausible that they do not seem to be real possibilities. However, it is not only from

the audience's privileged perspective that Sidney is being foolish, but from an in-universe one as well. This foolishness overrides our empathy and allows the audience to separate ourselves from the characters in the film. Since we would never act this way, we can criticize the characters, which is just like criticizing people on game shows or reality television for making mistakes. We might very well make the same mistake if stressed, but it is easy to shout at the television from the comfort of our living rooms.

Because the violence inside the SoSH is mitigated by various narrative features, the audience enjoys watching Ghostface kill everyone. Likewise, we enjoy watching Norman Bates of *Psycho* (1960), Leatherface of *The Texas Chainsaw Massacre* (1974), Damien Thorn of *The Omen* (1976), Carrie White of *Carrie* (1976), Michael Myers of the *Halloween* (1978-) series, the Xenomorph of the *Alien* (1979-) series, Jason Vorhees of the *Friday the 13th* (1980-) series, Freddy Krueger of the *A Nightmare on Elm Street* (1984-) series, Pinhead of the *Hellraiser* (1986-) series, Pennywise of *It* (2017, 2019), Hannibal Lecter of *The Silence of the Lambs* (1991), the titular Candyman, the titular Blair Witch, Samara Morgan of *The Ring* (2002), Alessa Gillespie of *Silent Hill* (2006), Sam of *Trick 'r Treat* (2007), Harry Warden of *My Bloody Valentine* (1981), Bughuul of the *Sinister* (2012, 2015) series, Bathsheba Sherman of *The Conjuring* (2013), the titular Babadook, Valak of *The Conjuring 2* (2016), Art the Clown of the *Terrifier* (2016-) series, Clown Mask of *Haunt* (2019), the Grabber of *The Black Phone* (2021), the titular Pearl, and the titular Abigail kill people. All these films contained violence that was mitigated enough to make them fall into the SoSH, which also includes many more films not mentioned here.

However, not all horror films are created equally. There are films that even hardcore horror fans like me watch only once or simply refuse to watch: torture porn, where the violence is so intense that no narrative features can mitigate it such that it feels fantastical.

5.2 Torture porn

In this section, I will employ Edelstein's (2006) phrase "torture porn" to refer to graphically violent films that lack comic relief and illogical character actions. I also defend this usage from Lowenstein's (2011) criticism that such a category does not exist. Torture porn does exist, and it does not fall into the summit of safe horror (SoSH) because such films are too intense to elicit excitement and instead elicit pity for the victims. I am referring to movies like Ruggero Deodato's *Cannibal Holocaust* (1980), Tom Six's *Human Centipede* (2009), and Srdjan Spasojevic's *A*

Serbian Film (2010). As these films are quite graphic, I do not want to discuss their content here. Instead, I will focus on a more palatable franchise that still falls outside of the SoSH: the *Saw* series.

The *Saw* sequels are particularly guilty of sacrificing narrative for more and more gruesome deaths. *Saw* (2004) presents Jigsaw's kills very briefly, as we spend most of the film with Dr. Gordon and Adam in the bathroom. At the end, there is a shocking reveal that Jigsaw has been in the room with them the entire time, posing as a corpse. Nothing this clever happens in the sequels. They are merely vehicles for increasingly violent deaths. In *Saw*, Paul Leahy dies of exsanguination because of the barbed wire maze he is placed in. This trap is specially designed for him because he had previously slit his wrist, allegedly for attention. Will he bleed again to save himself or did he really want to die? He must make a choice. The deaths in the sequels lack this poetic justice. In *Saw IV* (2007), Detective Eric Matthews has his head crushed by two blocks of ice, which has nothing to do with his crimes as a corrupt police officer.

Not only do the traps become more gruesome in the sequels, but it is hard to believe that anyone is supposed to survive any of these traps as they did in the original installment. In *Saw*, Amanda must retrieve a key from inside a live victim's stomach in order to unlock the reverse bear trap from her head. The other victim is anesthetized, which makes her job uncomfortable yet manageable. She succeeds and later returns in many of the sequels as one of Jigsaw's proteges. In *Saw 3D* (2010) two men and a woman, who has secretly been dating both of the men, are in a trap that is on display to the public. The two men must push saws towards each other or else the woman will be sawed in half. Unlike the original, there is no suspense as to whether someone will live or die. Instead, it is inevitable that one member of the trio will die. Indeed, the men, upset by her deception, agree to let the woman die.

Films like the *Saw* sequels fall under the category known as "torture porn", a phrase coined by David Edelstein. It is distinct from other sub-genres of horror, e.g., horror comedy or body horror, because the focus is on excessive gore. Edelstein focuses on *Hostel* (2005), *The Devil's Rejects* (2005), *Saw* (2004), *Wolf Creek* (2005), and *The Passion of The Christ* (2004). It isn't just that the gore is ramped up for Edelstein, but that the characters are too realistic:

Unlike the old seventies and eighties hack-'em-ups (or their jokey remakes, like *Scream*), in which masked maniacs punished nubile teens for promiscuity (the spurt of blood was equivalent to the money shot in porn), the victims here are

neither interchangeable nor expendable. They range from decent people with recognizable human emotions to, well, Jesus. (Edelstein 2006)

Notice that *Scream* is silly enough for Edelstein to contrast it with torture porn. Furthermore, he considers the characters in *Scream* to be expendable. Recall from the previous section that I noted several moments in *Scream* where we get frustrated with the character's stupidity. So, Edelstein would agree with me that comic relief and illogical character actions play a big part in making on-screen violence enjoyable.

Lowenstein argues that torture porn does not exist, and that such films are better described as spectacle horror. For Lowenstein, torture porn is "a needlessly brutal and graphically explicit image of mutilation" and spectacle horror is where "this investment in bloody, intense attractions is not gratuitous, but central to the pleasure of watching the film" (Lowenstein 2011, 46-47). He uses the scene in *Hostel* where Paxton is hiding on a cart of corpses and a severed hand that was blocking the cart's wheel gets placed by his face. Since this shot is from a point of view impossible for Paxton and the severed hand is not an important plot device, Lowenstein argues that this scene is specifically addressing the audience and how they enjoy this disgusting spectacle (Lowenstein 2011, 47). He objects to using "porn" "as a transparent term for self-evident moral disapproval [that] (...) ignores the possibility of spectacle horror's mode of feeling history as anything other than immoral and irresponsible" (Lowenstein 2011, 51). Lowenstein makes several comparisons between *Hostel* and the Iraq War that I completely missed, and I am now convinced that *Hostel* is a spectacle to reflect the brutality of the war and American privilege.

I agree that the brutality of *Hostel* is, in part, social commentary. However, unlike Lowenstein, I find that it leans too far into the gore and should be called "torture porn". Porn is not necessarily morally bad, even the sexual content typically categorized as such. It is a category of media that, like other categories, has morally good, morally bad, and morally neutral content. Torture porn is potentially problematic not because it is porn, but because of the unmitigated violence.

I am using "porn" as Thi Nguyen and Bekka Williams do, referring to something "engaged with for the sake of a gratifying reaction, freed from the usual costs and consequences associated with the represented content" (Nguyen and Williams 2020, 148). While they do not specifically mention torture porn, they differentiate between harmless

porn and harmful porn. Food porn, the viewing of images of expensive food, is harmless, while moral outrage porn, the simplification of an opponent's beliefs to magnify outrage, is harmful (Nguyen and Williams 2020, 149). So, contra Lowenstein, "porn" is a term that just refers to something we experience vicariously. Indeed, Nguyen and Williams mention baking porn, closet porn, headphone porn, and real estate porn. All of these are morally neutral. However, viewing excessive film gore because it offers all the excitement of committing violent acts without the risk of jail time will encourage a bad moral character to form.²

Let's return to the SoSH and why the *Saw* sequels and other torture porn falls outside of it, because it is not just about the graphic nature of the violence. There is no comic relief in the *Saw* franchise and people are murdered for all-too-common reasons, like doing their job as an insurance agent. Because Jigsaw is a main character, we are primed to feel pity for him that his cancer was not irradiated. However, people die of cancer all the time in the real world. Insurance agents simply assess claims and see if conditions are preexisting or not. There are no grounds for Jigsaw's vengeance against a lot of his victims. The only reason the audience would be on Jigsaw's side is because we know his backstory. If William Easton, the insurance executive from *Saw VI* (2009), were a main character whose backstory we saw, then we would not enjoy his death.

A film that supports this claim is *Drag Me to Hell* (Sam Raimi 2009). The main character, Christine, is a loan officer hoping to get promoted to assistant branch manager. She needs to show that she can be firm. So, she denies a third mortgage extension to an elderly woman, Sylvia. Sylvia curses Christine to be tormented and dragged to hell by the demon known as the Lamia. We follow Christine's story, as Sylvia dies early on. We are upset when Christine gets dragged to hell in front of her boyfriend at the train station, because she did what many of us would have done if in her shoes. If, however, we were following Sylvia's story, then we would root for Christine to get what's coming to her.

Drag Me to Hell falls in the SoSH because there is comic relief, and we are saddened that such an innocent mistake leads Christine to such a gruesome fate. *Saw VI* is not in the SoSH because the audience is led to feel bad for Jigsaw and not William. Moreover, *Drag Me to Hell* contains comic relief in the form of graphic scenes of Sylvia's corpse vomiting

² I will defend this claim in Section 6.

into Christine's mouth. These scenes are so over the top that they are played for laughs, something for which Raimi is famous.

6. Why the summit is safe

Let's take stock. We often take pleasure in viewing on-screen violence that is mitigated, i.e., violence which includes comic relief and foolish actions by the characters. However, we should not feel the same pleasure in viewing unmitigated violence. I will now explain why this is, and why our moral character is safe while viewing violence mitigated by narrative features. In 6.1 I will argue that, while feeling the Aristotelian mean of an emotion like fear is a good thing, feeling an excess or deficiency is when it begins to harm our character. In 6.2 I will argue that we watch and rewatch horror films not only to experience fear but because of the artistic value of clever storytelling, Easter eggs, and foreshadowing.

6.1 Aristotle and emotions

The idea that watching someone pretend to misbehave will make the audience want to misbehave dates back to Plato (*Republic* 391e-392a). While Plato might have a point about young, impressionable minds, I am concerned only with adult viewers. It is natural for adult viewers to enjoy mitigated violence like the kind that takes place in horror films inside the SoSH. Moreover, doing so will not negatively affect our moral character because emotions in themselves are not bad. Indeed, the appropriate amount of fear is good, because it keeps us alive.

In the *Nicomachean Ethics*, Aristotle tells us that emotions in themselves are not bad, but feeling too much or too little of them is:

I speak of moral character, for this is concerned with emotions and actions, of which there is an excess, a deficiency, and a mean. For example, one can be terrified, courageous, feel desire, anger, and pity, and experience pleasure and pain in general, either too much or too little, and in both cases wrongly. Whereas to feel these feelings at the right time, on the right occasion, towards the right people, for the right purpose and in the right manner, is to feel the best amount of them, which is the mean amount, which is the very thing that is excellence. (*NE* 1106b16-23)

So, being afraid can happen not enough, too often, or the right amount. Having the appropriate amount of fear and confidence is called courage

(*NE* 1115a6-7). Aristotle has a very specific definition of courage in mind: courage is shown when a man can defend himself nobly from death (*NE* 1115b3-4), because it is the most dreadful thing (*NE* 1115a26). So, only soldiers in his time could show courage. However, we now recognize that women and civilians can also be courageous. We can be courageous while fighting cancer, battling depression, or even watching a horror film (see also MacAllister this issue, 320).

I judge how good a horror film is by how nervous I am after I watch it. I don't like heights or driving fast cars, but I do like being scared by horror films. Following Aristotle, nothing is wrong with being scared so long as it is done in the right way, for the right amount of time, and in response to the right objects. I take it for granted that horror films are an appropriate object of fear and will focus on the appropriate amount and time.

Obviously, running in terror from fabricated on-screen violence is an excess of fear, because we are in no real danger at that time. Furthermore, doing so would cause us to be afraid of every raised voice and unexpected sound such that we would become cowardly. It would also be bad for our moral character for us not to react at all to the on-screen violence, for we could become desensitized to real-life violence. However, an increased heart rate, the desire to shut our eyes, and the feeling that something similar might befall us later is typical. I argue that this is the appropriate amount of fear to feel during a horror film.

Not fearing what we ought to fear is rashness. Aristotle says that excess fearlessness is madness (*NE* 1115b26). Someone who feared neither earthquakes nor floods would not be courageous, but rash (*NE* 1115b27-8). Indeed, our mortal flesh is no match for nature's power. Watching a horror film and then being inspired to go to a cemetery in the secluded woods in the middle of the night is not courageous; it is rash. The supernatural things we see in films will not occur, but we are making ourselves vulnerable to hungry animals or people with bad intentions that would find us alone. Moreover, if no one knows where we are, and there is no one around to find us, then we could get lost or hurt.

Having an appropriate amount of fear keeps us alive. Locking our doors is not cowardly, merely prudential. Wearing our seatbelt is not cowardly, we are merely attempting to avoid unnecessary bodily harm. A jolt in adrenaline from watching a scary movie reminds us that we are only human and can be undone in a multitude of ways. After *Final Destination 2* (2003), many drivers avoided being directly behind a log truck on the highway. Again, this is not cowardly, merely precautionary. Never driving again would be cowardly. Watching *Jaws* (1975) should make us wary of

swimming for a day or two, but never enjoying the ocean again would be cowardly.

The films outside of the SoSH depict violence so intensely that they might cause some people to feel the wrong amount of fear, either too much or too little. Consider the torture and murder of Josh in *Hostel* (2005). A Dutch businessman drills holes into his body and eventually slits his throat. Some people might react to such a scene as though they were viewing actual torture and get physically sick or cry. There is no comic relief, and the scene is very realistic. Certain audience members could be overwhelmed and traumatized by this scene. On the other hand, the very realistic nature of the scene might cause some people to become used to such graphic violence such that they would not react appropriately if this were actually taking place in front of them. This is exactly what Di Muzio is worried that all horror films do. However, the films in the SoSH will not do this. They are exciting enough to raise our heart rates, but not so much to overwhelm us.

The excitement of horror films is one reason for their popularity. However, there are other features that inspire repeat watching. In the next section, I will focus on two: foreshadowing and Easter eggs.

6.2 The art of horror

Di Muzio and others might object that if viewers aren't enjoying the simulated violence of horror films, then there is no reason to watch or rewatch them. However, Freeland (2000) argues that the audience appreciates the art form of filmmaking, not the violence itself. She proposes

[T]hat audiences do not simply relish the sorts of spectacles [she has] just described as a kind of vicarious violence. Rather, fans often look for the way that the numbers employ wit, parody, intertextuality, and cross-references to earlier films. (Freeland 2000, 262-263)

She is absolutely correct in her observation, because I rewatch horror films to search for hidden Easter eggs and foreshadowing all the time.

Alfred Hitchcock's cameos in his films are infamous. Whether he is walking by the main characters, playing cards on the train (as in *Shadow of a Doubt* (1943)), or his visage simply appears on a neon sign in the background (as it does in *Rope* (1948)), looking for him is a good reason to rewatch his films. Several other directors follow in his footsteps and

have their own cameos in their filmographies. Sam Raimi's 1973 Oldsmobile is practically a character in the *Evil Dead* series, and it also shows up in most of the other movies he directs or produces. James Wan hides Billy, the puppet from the *Saw* series, in a host of films he directs. *Scream*'s director, Wes Craven, appears as a janitor in Freddy Krueger's sweater in the original 1996 installment, which is a nod to his *Nightmare on Elm Street* series.

Observant fans can also benefit from clever foreshadowing horror films include. In *I Still Know What You Did Last Summer* (1998), Julie's friend Karla wins tickets to the Bahamas from a radio show. However, the trivia question she was asked is "What is the capital of Brazil?" and she incorrectly answers "Rio de Janeiro". So, we know early on that they have been lured to the island by the killer. It is no coincidence that Julie is the target of another attack; this incident is directly related to the last film. In *The Witch* (2015), twins Mercy and Jonas sing an ominous song about the family's goat, Black Phillip, that alludes to the ending where he turns out to be the devil. In *The Conjuring 2* (2016), Valak's name is spelled out in the background of multiple scenes, and knowing the demon's name is its ultimate undoing. The opening of *Midsommar* (2019) focuses on a tapestry that appears to show the end of winter and the coming of spring. It actually outlines the entire plot of the film, even Christian's fate inside the bear. Hidden details like this promote repeat viewing, not only the gore of the deaths.

Horror films are art, which contradicts Gaut's claim that most horror films lack artistic worth (2002, 298). So much creative effort goes into the acting, writing, special effects, cinematography, etc. that should not be ignored. Horror films are so diverse that to assume they are all slashers with no overarching theme or moral lesson is foolish. *Dawn of The Dead* (1978) ends with Fran and Peter in a helicopter escaping a mall overrun by zombies. Seconds earlier, Peter had contemplated taking his own life. Even if there is no safe place left for them, they do not give up. This is a life lesson that makes the film go from good to great.

Conclusion

I have defended most horror films from the claim that watching them makes us develop a bad moral character. I offered my view that most horror films fall into what I call the summit of safe horror (SoSH), wherein staged violence elicits excitement rather than pity. My view is rooted in the uncanny valley effect in robotics. I left torture porn out of the SoSH because the violence they depict is not mitigated, i.e., there is

no comic relief, and no foolish choices are made by the characters. As a result, we might not be able to repeatedly watch their antagonists commit atrocities because the real-world consequences are shown, and it could desensitize us to actual violence. We can, however, watch the antagonists of films in the SoSH be violent because the violence is clearly fictitious. Enjoying films in the SoSH is natural and will not desensitize us to actual violence.

Acknowledgments

I owe my thanks to Ryan Ross, the organizers of the Aesthetic Education and Screen Stories conference at the University of Rijeka, my many students and colleagues who have endured the presentation of these ideas, and the anonymous reviewers for their helpful comments on previous drafts. I especially owe my thanks to my husband, Vince, for always taking me to the movies to “do research”.

REFERENCES

- Aristotle. 1926. *Nicomachean Ethics*. Cambridge: Harvard University Press.
- Aster, Ari, dir. 2019. *Midsommar*. 2019; New York, US: A24
- Bousman, Darren Lynn, dir. 2007. *Saw IV*. Vancouver, Canada: Lions Gate Film.
- Carroll, Noël. 1990. *The Philosophy of Horror*. Oxfordshire: Routledge.
- Craven, Wes, dir. 1996. *Scream*. Toronto, Canada: Woods Entertainment.
- Derrickson, Scott, dir. 2021. *The Black Phone*. Los Angeles, US: Blumhouse Productions.
- Di Muzio, Gianluca. 2006. “The Immorality of Horror Films.” *International Journal of Applied Philosophy* 20 (2): 277-294.
- Edelstein, David. 2006. “Now Playing at Your Local Multiplex: Torture Porn.” *New York Magazine*. January 26, 2006. <https://nymag.com/movies/features/15622/>
- Eggers, Robert, dir. 2015. *The Witch*. New York, US: A24.
- Falconer, Pete. 2023. “Being in a Horror Movie.” *The Journal of Aesthetics and Art Criticism* 20: 1-13.
- Freeland, Cynthia A. 2000. *The Naked and the Undead*. Oxfordshire: Taylor and Francis.
- Gaut, Berys. 2002. “The Paradox of Horror.” In *Arguing about Art: Contemporary Philosophical Debates*, edited by A. Neil and A. Ridley, 317-330. London: Routledge.

- Gillespie, Jim, dir. 1997. *I Know What You Did Last Summer*. Los Angeles, US: Mandalay Entertainment. Film.
- Greutert, Kevin, dir. 2009. *Saw VI*. Vancouver, BC, Canada: Lions Gate Film.
- . dir. 2010. *Saw 3D*. Vancouver, Canada: Lions Gate Film.
- Hooper, Tobe, dir. 1974. *The Texas Chainsaw Massacre*. Austin, US: Vortex, Inc
- Lowenstein, Adam. 2011. "Spectacle Horror and *Hostel*: Why 'Torture Porn' Does Not Exist." *Critical Quarterly* 53 (1): 42-60.
- Martin, G. Neil. 2019. "(Why) Do You Like Scary Movies? A Review of the Empirical Research on Psychological Responses to Horror Films." *Frontiers in Psychology* 10: 1-22.
- Mori, Masahiro, Karl F. MacDorman and Norri Kageki. 2012. "The Uncanny Valley [From the Field]." *IEEE Robotics & Automation Magazine* 19 (2): 98-100.
<https://doi.org/10.1109/MRA.2012.2192811>
- Nguyen, C. Thi and Bekka Williams. 2020. "Moral Outrage Porn." *Journal of Ethics and Social Philosophy* 18 (2):147-172.
- Pascale, Marius A. 2019. "Art Horror, Reactive Attitudes, and Compassionate Slashers." *International Journal of Applied Philosophy* 31 (1):141-159.
- Peele, Jordan, dir. 2017. *Get Out*. Los Angeles, US: Blumhouse Productions.
- Plato. 2013. *Republic, Volume I: Books 1-5*, ed. C. Emlyn-Jones and W. Preddy. Cambridge, Mass.: Harvard University Press.
- Raimi, Sam, dir. 1981. *The Evil Dead*. Burbank, US: New Line Cinema. Film.
- , dir. 2009. *Drag Me to Hell*. Universal City, US: Universal Pictures.
- Romero, George A., dir. 1978. *Dawn of the Dead*. New York, US: United Film Distribution Company.
- Rose, Bernard, dir. 1992. *Candyman*. Culver City, US: TriStar Pictures.
- Roth, Eli, dir. 2005. *Hostel*. Vancouver, Canada: Lions Gate Film.
- Schneider, Steven J. 2003. "Murder as Art/The Art of Murder: Aestheticizing Violence in Modern Cinematic Horror." In *Dark Thoughts: Philosophical Reflections on Cinematic Horror*, edited by S. J. Schneider and D. Shaw, 174-197. Lanham: Scarecrow Press
- Wan, James, dir. 2004. *Saw*. Vancouver, Canada: Lions Gate Film.
- , dir. 2016. *The Conjuring 2*. Burbank, US: New Line Cinema.

CRINGE OVERHANG: THE PERLOCUTIONARY EFFECTS OF CRINGE COMEDY

Alexander Sparrow¹ 

¹Falmouth University, United Kingdom

Discussion – Received: 26/02/2024 Accepted: 30/08/2024

ABSTRACT

Cringe comedy can make people so uncomfortable, the cringe continues even after the comedy has stopped. This paper explains this effect, which I call “cringe overhang”. If audiences weakly connect to the characters, they laugh. If audiences strongly connect, they have a negative emotional response—say, struggling to watch, or wanting to leave the room.

Firstly, I argue cringe comedy jokes are illocutionary acts designed to provoke laughter through second-hand embarrassment (Austin 1975). Secondly, I acknowledge that these jokes don’t always produce the desired perlocutionary effect of laughter—sometimes the joke is unable to cut through the embarrassment, merely leaving the viewer in a state of discomfort. Thirdly, drawing on the benign violation theory of McGraw and Warren (2010), I explain that the surplus of embarrassment is due to maximising the violation in the comedy while adding comparatively little benign. Finally, I argue that cringe comedy’s funniness is reliant on its lack of social psychological distancing. By leaving no room between the viewer and the character, embarrassment is maximised, the comedy is less benign (i.e. a stronger violation) and more polarising as a result. This explains i) why cringe comedy produces a comedic “overhang” in some viewers, where they continue to cringe even after the comedy has stopped, and ii) why cringe comedy produces a laughter response in some audiences, and stress responses in others.

Keywords: speech acts; humour theory; benign violation theory; cringe comedy.

1. Introduction

“Cringe comedy” is a subgenre of comedy that uses the audience’s second-hand embarrassment to provoke laughter in the viewer (Hye-Knudsen 2018; Mayer, Paulus, and Krach 2021). This method can make some viewers so uncomfortable, it produces the side-effect of “cringe overhang”: a term I have coined to describe instances where the cringe continues for the viewer even after the comedy has stopped. It is possible to use J. L. Austin’s speech act theory to explain that jokes are subtextual invitations to laugh. This is why cringe comedy jokes that produce only second-hand embarrassment can be said to have failed: the subtextual invitation has been refused and laughter has not been provoked. Having used Austin’s work to explain *why* a joke has failed, I then use Peter McGraw and Caleb Warren’s benign violation theory to explain *how* the jokes have fallen short. They fail to balance the strong violation of second-hand embarrassment with sufficient “benign”, a noun used to denote the editing a joke can undergo to make its violations simultaneously interpretable as safe, playful, or nonserious. This lack of benign leads to a negative emotional response in place of laughter. If audiences feel *some* second-hand embarrassment for the characters, they laugh. If audiences feel a *strong* second-hand embarrassment, they have a negative emotional response—say, struggling to watch, or wanting to leave the room. When combined, the two theories explain how cringe comedy works, and one of the pitfalls of failure: inducing cringe overhang in some audience members.

The UK sitcom *Peep Show* (Armstrong and Bain 2003-2015) is a perfect cringe comedy case study: shot from the point of view of its two main characters, it also presents their inner monologues, allowing for their embarrassment to be felt *and* “thought” by the viewer. This will be the main cringe comedy I draw examples from, though my findings apply to cringe comedy in all performance capacities, i.e. where the cringe humour is created *deliberately* for an audience, as opposed to situations in everyday life that happen to be both embarrassing for one person and humorous for those who observe it. People sometimes laugh through second-hand embarrassment in response to real-life events, i.e. non-performative and unintentional ones. The difference in intent between comic performer and embarrassed individual, where the former’s acts are purposeful and the latter’s are accidental, may mean a different theoretical explanation is needed. Accordingly, I have focussed on intentionally performed cringe *comedy*, rather than accidentally enacted cringe *humour*.

In order to explain the phenomenon of cringe overhang, I explain why cringe jokes can be understood to have failed, how that failure happens, and why that failure leads to cringe overhang. Firstly, I argue cringe comedy jokes are illocutionary acts designed to provoke laughter through second-hand embarrassment—explaining *why* cringe jokes that don't produce laughter have failed (Austin 1975). Secondly, I draw on the benign violation theory to explain *how* cringe jokes can fail: by maximising the violation and minimising the benign, leading to an excess of embarrassment (McGraw and Warren 2010). Thirdly, I acknowledge that these jokes don't always produce the desired perlocutionary effect of laughter: sometimes the joke is unable to cut through the embarrassment, and results in cringe overhang, leaving the viewer in a state of lasting discomfort. Finally, I argue that cringe comedy's funniness is reliant on its *lack* of social psychological distancing. By maximising second-hand embarrassment, the comedy is less benign (i.e. a stronger violation) and more polarising as a result. This explains why cringe comedy produces a cringe overhang in some viewers, where they continue to cringe even after the comedy has stopped—and why this phenomenon is an unsurprising, and for some audiences *unavoidable*, part of this comedic subgenre.

2. Cringe comedy as illocutionary acts and subtextual actions

In his seminal work on the philosophy of language, *How to do Things with Words*, J. L. Austin coined the term “illocutionary force”. This is where the speaker makes an utterance that demands something of the listener through subtext. Take the phrase, “The oven's off.” That could of course mean that the oven hasn't been switched on—but this sentence implicates that the speaker's intention is to get the listener to turn the oven on. The illocutionary force of an utterance provides a subtextual invitation for a listener to act on, which goes beyond the literal spoken words (Austin 1975, 99-100). This is a helpful way to think of jokes in comedy. Building on Austin's work by applying it to comedy, I define “joke” as an illocutionary act inviting the audience to laugh.

Comedic actions are similar to jokes. They use subtext to *invite* the viewer to laugh. They do not tickle the audience and force them to laugh, nor do they actively ask or command the audience to laugh. They present the audience with characters performing actions which, like jokes, use subtext to provoke the desired laughter response. In the first episode of season one of the UK sitcom *Peep Show*, the character Mark Corrigan is observed by his office crush chasing children and threatening them with a metal pipe (Armstrong and Bain 2003). This happens because he finally

snaps after the gang of youths have repeatedly harassed him over a period of several days. The viewer is not meant to sit and indifferently take the child chasing scene in; nor is the scene supposed to produce emotional distress or tears or a desire to hide (like a similar scene in a horror film might). Instead, Mark's action of chasing the children—and unknowingly embarrassing himself in front of Sophie, the woman he is infatuated with—is designed to invite the viewer to respond by laughing.

So far, we know comedy is based around a desire to provoke laughter; either with illocutionary force from performed jokes, or through a subtextual invitation from performed actions. Note that I specify comedy must be performed. This helps differentiate physical comedy, like slapstick, from situations in everyday life where a person undertakes an embarrassing action that unintentionally provokes laughter. In comedy, if you chase a child with a metal pipe, that can be funny. In real life, if you chase a child with a pipe, it takes on an entirely different tone. The former is done with the intention of provoking laughter in an audience, while the latter generally isn't. Of course, onlookers may still find real-life pipe wielding child-chasing funny. It is important to note again that the focus here is on performed cringe comedy. The clear difference in intent between comic performer and embarrassed individual could mean there are also differences between audience members and real-life onlookers, e.g. a lack of empathy in the latter. My investigations have focussed on performed comedy, so I make no arguments about everyday events.

Cringe comedy is made up of joke illocutionary acts, and performed subtextual comedic actions, which are designed to provoke laughter through second-hand embarrassment. The subgenre tends to depict characters suffering through embarrassment *in the moment*, rather than recounting the emotionally distressing scene after the fact. It seems that an important feature of cringe comedy is getting the audience as close to the embarrassment as possible. This is easier when the audience sees the embarrassing joke delivered, or action undertaken, in real time. An episode of *Peep Show* where Mark had instead chased the children with a pipe *offscreen* and then had the story recounted to him afterwards by Sophie may still have been funny. However, the embarrassment and the laughter it provokes by actually watching the action play out heightens the effect produced. Embarrassment, like other emotions, is heightened through direct experience. The goal is to make the audience empathise or feel the character's embarrassment so deeply that it afflicts them like they are directly experiencing it themselves. This discomfort can provoke laughter.

In this section, I have defined cringe comedy as a comedic subgenre where second-hand embarrassment is used to provoke laughter in the viewer. In the next section, I will outline the benign violation theory and show how it can be used to explain how jokes work. This is important for defining and explaining cringe overhang because benign violation theory will later be used to show how cringe comedy jokes can fall short of their laughter-provoking purposes.

3. Benign violation theory

Expanding on prior work by Tom Veatch (1998), Peter McGraw and Caleb Warren argue in *Benign Violations: Making Immoral Behaviour Funny* that all instances of humour are brought about through the “benign violations” that occur when an utterance or an action is perceived as violating our expectations in a benign way (2010, 1142). A “violation” is anything that threatens our idea of how the world “ought to be”, by being unsettling or disquieting. This can be through breaking taboos or moral norms, but extends much more broadly to our expectations of daily life in all areas like language, how we move, table manners, etc. “Benign” means “safe, playful, [or] nonserious” (2010, 1142). They emphasise that violation and benign interpretations must be perceived simultaneously in order to be considered humorous.

Examples drawn from various types of joke will help illustrate the benign violation theory in action. Firstly, I’ll use a pun: “*I like religious humour, but catholic jokes make me cross.*” This joke violates our expectations of language use, because “cross” can mean both i) angry, and ii) a religious gesture. It is sufficiently benign because many of us lack ties to the catholic faith. Those who do have ties recognise that the punchline makes light of a quirk of language, rather than the religion itself.

In a slapstick comedy film, a clown might fall down a well—then after some seconds, suddenly pop back up again, surprisingly unharmed. Here, the fall violates our expectations of how the world ought to be in two ways. Firstly, we don’t expect people to fall down wells, especially unintentionally and out of the blue. This violates our expectations of the clown’s awareness of their surroundings (as well as our expectations of well-placement, assuming it’s somewhere unusual, like the middle of a busy street). Secondly, we don’t expect such a dangerous feat to be pulled off without injury—violating our expectations of physics. It is made sufficiently benign when it is revealed no real harm has been done, in spite of our very reasonable assumptions otherwise.

In short, the benign violation theory is a balancing act. It requires a violation of our expectations to be made playful. This is easy in the joke and slapstick comedy above, because the violations aren't that intense. When the violation comes across a lot stronger, say, invoking a recent tragedy or morally questionable stance, increasing the benign takes more work. "Temporal, social, spatial, likelihood, [and] hypotheticality" are means of psychological distancing comedians can draw on to render a violation simultaneously benign (McGraw and Warren 2010, 1146). These methods are meant to create a feeling of space between the violation and the audience—allowing room for the violation to be interpreted as safe, playful, or nonserious. I will briefly explain how each of these methods work, along with examples of jokes that make use of them.

i) Temporal psychological distancing is where the joke-teller uses the passage of time to create a playful or nonserious buffer around the violation. In other words, it's easier to make a joke about a tragedy sufficiently benign when said tragedy took place in the distant past. An example follows: "*My house flooded the other day. It was like Hurricane Katrina, but worse—because I knew the people affected.*" Hurricane Katrina was a tragedy that claimed hundreds of lives nearly 20 years ago. The passage of time has weakened audience responses to the violation, allowing it to become fodder for jokes.

ii) Social psychological distancing is where the joke-teller makes jokes about people the audience do not know, or do not have strong ties to. Audiences are less likely to feel the violation as strongly when they don't personally know the joke targets. For example: "*Fame affects people differently. George Clooney has aged like a fine wine; Johnny Depp has aged like my grandma.*" Many people know of American actor Johnny Depp, but few know him personally. This social distancing allows us to enjoy jokes at his expense without feeling the violation too strongly.

iii) Spatial distancing relies on the topic of violation being far away from the audience. Because the violation isn't in close vicinity, it is less likely to have personally affected any given audience member—thereby increasing its chance of being sufficiently benign. One example is, "*I was doing a gig at a university and everyone thought I was American. You'll never guess how many people I shot.*" This uses the many school and university shootings in the USA over the last several years as joke fodder. The joke worked at my New Zealand gigs because the United States is so far away, audiences didn't feel the mass shooting violation as strongly.

iv) Likelihood psychological distancing is where the joke-teller hopes the violation within the joke comes across as somewhat unlikely. If the audience doubts the veracity of the joke's contents, it will be easier to laugh at its unpalatable imagery. For example, "*My friend's dad was hit by a car. She was distraught. He was crushed.*" Here, the flippant treatment of the topic at hand, as well as the silly pun the punchline relies on, makes it seem unlikely the car accident (the violation) actually happened. Because we doubt anyone was hit, the audience feels the violation has been made sufficiently benign.

v) Hypotheticality psychological distancing entertains a given violation, without delivering it like an apparently truthful anecdote. Because it's just an "imagine if" scenario, audiences can interpret it as benign. For example: "*Teenagers wanting to be rebellious always get piercings without their parents' permission. Imagine piercing your parents without their permission! That's rebellious.*" This joke doesn't suggest non-consensual parental piercing has happened; it just asks audiences to *contemplate* the idea. This keeps it sufficiently playful/benign.

Simple and precise, the benign violation theory seems to best encapsulate the various occasions of humour—ranging from tickling and slapstick comedy, to puns, blue humour, and workplace humour. In short, it explains enough instances of humour to make offering tweaks and improvements to perfect it a worthwhile exercise.

There are some minor issues with the benign violation theory. Some of McGraw and Warren's suggestions for achieving psychological distance work better than others. They later amended their arguments about temporal distancing, attempting to add some precision to the time limit within which a joke can be predicted to work (McGraw, Williams, and Warren 2014, 566). It can be argued that their "improvements" fall short by using psychological analysis of audience responses to explain when jokes should be performed. Instead, they could use the theory more practically to explain how jokes can be adjusted so they can be performed *at any time*.

Some thinkers take issue with the theory due to the power inequality that can exist between the joke-teller and the listener (Kant and Norman 2019).

Social psychological distancing is also readily open to misuse, making jokes less effective, rather than funnier. When a joke target is in the same space as the audience, their reaction is observable. When a comedian regales an audience with stories of joke exchanges targeting people who

aren't there, hoping social psychological distancing will make the joke sufficiently benign, likelihood and hypotheticality distancing risk also unintentionally undermining the joke. When the joke target is somewhere else, the audience has reason to doubt whether the exchange took place (likelihood) and if this is just a scenario made up by the comedian (hypotheticality). Sometimes comedians must joke about people in the same room—and there's a good chance of producing jokes that are less funny if social psychological distancing is rigorously adhered to.

In defence of the benign violation theory, Kant and Norman's concerns of power differentials between joke-teller and listener are more relevant to a social or workplace setting, say, where the joke teller is the boss and the listener a member of their staff. Similarly, the issue I have with social psychological distancing is most likely to be an issue specifically when applied to stand-up comedy.

Despite these criticisms, the benign violation theory can be used to explain cringe comedy, how it works, and why attempts at cringe comedy sometimes fail. In the next section, I illustrate how the benign violation theory is able to do this, by creating a benign-violation graph on which cringe comedy jokes can be placed and moved as they are made more or less benign.

3.1 Benign violation theory in action

As stated earlier, for a joke to provoke laughter, it must violate our expectations of how things ought to be, while *simultaneously* being open to interpretation as safe and playful. This usefully restricts comedic writing by providing parameters within which we can place any given joke—including those used in cringe comedy. The theory can be illustrated with a graph showing level of violation on the vertical axis, and the increasing amount of psychological distancing on the horizontal axis. An effective cringe comedy joke will find the point where these lines intersect, ensuring the joke is benign-violation balanced.

A successful joke can neither be solely benign, nor solely a violation. If an utterance landed on either of these extremes, it would lack the simultaneous bite or playfulness, respectively, to form the unique wordplay agreeable to laughter. A comment like, "Gosh, it's chilly today", is solely benign. Yelling sexist remarks in the street is solely a violation (Sparrow 2024, 32). Neither provokes laughs. To be clear, a failed joke that ends up accidentally falling at either extreme, does not automatically become a statement or opinion in light of the fact it failed to produce laughter. While that should be taken into account, it also

doesn't mean any offensive remark that fails to produce laughter can be defended as a failed joke: the structure, timing, and location of an utterance usually provides sufficient clues as to whether someone is a poor joke-teller, or trying to avoid responsibility for their offensive speech.

Different joke balances are illustrated in **Figure 1** below.

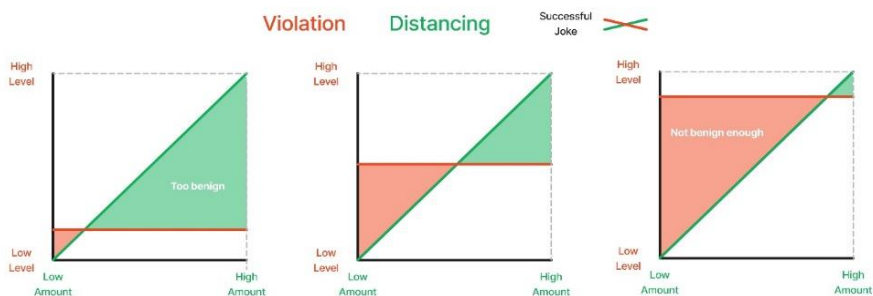


Figure 1

The intersection of violation and benign is where a joke balances. As the graphs show, generally speaking, the stronger the violation, the more psychological distancing is required to effectively balance it. While jokes are subjective, and the strength of a given violation varies from person to person, the graph can still help comedians and thinkers visualise what changes need to be made if a joke isn't working. The vertical axis shows the strength of the violation. Weaker violations, like jokes about work or the weather, will sit low. Stronger violations, like mentions of a current war or sexual assault, will sit higher. The horizontal axis shows the amount of psychological distancing used. If very little effort has been made in creating psychological distance, the joke will be towards the left. If much effort has been made to create psychological distancing—say, social, temporal, and likelihood distancing have *all* been used—then the joke will move to the right.

Jokes fit under various comedic subgenres, all of which balance the extremes of benign and violation differently. Comedic subgenres (from now on referred to as subgenres) will, generally speaking, align more closely with either benign or violation. For example, puns can often be illustrated with **Figure 2**.

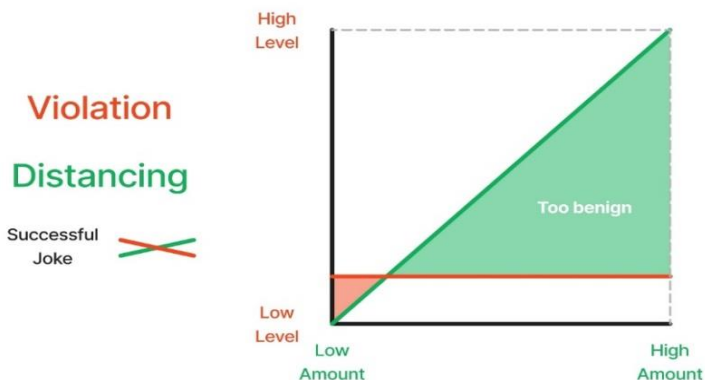


Figure 2

Puns violate our expectations of language use, and the way we understand words. Although puns can be used in violation-heavy subgenres like blue humour and gallows humour, we tend to think of them as comparatively light or silly (Sparrow 2024, 33). Because of this, puns will often be placed low on the violation axis, and require less psychological distancing to make them sufficiently benign.

In contrast to this, stronger violations, as shown in **Figure 3**, tend to be present in jokes about more challenging or controversial topics like war and sexual violence.

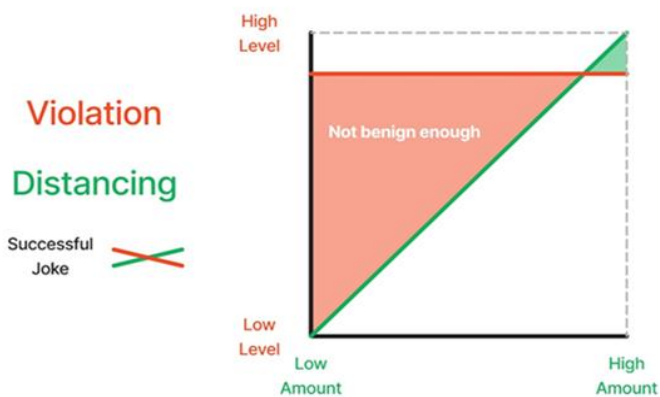


Figure 3

Jokes in the subgenres of dark, gallows, blue, and political humour are typically thought of as stronger violations—meaning they are higher on

the vertical axis, and require more psychological distancing to make them sufficiently benign.

The benign violation graph provides a basic framework for how jokes can be fixed when they fail to provoke laughter. Jokes can fail for a wide variety of reasons, such as unclear premises/set-ups, using too many words, and having predictable punchlines. If structural tweaks aren't providing any answers, audience response to jokes can provide useful clues as to whether the joke is appropriately balanced. If they angrily leave mid-show, or heckle, the violation could be overpowering the joke. By identifying that a joke is too violation heavy, a comedian can work to increase the amount of psychological distancing, thereby balancing the joke. If audience members are constantly checking their watches, or are on their phones, the joke might be too benign, lacking the violation necessary to cause surprise and laughter. To remedy this, the comedian can reduce the psychological distancing to allow for the violation to be more obvious (Sparrow 2024, 34).

Having shown the benign violation theory in action, I will now demonstrate how cringe comedy relies on strong violations to produce effective jokes.

3.2 Benign violation theory's use of violation

Strong violations occur in cringe comedy, many of which, if delivered with anything other than deft control, can come across too strongly for audiences. Comedians—a term I'm using for directors, writers, actors, and stand-up comics—can work to increase the playful, nonseriousness of the benign in a given joke in order to make the violation more palatable. This is more effective than watering the violation down or removing it entirely. A clear example of this is the sixth and closing episode of the third season of *Peep Show*, where Mark realises while on a weekend with Sophie that he is not in love with her, and they are not remotely compatible (Armstrong and Bain December 16, 2005). He realises this while lost in a field on a weekend away where he had intended to propose to her. When he finally manages to get back to the hotel, however, Sophie has found the ring and accepts the “proposal”, even though one was never offered. Mark “accepts the acceptance” out of sheer embarrassment and discomfort at the prospect of telling her he actually wants to end the relationship. The violation here is a man struggling to express his emotions and being trapped by them as a result. Rather than watering down the violation of a loveless marriage, the writers of the show leaned into it by showing how such marriages can begin—out of embarrassment, or a misplaced sense of duty, or from a

perceived lack of alternatives. As a season closer, the prospect of an entirely avoidable miserable life ahead is poignant, as well as hilariously embarrassing.

While benign violation theory is not an attempt to explain away all violation-heavy joke-attempts as important and worthwhile social criticism, it can allow comedians a framework for sensitively approaching polarising, controversial, or otherwise difficult topics in an appropriately comedic way. Comedians need violations like the one above if they are to produce jokes capable of criticising or making light of the status quo, questioning the rationality of our commonly held beliefs, and thinking of novel solutions to difficult problems. Without the bite of a violation flipping our world upside down for a moment, a joke collapses.

In this section, I defined the benign violation theory, explained its parameters with a graph upon which jokes can be placed and moved according to how benign- or violation-heavy they are, and outlined the essential part violations play in jokes. In the next section I show how cringe comedy in particular can fail to be sufficiently benign-violation balanced.

4. Cringe comedy's failed perlocutionary result

As previously mentioned, illocutionary acts are utterances spoken to get the listener to perform an act. If I want you to open your present, I might say, "*You haven't opened your present*". The illocutionary force of the utterance is intended to get you to open said present. The perlocutionary response is what the listener does in response to the illocutionary act. If you open the present after I tell you that you haven't opened it, that is the perlocutionary result. If you say, "*You can open it if you want*", that is the perlocutionary result. Note that in the second example, the perlocutionary result does not match the illocutionary force, i.e. my utterance has failed to garner the desired result.

Failure to achieve the desired perlocutionary effect (provoking laughter) is not a phenomenon exclusive to cringe comedy. At an open mic night, comedians will try out jokes for the first (or second, or third) time. Sometimes the audience finds the jokes funny and laughs. That's great. Illocutionary force: aiming to provoke laughter. Perlocutionary result: laughter. They match. The jokes are a success. This is the same in comedy films and sitcoms. Illocutionary force: produce laughter through comedic dialogue. Perlocutionary result: laughter. Again, a comedic

success. On the occasions where audiences *don't* laugh, the illocutionary force, or subtextual invitation inviting laughter, has failed to achieve the desired perlocutionary result. Instead, it has produced frowns, heckling, silence, or some other unintended response. It is worth noting again here that while a comedic action can't have an illocutionary force—which is a quality limited to utterances—comedic action has subtext, which can similarly fail in its invitation to laughter.

If cringe comedy jokes or actions are too embarrassing, they produce cringe overhang. Second-hand embarrassment can be so strong a violation of norms for certain audience members that it ceases to be laughable. One scene of *Peep Show* some people find difficult to watch is in season three, episode four, where Mark mistakes his infatuation for Big Suze—his roommate Jeremy's ex-girlfriend—for love (Armstrong and Bain December 2, 2005). When Jeremy discovers Mark and Big Suze hugging in Mark's room, Jeremy accuses Mark of being in love with Big Suze, then storms off. When she asks Mark if he really does love her, in an effort to avoid embarrassment, he goes way overboard in his protestations, saying, among other things, "*God no... Honestly, Suze, I like you, sort of, but not even really that much. You're very, y'know, you're the horsey type... Big stupid posh-head, that's you*", which scares her off. Here, the embarrassment for some viewers is so strong that it chokes off the possible interpretation of the embarrassment as simultaneously benign (and therefore funny). In these cases, cringe comedy doesn't produce the desired perlocutionary effect of laughter—the joke is unable to cut through the embarrassment, consequently leaving the viewer in a state of discomfort. This causes the actual perlocutionary response (looking away, leaving the room, blushing) to fail to match the *desired* perlocutionary result of laughter. In short, the joke seems to some audience members to act *only* as a violation.

In this section, I explained that cringe comedy can fall short of its illocutionary aim of provoking laughter by being too embarrassing. Strong violations can appear insufficiently benign—rendering them too embarrassing, causing cringe overhang. In other subgenres, this shortcoming can normally be fixed by working to increase the benign of the joke in question. However, cringe comedy, due to its genre constraints, is not always able to do this. I explain why this is the case in the following section.

5. Cringe comedy's reliance on its *lack* of social psychological distancing

You can break down cringe comedy like this:

1. Cringe comedy jokes carry an illocutionary force that aims to elicit laughter through feelings of second-hand embarrassment. Similarly, cringe comedy actions carry a subtextual invitation to laugh at second-hand embarrassment.
2. Cringe comedy can fail to achieve its illocutionary aim of laughter through this second-hand embarrassment. Instead of producing *both* laughter and second-hand embarrassment, the perlocutionary effect is solely the latter.
3. The benign violation theory suggests the excess embarrassment and dearth of laughter is caused by maximising the violation and failing to balance it with sufficient benign.
4. The violation of cringe comedy is deliberately caused by its *lack* of social psychological distancing. The audience members are made as embarrassed for the character as possible, maximising the violation. Therefore, cringe comedy is less benign and more polarising than other subgenres. (Sparrow 2024, 35-36)

This explains how cringe comedy is capable of provoking laughter in some viewers, and cringe overhang in others. When the viewers feel *some* second-hand embarrassment on the character's behalf, they are still able to laugh through it. To these viewers, cringe comedy jokes clearly feature a violation—but they're still funny. This is because while the violation might feel more intense than, say, a childish pun, the violation is still able to be simultaneously interpreted as sufficiently benign. This is illustrated with **Figure 4**.

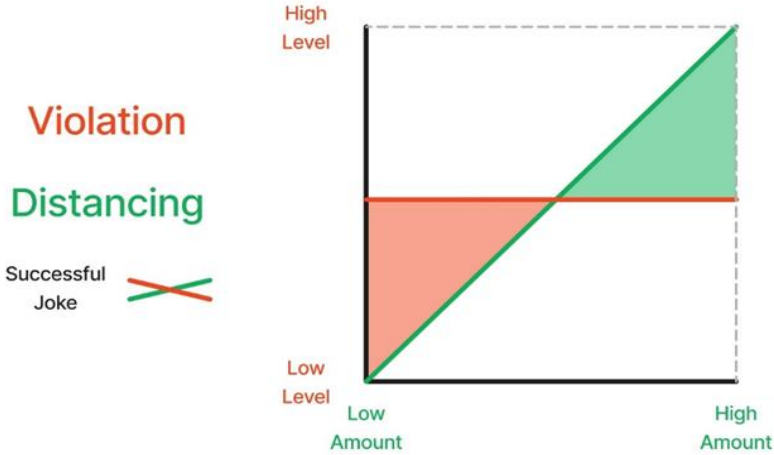


Figure 4

Typically, audiences experience cringe comedy as a relatively intense violation. The benign and the violation lines still intersect, so the joke remains a success. This situation is illustrated in **Figure 5**:

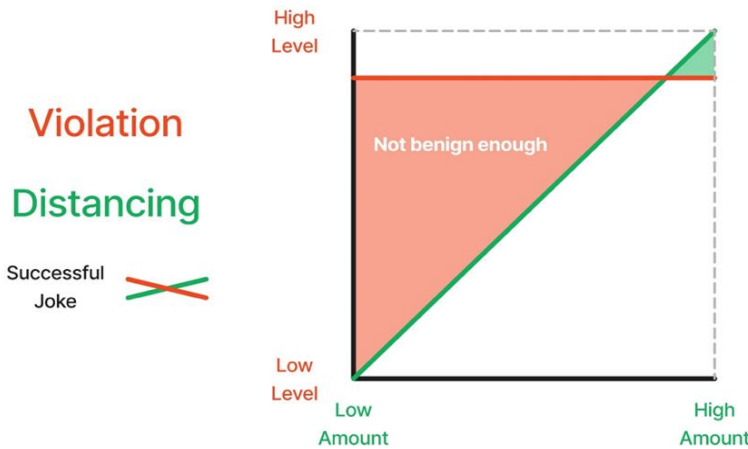


Figure 5

On some occasions, an audience member may experience the second-hand embarrassment so intensely that the violation is overwhelming. They might want to avert their eyes or leave the room. In these cases, an audience member might even fail to note where the joke is. The joke-attempt has overstepped the mark and is now only experienced as a violation: there is no intersection between the violation and the psychological distancing. They are experiencing cringe overhang. This is illustrated in **Figure 6**:

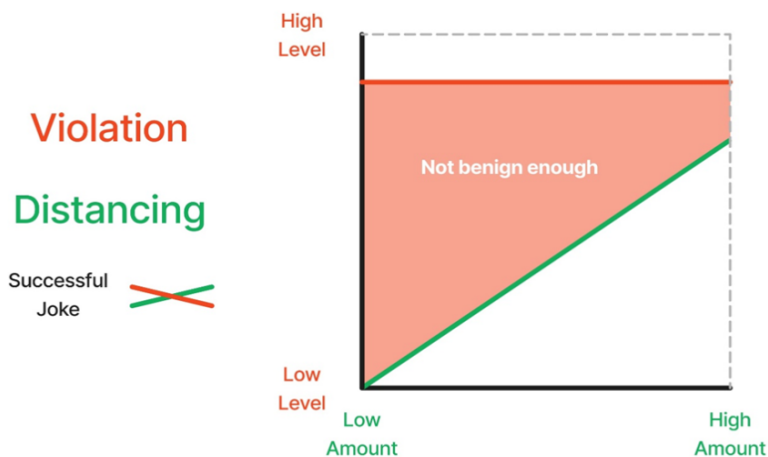


Figure 6

The responses above are not a reflection of the moral rightness or wrongness of cringe comedy—they just show how differently the same set of violation-heavy jokes can be perceived by different members of any given audience.

Cringe comedians cannot readily adapt their polarising material for unforgiving audiences, because without the second-hand embarrassment and resulting maximised violation, they don't have any jokes at all. If they added social psychological distancing to create the relief of additional distance from the characters' plights, they'd cease to operate within their chosen subgenre. Cringe comedy has a smaller audience than other subgenres because the violation essential to its existence is too overpowering for some audience members to forgive.

Again, perception of the funniness of jokes is subjective. Comedy, like other artforms, is subject to audience taste—and few people would say that all art must appeal to *everyone*. Some people may not be strongly affected by embarrassment (second-hand or otherwise), and so may not view cringe comedy and *Peep Show's* exploits as violations at all. Conversely, some people may find observational jokes on innocent-seeming subjects a strong violation because of their interest in that subject. Ryan Hamilton jokes about this phenomenon in his Netflix special, *Happy Face* (Raboy 2017, 41:00)—where a couple come up to him following his set about hot air balloons, claiming to be offended by the misrepresentation of their pastime. This seems to be a true story, but true or not, it shows how audience members can be made uncomfortable

by the comedic treatment of subjects they value or norms they hold: comedic treatment which may seem perfectly innocent to other people. It seems that while comedy is subjective, subgenres that intentionally create stronger violations can expect to produce stronger negative responses in audience members who find those violations especially confronting.

6. Conclusion

In this paper, I have drawn on speech act theory and benign violation theory to explain the phenomenon of cringe overhang. Speech act theory explains cringe comedy's illocutionary force as a subtextual invitation to laughter through second-hand embarrassment—which is why cringe comedy can be considered a failure when it only embarrasses the audience, but doesn't provoke laughter. Benign violation theory explains cringe comedy's method of inviting laughter as the production of non-socially psychologically distanced violations—in the form of second-hand embarrassment—accompanied by the smallest amount of benign to make it laughable. When a cringe comedy joke is experienced solely as a violation, the benign violation theory explains how cringe comedy fails to produce laughter: the joke is not benign-violation balanced. Having failed to achieve its desired perlocutionary effect of laughter, the joke's violation instead causes the lingering discomfort of ongoing second-hand embarrassment.

If you were to take a bite out of an apple, only to find half a worm in it after you had swallowed, you would likely feel a sense of disgust—even some time later, the very *thought* of eating said worm is likely to produce remnants of that same disgust. This half-worm hanging on in the back of the audience member's mind is cringe overhang.

Viewers who feel some second-hand embarrassment for *Peep Show*'s Mark Corrigan experience the violation as intended: sufficiently balanced with benign. Viewers who are stuck with a strong second-hand embarrassment struggle to laugh at his aborted proposal, chasing children with a metal pipe, and insulting his way out of an intimate conversation with his roommate's ex-girlfriend. For these viewers, the violation—instead of producing laughter—provokes cringe overhang. Due to the risk of experiencing this undesirable phenomenon, cringe comedy is a subgenre that, due to its genre constraints, just *isn't* for everyone.

Acknowledgments

For their many useful suggestions and valuable feedback, I would like to thank Richard Joyce and Edwin Mares; the 2023 conference convenors and attendees from both the Aesthetic Education through Narrative Art (Croatian Science Foundation) and International Society for Humor Studies organisations; Murray Smith, Oliver Double, Graeme A. Forbes and staff and students at the University of Kent; Katie Boyle; and especially the reviewers of this manuscript.

REFERENCES

- Armstrong, Jesse, and Sam Bain, creators. *Peep Show*. Aired 2003-2015, Channel 4. 4DVD.
- , writers. *Peep Show*. Season 1, episode 1, “Warring Factions.” Aired September 19, 2003, on Channel 4. Directed by Jeremy Wooding, featuring David Mitchell and Robert Webb. 4DVD, 2006, DVD.
- , writers. *Peep Show*. Season 3, episode 4, “Sistering.” Aired December 2, 2005, on Channel 4. Directed by Tristram Shapeero, featuring David Mitchell and Robert Webb. 4DVD, 2006, DVD.
- , writers. *Peep Show*. Season 3, episode 6, “Quantocking.” Aired December 16, 2005, on Channel 4. Directed by Tristram Shapeero, featuring David Mitchell and Robert Webb. 4DVD, 2006, DVD.
- Austin, John Langshaw. 1975. *How to do Things with Words*. The William James Lectures Series. Oxford: Clarendon Press.
- Hye-Knudsen, Marc. 2018. “Painfully Funny: Cringe Comedy, Benign Masochism, and Not-So-Benign Violations.” *Leviathan: Interdisciplinary Journal in English* 2: 13-31.
<https://doi.org/10.7146/lev.v0i2.104693>
- Kant, Leo and Elisabeth Norman. 2019. “You Must Be Joking! Benign Violation, Power Asymmetry, and Humour in a Broader Social Context.” *Frontiers in Psychology* 10. Article 1380.
<https://doi.org/10.3389/fpsyg.2019.01380>
- Mayer, Annalina Valpuri, Frieder Michel Paulus, and Sören Krach. 2021. “A Psychological Perspective on Vicarious Embarrassment and Shame in the Context of Cringe Humor.” *Humanities* 10 (4). Article 110. <https://doi.org/10.3390/h10040110>
- McGraw, Peter, and Caleb Warren. 2010. “Benign Violations: Making Immoral Behaviour Funny.” *Psychological Science* 21 (8): 1141-1149.

- McGraw, Peter, Lawrence E. Williams, and Caleb Warren. 2014. "The Rise and Fall of Humour: Psychological Distance Modulates Humorous Responses to Tragedy." *Social, Psychological, and Personality Science* 5 (5): 566-572.
- Raboy, Marcus, director. *Ryan Hamilton: Happy Face*. Comedy Dynamics. Aired August 29, 2017 on *Netflix*, <https://www.netflix.com/hr-en/title/80184834>
- Sparrow, Alexander. 2024. "Humour and the Appearance of Authenticity in Live Comedy." Master's Thesis, Victoria University of Wellington. Open Access Te Herenga Waka-Victoria University of Wellington. <https://doi.org/10.26686/wgtn.25304839>
- Veatch, Thomas C. 1998. "A Theory of Humor." *Humor: International Journal of Humor Research*, 11 (2): 161-215 <https://doi.org/10.1515/humr.1998.11.2.161> 199



WHAT COUNTS AS (EVIDENCE OF) NARROW AESTHETIC COGNITIVISM

Mario Sluga¹ 

¹Queen Mary University of London, United Kingdom

Original scientific article – Received: 09/02/2024 Accepted: 06/07/2024

ABSTRACT

In this paper I argue that the existing arguments for narrow aesthetic cognitivism are not valid. The reason is that the proponents of the view have mostly focused on theoretical debates rather than on empirical studies of the matter. I make my argument in five steps. First, I distinguish narrow from broad aesthetic cognitivism. Second, I specify the scope of the narrow claim and suggest that narrow aesthetic cognitivism is not about art in general but about narrative fictional works. Third, I point out that arguments in favor of the possibility of such works expressing truth do not discuss how truth is justified. Fourth, I concede that justification of truths in fictional and nonfictional works in the end rests on the experiences outside of the work but argue that it is an empirical matter whether belief has been acquired or not. Fifth, I demonstrate that the most recent analysis on belief-acquisition evinces that fiction is not an effective vehicle for imparting beliefs. I conclude with a discussion of what would count as evidence in favor of aesthetic cognitivism.

Keywords: aesthetic cognitivism; empirical studies; knowledge; justification; fiction.

1. Introduction

The debate about the potential value of art goes at least to Plato's banning of artists from his ideal *Republic* for misrepresenting the world and thereby corrupting its citizens (Plato 1997). In doing so, Plato also set the terms of the debate which has since mostly revolved around whether art can be a source of epistemological and/or ethical knowledge (Aristotle 1987; Horace 1990; Boileau 1674; Schiller 1794; Spivak 2012). Put differently, the value of art has traditionally been construed in terms of the potential, on the one hand, for conveying truths about the world (Walsh 1969; Lamarque and Olsen 1994) and, on the other, in promoting ethical development (Shelley 1891; Levinson 1998). While in its broader sense, aesthetic cognitivism may be said to cover both these claims, in this paper I want to investigate only the former—I will refer to this as aesthetic cognitivism narrowly understood.

In what remains, I will argue that, while narrow aesthetic cognitivism is certainly a possibility, as it stands philosophers subscribing to it have not produced convincing arguments in favor of it. That is because they have primarily engaged in theoretical debates rather than in discussions of existing empirical literature on the subject. To make my case, I will first distinguish between narrow and broad aesthetic cognitivism. I will then propose that the scope of the narrow claim needs to be quantified with more care than has been hitherto. Having done so, I will survey different types of arguments in favor of the narrow claim and demonstrate that they fail to accomplish what they set out to do. I will conclude with a discussion of what would count as evidence in favor of aesthetic cognitivism.

2. Weak, broad, and narrow aesthetic cognitivism

Here is a sample of explicit articulations of aesthetic cognitivism.

Aesthetic cognitivism may be said to be the view that art at its best is a form of understanding and as such, though it differs greatly in other respects, is to be accorded the same evaluative status as science, a status which its undoubted capacity to entertain and give us pleasure could not justify. (Graham 1996, 1)

Aesthetic cognitivism (...) is best thought of as conjunction of two claims: first, that art can give us (non-trivial) knowledge,

and second, that the capacity of art to give us (non-trivial) knowledge (partly) determines its value *qua* art, i.e. its aesthetic value. (Gaut 2005, 437)

Aesthetic cognitivism and aesthetic anti-cognitivism can be defined as opposing responses to two questions: (1) Can art provide knowledge? And, if it can, (2) how is this aesthetically relevant? (Thomson-Jones 2005, 376)

[T]he claim that narrative art is a source of truth, capable of transferring knowledge and other cognitively relevant states (aesthetic cognitivism). (Vidmar Jovanović 2019a, 18)

One possibility is that art can promote new knowledge and understanding, a notion referred to as *aesthetic cognitivism*. (Christensen, Cardillo, and Chatterjee 2023)

As we can see the claims are both quite broad and quite weak. In the former case, Gaut and Thomson-Jones point out that the key component of aesthetic cognitivism is not only the epistemic import of art but that this epistemic payoff is also a key part of the aesthetic value of art. Concerning the latter, their proponents explicitly state that only some art conveys knowledge or that art can on occasions be of epistemic value by conveying knowledge.

Here, I will be focusing exclusively on the question of epistemic value without exploring whether or how it relates to the problem of aesthetic value. But I want to be even narrower than that. I also want to delimit “knowledge” that I will be addressing from those more broadly understood “form[s] of understanding” and “other cognitively relevant states”. Here I am discounting potential wider cognitive benefits such as the positive impact of art on imagination, emotions, curiosity, reasoning, etc. I will also not be addressing the debate on whether art conveys moral knowledge (for discussion, see Nussbaum 1990; Gaut 2005; Plantinga 2018). For one, a recent research project led by Vidmar Jovanović explicitly treats moral knowledge under a banner distinct from aesthetic cognitivism which they refer to as aesthetic education: “the exploration of the transformative power of aesthetic experience with respect to one’s ethical development” (2020).¹ The other reason is that skepticism of moral knowledge is far more prevalent than skepticism about traditional

¹ https://aetna.uniri.hr/?page_id=30

knowledge making the debate potentially moot (Campbell 2019). Instead, I am zeroing in on less controversial forms of knowledge usually captured under the banners of *know that*, *know how*, and *know what it is like* that I will outline in the next section.

The second point I want to address pertains to the “weakness” of the claim. In the forms expressed above, it is hard to deny that art *can* and does convey knowledge. Religious art regularly conveys religious truths. Michelangelo’s *Sistine Chapel Ceiling* (1508-12) imparts, among other things, the knowledge that in the Christian tradition God created Adam. Calvary paintings standardly inform the viewer that in the same tradition Jesus was crucified and died on the cross. Portraiture usually provides information about how the subject looked like. Jacques-Louis David’s painting *Portrait of Pope Pius VII* (1805) tells us a lot about the facial features of the eponymous sitter at the time the painting was made. Eddie Adams’ photograph *Execution of Nguyễn Văn Lém* (1968) informs us not only about the appearance of both the victim and his executioner but also a number of things about the act of killing (type of gun, place, its summary nature, etc.) as well as of the emotions of the two people at the fateful moment. *Grizzly Bear* (2005), a documentary by Werner Herzog, imparts next to, again, a range of visual information (about how its protagonist Timothy Treadwell looked like or the places he visited), also the information about Treadwell’s life, such as his enthusiasm for grizzlies and eventually death at the hands of one. Truman Capote’s 1966 non-fiction novel *In Cold Blood* teaches the readers about the particularly gruesome murders that took place in Holcomb, Kansas in 1959. Under this analysis, the weak version of aesthetic cognitivism is trivial, and the underlying claim must be stronger.

That this is the case and that the quoted general formulations are somewhat disingenuous is evinced by the fact that the discussions that follow regularly turn to fictional narrative arts such as literature and film. Graham (1996) immediately turns to literary fiction as does Gaut (2005) and Vidmar Jovanović (2019a) is explicit that she is concerned with narrative art and in practice tackles fictional narratives.² In other words, the non-trivial discussions about whether art imparts traditional forms of knowledge are really about whether fictional works of art or art fiction does so. In the first instance, therefore, I take (non-trivial) narrow

² Thomson-Jones (2005) is an exception, but her real focus is the second claim about the aesthetic value and how it relates to the broader version of epistemic part of aesthetic cognitivism. Christensen, Cardillo, and Chatterjee (2023) are similarly interested in this broader version of the epistemic value of aesthetic cognitivism.

aesthetic cognitivism to mean something like art fiction conveys knowledge.³

3. Narrow aesthetic cognitivism

In the previous section the problem of quantification already arose, but there I discussed it in terms of the claims' "weakness". While modals and quantifiers are not the same thing, the way in which proponents of narrow aesthetic cognitivism articulated in terms of "can" proceeded to evince their position was to find examples of at least *some* art fictions which convey knowledge. This is why I also framed my discussion as a list of examples vindicating the weak thesis. And while I have avoided modals in my formulation of narrow aesthetic cognitivism the problem remains. Would one fictional work of art suffice to evince narrow aesthetic cognitivism? Would one person learning something be sufficient? Who is this person who acquires knowledge? How much knowledge and, or, what is the relevant type of knowledge that should be acquired here? Let us address these questions in turn.

3.1 Art fiction

Proponents of narrow aesthetic cognitivism do not usually explicitly address what subset of art fiction is relevant. One option is that the discussion here is really about high art fiction. The examples that reappear in the discussions such as Tolstoy and Dickens are certainly consistent with this version. However, when narrow aesthetic cognitivists address other media more middle brow example start appearing including *Mission Impossible* and *Alien* series as well as *Game of Thrones* (HBO 2011-2019). This suggests that the implicit relevant subset is at least something like high and middle brow art fiction. Importantly, we are talking about art fiction stories. When in the rest of the paper I write something like art fiction imparts knowledge I take this to mean a significant subset of art fiction stories of at least high and middle brow nature imparts knowledge.

³ In a Waltonian (1990) framework all paintings and films are fictional. I defend a more moderate modification of the framework that corresponds with the ordinary distinction between fiction and nonfiction (Sluga 2019a; 2021b; 2023; forthcoming 2025).

3.2 Benefactors

The statement that art fiction confers knowledge does not specify who the beneficiaries are. I suggest that we should be interested here more in the audiences than in the makers of art fiction because it is uncontroversial that making an art fiction involves a more sustained engagement than reading, watching, or listening to it. For instance, writers regularly learn about the subject of their writing. It is hard to deny that in writing three books about Thomas Cromwell, Hillary Mantel acquired substantial knowledge about her protagonist and 16th-century England. The same holds for directors and screenwriters. In making *Oppenheimer* (2023), Christopher Nolan must have learned a lot about the eponymous physicist's life. In bringing characters to life through performance, actors learn about the people they portray. Through filming *The Iron Lady* (Phyllida Lloyd 2011), Meryl Streep acquired, among other things, knowledge on both how Margaret Thatcher's movements look like and how her voice sounds like as well as how to imitate them.

It might be objected that it is not really art fiction making that confers knowledge here but the *preparation* for the production of an artwork i.e., the research that precedes it is the relevant process. On this account Mantel learns about Cromwell and his times not by writing proper but in reading the books about the man and the period which she only later applies to writing. The same holds for Nolan and other biopic makers. And for Streep, similarly, it is watching and listening to the tapes of Thatcher that secures the actor's knowledge about her subject, not the performance itself.

The objection holds only under an improperly limited understanding of what it means to make art fiction. This cannot exclusively refer to the acts of putting certain words on paper, instructing the film crew what to do, or making certain bodily movements in front of the camera. Making art fiction also entails researching one's subject, thinking about how to execute one's craft, drafting, and redrafting what has been done with reference to previous research and attempts, etc. This is no different from writing an academic paper—in writing this I have learned a lot about aesthetic cognitivism. But on the limited view of academic writing which entails only the typing of the words and leaves out the preceding research I would have acquired no knowledge. Moreover, even this limited view which considers writing only to be the act of putting words on paper, must admit that this act of writing is not merely the movement of hands but also involves a lot of mental processing such as organizing of thoughts and articulation of those thoughts in words. Similarly, the very act of executing one's craft, be it novelists' and scriptwriters' writing,

instructing the film crew, or moving for the camera, also brings clarity to the ideas that one has about the subject of that craft. So even in putting the words on paper and formulating her sentences Mantel must have formed new connections and gained a clearer picture of Cromwell and early modern England. In instructing Cillian Murphy on how to perform, Nolan certainly arrived at a firmer grasp of Oppenheimer and his motives. In performing for the camera through multiple takes Streep developed her knowledge of how to sound and move more like Thatcher.⁴

Therefore, it is hard to deny that art fiction regularly confers knowledge to its makers. The more controversial question—and certainly the one that has garnered more attention but remains underspecified in the definitions of aesthetic cognitivism—is whether art fiction conveys knowledge *to its audiences*.

Once this is determined, we also need to ask what the size of the effect is. In other words, what subsets of audiences should gain knowledge for narrow aesthetic cognitivism to hold? There appears to be agreement that the audiences who will profit are those who are sufficiently attentive. Gregory Currie, for instance, talks about “suitably prepared readers” (2020, 158). Iris Vidmar Jovanović has in mind “an active, reflective reader” (2019b, 366). Stacie Friend, similarly, argues that all audiences are expert audiences for at least some genre (2014, 244). At the same time, narrow aesthetic cognitivists would also probably allow that art fiction imparts knowledge of the aesthetic world more easily than knowledge of the non-aesthetic one. Put differently, while these attentive audiences will regularly gain knowledge about the story-world this will be less frequent for knowledge about the actual world. Put in yet another way, under this view learning from art fiction is something like learning through classroom lectures—some students will be able to say what the lecture was about and others not, and a subset of the first will also have learned about some claims made in the lecture while another subset of the same will not. When I write that art fiction imparts knowledge from hereon, then, this is shorthand for saying a subset of art fiction stories of at least high and middle brow nature imparts knowledge to a significant subset of audiences.

3.3 Knowledge

It is also uncontroversial to say that all art fiction conveys knowledge, if we do not specify what kind of knowledge we are interested in. Put

⁴ Makers of art fiction arguably also become better at making that art through continuing practice.

otherwise, if we take knowledge simply to mean some form of justified true belief, then art fictions regularly convey knowledge about what is true in the world of fiction. Upon reading *Anna Karenina* (1878) audiences learn, among many other things, that Tolstoy's heroine killed herself by jumping in front of a train. Furthermore, audiences regularly acquire knowledge of genre conventions through engaging its members. After a certain number of contemporary zombie horror films, audiences recognize that in these films humans are oftentimes a greater danger to protagonists than zombies. Having seen a sufficient string of whodunnits, audiences will know that the convention of the genre is an elaborate revelation of how the crime was committed and by whom in the last act by the protagonist. By engaging with an artform, the audiences will also gain knowledge of the stylistic features and narrative structure of that artform such as that the events depicted through parallel editing usually take place at the same time or that there is a certain class of narrators who are not to be trusted. While the knowledge about the fictional story might be trivial, the ensuing examples become less and less so, essentially moving into some basic, but crucial, territory of formal analysis and narratology. Yet, I suggest, because the acquisition of such knowledge is not controversial and/or because all these types of knowledge are about a specific work of art fiction or art fiction in general, discussion of narrow aesthetic cognitivism have mostly skirted the subject.⁵ Put differently, narrow aesthetic cognitivism is dominantly interested in whether art fiction conveys non-aesthetic knowledge to its audiences, i.e. the knowledge of the world (minus the world of art).

But I should be more precise and list at least some of the relevant types of knowledge that narrow aesthetic cognitivism is interested in within that non-aesthetic brand. These are: perceptual, propositional, experiential, social, and how to knowledge.

Perceptual knowledge relates to how something appears to the senses: how it looks, sounds, smells, feels or tastes. At present, art fiction standardly conveys only visual and sound information. As such, we may, for instance, be interested in whether *Jurassic Park* (Steven Spielberg 1993) conveys knowledge about how Tyrannosaurus Rex looked and sounded, or, to use the above example, how Margaret Thatcher appeared, moved, and spoke. But there are some relevant examples of knowledge import which are difficult to deny. For instance, *Vertigo* (Alfred Hitchcock 1958) confers, among other things, knowledge of how James Stewart and Kim Novak looked at the time of the shooting, as well as

⁵ For one notable exception discussing narratological knowledge see Carroll (2013).

how many of the places in San Francisco appeared at the same time. More generally, every photographic fiction film conveys a plethora of historical information on the visual aspects of various entities including the actors, locations, and whatever was in front of the camera at that moment.⁶

Does this not mean that narrow aesthetic cognitivism is vindicated for there is a significant subset of art fiction—namely photographic film fiction—which provides nontrivial non-aesthetic knowledge to its audiences? I suggest this depends on what aspect of art fiction narrow aesthetic cognitivism takes to be important for conferring knowledge. The absence of discussions about the above point suggests that narrow aesthetic cognitivism is not really interested in this type of knowledge. I assume this is because the reason behind the imparting of knowledge in this case are medium properties—the photographic nature of traditional fiction film—rather than art fiction *qua* art fiction. The perceptual knowledge is conferred because conventionally photographic fiction film hinges on verisimilar photography. Except for, perhaps, radio drama there is no equivalent of imparting perceptual knowledge among other art fictions.⁷ Were the acquisition of knowledge due to aspects different than basic medium properties, I assume those debating narrow aesthetic cognitivism would have taken more interest. Instead, the implicit demands seem to be that the imparting of the knowledge should come through art fiction *qua* *fictional representation*.

What certainly commands considerable if not the most interest in these debates is propositional knowledge of the form knowing that *p* (Stolnitz 1992; Gaut 2005; Gibson, Huemer, and Pocci 2007; Vidmar Jovanović 2019b; McGregor 2021a; Sluga 2021a). Here the discussion is most often about statements about the world as it is. The question here is, do audiences, by reading Mantel’s trilogy about Cromwell and watching Nolan’s film about Oppenheimer acquire knowledge about these and related historical persons. When reading the sentence from the second instalment *Bring Up the Bodies* (2012)—“Cromwell has the skin of a lily’, the king pronounces”—do audiences learn something propositional about Cromwell or Henry VIII such as the shade of Cromwell’s skin

⁶ I am deliberately not including the discussion of how objects sound because these regularly come from pre-recorded studio sound libraries which were often made with very different objects. For voice, see the next footnote.

⁷ The reason I hesitate is that in radio drama actors often deliberately try not to sound like they usually do. The same often holds for fiction film where actors regularly sport accent different from their own. Certainly, even then fiction films and radio dramas confer perceptual knowledge about how the actors sounded at the time of the performance, but that type of knowledge seems to be less relevant because it is not really knowledge about how the actor sounds in general.

and/or his king's opinions about it? When watching the opening of *Oppenheimer* and seeing the titular character read the statement at a hearing do audiences learn something specific about the theoretical physicist such as the exact words that he said at the time in question or how he felt when saying them? Propositions can, of course, be broader and relate to general social and human condition such as the opening sentences of *Anna Karenina*—"All happy families are alike; each unhappy family is unhappy in its own way"—raising the question of whether novels impart such lessons. And propositions need not be explicit but may be implicitly conveyed through art fiction such as "drug crime is a structural problem" that is generally agreed to be articulated in *The Wire* (HBO 2002-2005). In the next section, when we discuss strategies for answering these questions, we will see that the crucial thing to remember here is that audiences need to learn from art fiction unlike makers who learn from the process of making art fiction based on research.⁸

An important subclass of propositional knowledge is the knowledge of possibilities, hypotheticals, and counterfactuals. In this case, it has been debated whether art fiction confers knowledge of what might happen and what could have happened (Putnam 1978; Gaut 2005; McGregor 2021a). Does, for instance, Dora Lessing's *The Golden Notebook* (1966) account of her heroine impart knowledge of a possible person in a specific historical period? Does Marlon James' *A Brief History of Seven Killings* (2014) confer CIA as a possible cause of crime and social harm in 1970s Jamaica?

Another type of knowledge relevant for narrow aesthetic cognitivism is experiential or phenomenal knowledge. This is the knowledge of what it is to have experiences more complex than perceptual ones such as what is to be in love, what is to be in mourning, as well as, less dramatically, how is to ride a horse or go deep diving (Walsh 1969; Putnam 1978; Gaut 2005; McGregor 2021a; Sluga 2021a). This type of knowledge also covers how a particularly valued knowledge among aesthetic cognitivists of how it is to be in somebody's else's shoes (Smith 2017). One question,

⁸ There seems to be an equivalent to medium-based perceptual knowledge in the propositional domain. In literary fiction, it is hard to deny that the texts qua texts impart a range of non-aesthetic information about syntax and semantics. Ernest Hemingway's fictional works, for instance, teach us a lot about how a grammatical sentence in American English of the first half of the 20th century looks like. The main difference is that this type of knowledge is not as easily extractable for the audiences as perceptual knowledge in film fiction is but would rather need more expert engagement by linguists or students thereof. (In another version of this, foreign language speakers can acquire knowledge about a language through reading or watching films in that language but again the learning is due to texts qua texts.)

for instance, might be does *Memento* (Nolan 2000) teach audiences how it is to have anterograde amnesia. Another: does *Requiem for a Dream* (Darren Aronofsky 2000) confer how the high of using controlled substances feels like?

The last type of knowledge falling under narrow aesthetic cognitivism I will address is how to or practical knowledge. This knowledge type involves non-propositional knowledge of how to undertake certain activities such as how to type with all ten fingers, how to play the guitar, how to tie a tie, how to score a penalty or how to dance. While it is not impossible that this type of knowledge could be articulated in propositional terms, it is often the case that the people who possess it are unable to do so. From the perspective of this knowledge, perhaps one can learn how to fight by watching kung fu films or how to dance and sing by watching musicals.

There is, admittedly, at least one other form of knowledge which bridges the propositional and practical type and generally refers to the domain of social relations. These include knowledge about human motivation and folk psychology more generally, knowledge of societal norms, how to behave in certain situations, and consequences of not doing so, knowledge about what it means to be a good parent or spouse, etc. It has been argued among those interested in evolutionary aspects of narrative and fiction that these are the reasons why the capacity to tell narratives and convey fictions evolved in the first place (Hutto 2008; Gottschall 2012; Boyd 2018).

While there are undeniably other forms of knowledge such as embodied knowledge, I will focus on the acquisition of propositional knowledge as this type is most often debated in literature and arguably the easiest to demonstrate.⁹ To do so, I will discuss the general structure of the argument in favor of narrow aesthetic cognitivism and argue that what they present as evidence of the claim is invalid. Many only argue that art fiction expresses truth. This fails because for knowledge to be obtained it also needs to be justified. Some also address the issue of justification. But even they regularly assume that justified truth will be automatically believed in by the audiences. There is an even smaller group who addresses potential mechanisms of belief acquisition. However, the empirical evidence they cite is selective and leaves out the most recent

⁹ I thank a reviewer for drawing my attention to this. They are also to be acknowledged for pointing out that propositional knowledge may be embedded in non-propositional form (rhythms, patterns, etc.). The reason why I focus on propositional knowledge deriving from implicit or explicit propositional representations remains the same as above.

meta-analysis of the problem which contradicts their argument. In other words, to argue that art fiction confers knowledge on its audiences, one needs to argue that it expresses truth, justifies the truth it conveys, and instils belief in this justified truth.

4. Art fiction expresses truth

In a recent contribution to aesthetic cognitivism Rafe McGregor (2021a) has argued that fiction imparts at least three types of knowledge about causes of crime and harm—phenomenological, mimetic, and counterfactual. In my typology, McGregor’s phenomenological knowledge (e.g., what is it like to live in a war zone through a graphic novel *The Sheriff of Babylon* [Tom King and Mitch Gerard 2018]) corresponds to experiential knowledge, mimetic or knowledge of everyday reality (e.g., the resources at the disposal of criminal cartels and their appearance via *Miami Vice* [Michael Mann 2005]) to a combination of perceptual and propositional knowledge, and counterfactual knowledge (e.g. CIA as a possible counterfactual cause of crime in Jamaica in the 1970s through *A Brief History of Seven Killings*) to a subtype of propositional knowledge.¹⁰ While it is certainly possible that all these fictional representations are true, McGregor is generally uninterested in demonstrating the way in which fiction justifies its truth. *Miami Vice*, undeniably, depicts in a rich perceptual way a range of material resources at the cartel’s disposal—from motorcades transporting new recruits through mobile jammers to hundred meters killing fields. But given that the audiences are watching a fiction film, what is to justify them in believing that what is depicted is the case in the real world? McGregor does not tackle the question. Yet, if knowledge is some form of justified true belief, proponents of narrow aesthetic cognitivism need to address the question of how fiction justifies its truths.¹¹

A critic of the idea that the discussion of propositional knowledge necessarily needs to involve a discussion of justification, might point out that conceptual knowledge also falls under the banner of propositional knowledge. The critic will point out, this type of knowledge requires no justification through empirical investigation of reality. Gaut (2005), for instance, argues that the knowledge of a definition of a term is conceptual knowledge. If we have been acquainted with the articulation of, say, the

¹⁰ McGregor also discusses moral knowledge in part under counterfactual knowledge or knowledge of how things should be.

¹¹ For a more detailed debate, see Sluga (2021a) and McGregor (2021b).

notion of “crystallisation” that Stendhal proposes and elaborates upon in his *On Love* (1822) we get to know what crystallisation is (Gaut 2005). While this is true in non-aesthetic settings, the case is different in art fiction.

First, it remains the case that the question of whether crystallisation—that love essentially rests on attribution of imaginary qualities to the object of love, qualities which are not actually possessed by it—is true or not still has not been settled. Only if it is true, do we gain knowledge of the non-aesthetic world through acquaintance with the term. And Stendhal is offering the concept in a fictional setting meaning that the main examples on which he articulates it are fictional. What we really have here is the knowledge of the fictional world in which crystallisation is the essential feature of love. And this is in fact not much different from the question of whether Tolstoy’s pronouncement about happy and unhappy families is true. In *Anna Karenina* it is, but it is an open question whether it is true in the actual world. The main difference is that Stendhal has come up with a word for his general claim about some aspects of human nature—“crystallisation”.

Gaut might retort that I have missed the point. The relevant knowledge acquired here is simply what the term means, not whether the proposition the term expresses is true or not. But for the knowledge not to be only about the fictional world, the same concept needs also to be about the non-aesthetic world. And if this is so, then there is no need to draw out such a specific example as Stendhal’s. Any instance in which an art fiction explains what a word means should under Gaut’s view count as a case of acquiring conceptual knowledge. *Pulp Fiction* (Quentin Tarantino 1994) opens with a dictionary style definition of “pulp”. *The Big Short* (Adam McKay 2015) has Margot Robbie explain what “subprime mortgages” are and what “shorting” means. However, that Gaut (2005, 441) continues his discussion of the defense of conceptual knowledge by saying that the defense may go too far suggests that the truth of the matter is relevant after all. He talks about how the defense fails when applied to know how and instances in which it turns out that a detailed fictional account of plumbing is completely inaccurate and causes a flood when “lessons” learned are applied in one’s real home. From this perspective, conceptual knowledge can also fail when applied to real world. It is true it does not fail in Stendhal, Tarantino, and McKay, but that is only accidental. The definitions provided could well have been invented. And given that they occur in fiction, it is a reasonable possibility that they are. This is not different from biopics or historical fiction in which it is never quite clear what piece of information provided is only fictional and what is also true.

Gaut might again object that I have read too much into his later discussion of the failure of defense of conceptual knowledge. Truth of the claim is beside the point. And he could say there is a difference from *Pulp Fiction* and *The Big Short* insofar they do not invent new concepts while *On Love* introduces a term to the non-aesthetic world, regardless of whether the proposition it means is true or not. But even in that case there is a need for justification. What needs to be justified is not whether crystallisation accurately describes key features of love in the non-aesthetic world but whether crystallisation has the same meaning in the fictional world and in the non-aesthetic world. To Gaut this is obvious because he knows the history of the term. But to a less astute reader this is far from evident.¹² And from this perspective precisely because it is found in art fiction and developed in an art fictional context, there is no guarantee that this novel concept also circulates in the non-aesthetic world and that it circulates with the same meaning. After all, Stanisław Lem also develops a term “solaristics”—a scientific discipline focusing on the planet Solaris—in his eponymous novel, but the concept has no non-aesthetic equivalent.

But even if Gaut is correct that there is no need for justification of conceptual knowledge the question whether audiences reading *On Love* acquire knowledge of crystallisation and whether audiences engaging fictions acquire any knowledge in general is empirical. And rarely has it been tackled by philosophers in a sustained manner as we will see in the section after next.

Before proceeding, it is necessary to say a word or two about knowledge of possibilities as a subset of propositional knowledge. Like conceptual knowledge, it could be said that a knowledge of possibility only requires acquaintance with the possibility given that the space of possibilities is much wider than what is actually the case. And art fiction excels in conveying detailed stories which are possible, be it very probable historical accounts, plausible gangster and romantic narratives, or improbable but not impossible horror, fantasy, and science fiction tales. On this account, by reading *A History of Seven Killings* audiences become acquainted with a possible history in which CIA has supplied weapons for the assassination attempt on Bob Marley. Two points should be made here.

¹² Truth be told, when I first read Gaut I thought he was talking about how Stendhal is explaining the meaning of a physical phenomenon of “crystallisation” in his novel. I have never before heard of the term meant in Gaut’s and Stendhal’s sense.

First, there is much art fiction which conveys impossible worlds, be it logically or against the laws of physics. Science fiction often involves travel at or higher than the speed of light which is not allowed by laws of physics. Other art fiction involves stories which are logically inconsistent either by mistake or design. Acquaintance with such stories could bring no knowledge of possibilities because what is depicted is impossible. Moreover, if the idea is that acquaintance with fiction in general leads to knowledge of possibilities, then there will be many art fictions which form inaccurate beliefs about what is possible.

Second and more importantly, knowledge of possibilities is arguably considerably less weighty than that of what actually is the case and is in a sense trivial. Certainly, upon reading James' novel audiences might arrive at the knowledge that it is possible that the CIA was involved in Marley's assassination attempt. But so could have KGB, MI5, or any other intelligence service for that matter. If an art fiction about a game of dice specifies that the sequence was 3-4-3-2 this does not seem to impart any special knowledge because audiences generally know that any sequence is possible in such a game. It might be retorted that the value is in a level of detail with which this sequence has been represented, how much the players bet, how they felt when betting, what the win or loss meant for them. But that does not change the fact that what is reported is a space of possibilities of human action and sentiment and that this space of possibilities is in a good sense already known. To put it in Jerome Stolnitz's terms, these are trivial truths which "resemble garden variety truths" (1992, 192).

Lastly, proponents of narrow aesthetic cognitivism also usually focus on propositional cases which are plausibly true but there is another tradition, running from Plato, which argues that art regularly expresses falsities. In screen studies, for instance, there is a body of work usually dubbed screen theory which argues that films excel in misrepresentations especially when it comes to underrepresented groups and minorities (women, ethnic, and sexual minorities, etc.) (see Sluga 2019b). In other words, proponents of narrow aesthetic cognitivism who are uninterested in the way art fiction justifies its truth would end up with art fiction which, arguably, imparts inaccuracies more often than it does knowledge for, on average, fiction certainly involves more inventions and misinformation about the non-aesthetic world than truths.

5. Art fiction justifies its truth

Stolnitz (1992) has produced one of the strongest critiques of narrow aesthetic cognitivism by arguing that art fiction, unlike science and history, does not have established procedures for justifying what it conveys is true. Art fiction regularly mixes truth with other elements, be it inventions or inaccuracies, but for the audiences who consume it, it does not provide any tools for disentangling the two. In other words, while it is the case that art fiction can convey truth it cannot convey any knowledge because whatever belief audiences acquire through their engagement with art fiction, this belief is not justified. There have, of course, been responses to Stolnitz's proposal arguing that art fiction does justify its truth.

One line of defense has been to argue that there are subsections of art fiction which audiences can assume are truthful. Proposed candidates have included realist fiction (Gaut 2005; Vidmar Jovanović 2019b) and sub-genres such as works by Dickens (Friend 2014). Vidmar Jovanović argues, for instance, that writers committed to realism such as Austen will accurately depict social and economic relations of their time. For Friend, we can assume Dickens to be a truthful guide on social facts but less so on scientific ones (given his recourse to spontaneous self-combustion in one of his novels). In other words, membership of a specific (sub)genre is what justifies truths in these fictions. From Stolnitz's perspective, however, the problem remains because even the most realist (sub)genres involve admixtures of other elements. Despite what the opening of Dickens' *Our Mutual Friend* (1865) suggests, for instance, there was no such a thing as collector of dead bodies from the banks of Thames. Put differently, for Stolnitz no art-fiction (sub)genre justifies its truth because it is a typical feature of all art fiction (sub)genres to involve inventions and inaccuracies.

Gaut has provided perhaps the strongest response to Stolnitz by arguing that while there is an institutional difference between non-fiction and art fiction in terms of immediate justification, in the end the deeper justification of both non-fiction and art fiction lies in one's experience. Gaut acknowledges that, unlike for fiction, non-fiction which standardly makes a claim to imparting knowledge is vetted through some kind of expert process. In academia, for instance, this is through peer-review while in journalism it is through editorial fact-checking. However, Gaut points, there are still non-fiction works which have escaped vetting either by design (frauds like *Hitler Diaries*) or mistake (arguably, even the best history books involve some inaccuracies). And just by looking at such non-fictional work audiences cannot know whether the vetting has been

successful. To ascertain that audiences need to go outside the text, examine the conditions of its making, check its claims against other sources, i.e. appeal to experience. In that sense there is no substantial difference with fiction because in both cases the ultimate justification lies in the appeal to experience.

While justification does ultimately lie with experience, the institutional guarantee does play an important role that is not addressed in Gaut's retort. Justification pertains to ascertaining whether something is true or not. But knowledge also involves belief and just because a truth is justified it does not mean a belief about it will be formed. In other words, narrow aesthetic cognitivists also need to explain how audiences come to acquire beliefs and whether they do acquire them in the first place. Institutional guarantee that comes with non-fiction but is absent from art fiction is arguably part of a mechanism for acquiring beliefs. Put in yet another way, the underlying contract for consuming non-fiction is that one is supposed to believe in it. There is no similar underlying contract for art fiction not even in its most realist subsets.

6. Art fiction imparts beliefs

A minority of proponents of narrow aesthetic cognitivism have recognized that it is also necessary to investigate how and whether art fictions influence beliefs. They usually cite a selection of empirical studies and propose certain mechanisms of belief acquisition (see Friend 2014; Carl Plantinga 2018). The work that is usually put forward to support the idea that art fictions change beliefs comes from a few psychologists including frequent collaborators Deborah Gerrig and Richard Prentice (Gerrig 1991; Prentice et al. 1997; Prentice and Gerrig 1999), Melanie Green and Timothy Brock (2000), and Elizabeth Marsh (Marsh et al. 2003; Marsh and Fazio 2006). These are usually studies of literary art fiction where a general knowledge test follows the exposure to a fictional text set against a control group.

In one string of studies one group reads fictional stories with general statements about the non-aesthetic world which were either true (e.g., mental illness is not contagious) or false (e.g., mental illness is contagious). The control group reads either unrelated stories or no stories at all. Those exposed to fictional texts containing the above general statements (the test group) were statistically more likely to agree with the claims in the follow-up questionnaire so long as they were consistent with the information in the text, even if the information was false (Gerrig 1991; Prentice et al. 1997; Prentice and Gerrig 1999). When participants

read one and the same story, but one group was told the story is fictional and the other nonfictional those in the fictional group were more likely to agree with the statements they have read (Prentice and Gerrig 1999). In another set of experiments, the general statements about the non-aesthetic world which could be either true (e.g., extant is used for navigation via stars) or false (e.g., compass is used for navigation via stars) were only peripheral or in the background of the fictional story. Much like in the previous studies, those exposed to the general statements did significantly better (in case of true statements) or worse (in case of false statements) than the control group on a follow-up general knowledge test (Marsh et al. 2003; Marsh and Fazio 2006). In other words, a significant subset of audiences acquires propositional beliefs from fiction. From this perspective, fiction is even more persuasive than non-fiction in imparting beliefs about propositional claims. Interestingly, even those more skeptical of knowledge acquisition like Gregory Currie have argued that “[i]t is an obvious truth that fictions affect the beliefs of those who come into contact with them” (2020, 161).

This, however, is belied by the currently most comprehensive meta-analysis of persuasive effects of fiction (Braddock and Dillard 2016). In the last four decades meta-analysis has become a key method of summarizing research on a given subject in various disciplines including psychology and an important way to tackle replication crisis in those disciplines (Sharpe and Poets 2020). Meta-analysis as a method identifies different studies focusing on a similar question and statistically combines the results of these studies to compute whether there is an effect and the size of the effect (Borenstein et al. 2011). Given that results of single studies are statistical analyses themselves based on sample sizes and quantified responses which provide values for whether there is an effect and what its size is, with right parameters and weightings these analyses can be combined into an overarching statistical analysis to ascertain whether the effect holds more broadly or not and its size if it does. As such meta-analyses are not competing studies which look at the same question and find the effect (size) or not. They are analyses of the whole body of research on a topic of interest and their results hold more generally than any single study.

Braddock and Dillard have produced a meta-analysis of 74 empirical studies investigating effects of fiction across media (visual, audio, theatre, literary) on four different aspects of persuasion—attitudes, desires, behaviors, and beliefs. For each of these aspects of persuasion they have made a separate analysis to see whether there is a difference in how fiction and nonfiction influences a given aspect. For beliefs, they have found that “nonfictional narrative stimuli significantly affect belief

change, but fictional narrative stimuli do not” (2016, 457).¹³ In other words, while about dozen different studies that Friend et al. cite evince the sought-out effect—that art fictions change beliefs—the meta-analysis of 74 studies on the subject demonstrates that there is in fact no such effect—art fiction does not change beliefs. From a statistical perspective, the reason is that the effect sizes and sample sizes in studies reported by Friend are outweighed by the lack of effect and sample sizes in the studies she does not cite. In fact, the most recent non-statistical literature review of the subject arrives at the same conclusion (Dubourg and Baumard 2023).¹⁴

Moreover, Braddock and Dillard have also considered whether each of the four aspects of persuasion effects—attitudes, desires, behaviors, and beliefs—depends on the medium. They have found that “*existing data support the conclusion that medium of presentation is unimportant for inducing persuasion via narrative*” (Braddock and Dillard 2016, 462, italics in the original). Put differently, that fictional narratives do not change beliefs cannot be explained by the claim that one type of fictional narratives, say the textual ones put forward by Friend, change beliefs whereas the others do not. In general, in the fictional case no medium type changes beliefs (in the nonfictional case they all do).

Braddock and Dillard’s findings also problematize some of the proposed mechanisms of belief acquisition. Building on the work of Prentice and Gerrig (1999) and Gilbert (1991), Friend (2014) has advocated for scrutiny lowering—an idea that the very label “fictional” lowers a more critical outlook on what is consumed leading to easier persuasion than in non-fiction. This however is contradicted by the above meta-analysis which has the opposite results of what the scrutiny lowering theory predicts.¹⁵

Other mechanisms such as Currie’s (2020) narrative value have different issues. While Currie is skeptical of a general belief acquisition mechanism, he still proposes a typical way in which audiences infer beliefs about the non-aesthetic world from beliefs about fiction: “Deviations in fiction from truth are expected to be justified by their

¹³ Unlike for beliefs, they determine that attitudes, desires, and behaviors are influenced by fiction suggesting that broad aesthetic cognitivism and in particular aesthetic cognitivism focusing on morality stands on a firmer ground.

¹⁴ Interestingly, they do not mention Braddock and Dillard (2016) but provide proposals for why studies which report effects should instead be construed as showing something else.

¹⁵ Braddock and Dillard’s findings also contradict proposals such as the availability heuristic (Tversky and Kahneman 1973) and the acceptance by default (Gilbert 1991) as both would predict that, all other things being equal, effects of fictions and nonfictions are the same.

narrative payoff” (2020, 161). In other words, if information is simply in the background, then the tacit assumption is that it is true in real life. Because French interpolations by Russian aristocracy in Tolstoy’s *War and Peace* (1869) are not of particular narrative importance (they could have been in, say, Italian and nothing much would change narratively speaking) but happen in the background, audiences generally infer that 19th century Russian aristocracy spoke French. By contrast, because Dickens’ *Our Mutual Friend* opens in a particularly dramatic fashion with a collector of dead bodies from the Thames audiences can assume that the collector is not a historical profession. Counterexamples, however, abound. In the case of background, in historical fiction films characters regularly speak in contemporary languages rather than languages of the epoch depicted yet a diet of Hollywood epics on Rome will not lead to an inference that Romans spoke English. Similarly, irrespective of their occupation characters in American TV shows regularly look like models yet this hardly leads to inferences that a typical policeman, doctor, or a lawyer conforms to contemporary standards of beauty much more than an average person. When it comes to the foreground, Currie himself admits that when he first read *Our Mutual Friend*, he came to mistakenly believe that there was such a profession as collector of bodies from the Thames. This means that the difference between foreground and background is much less clear than he proposes leading to problems in determining what to believe in. There is also a matter of middle ground: what do we do with fictional truths which are neither crucial for the storyline but are at the same time not unimportant such as that queen Anne was a lesbian (*The Favourite* 2018, Yorgos Lanthimos) or that first inoculation took place on the Danish Court (*A Royal Affair* 2012, Nikolaj Arcel)?

In short then, existing studies disprove the claim that art fiction imparts beliefs making general proposals about belief acquisition moot. It is true that virtually all the studies in the meta-analysis are short term studies in the sense that they only look for effects immediately after the exposure to art fictions. But this arguably makes things even more difficult for narrow aesthetic cognitivists because it is unlikely there would be long term effects on beliefs without there being short term ones. In other words, the biggest hurdle for proponents of narrow aesthetic cognitivism is that empirical findings contradict the view that art fictions impart beliefs.¹⁶

¹⁶ It might be objected that I am taking narrow aesthetic cognitivists to a different standard than the one I have produced in this paper. I have referred to no empirical studies when claiming standard knowledge acquisition among art fiction makers and aesthetic knowledge acquisition among audiences. That is true, but in these cases the situation is different in the sense that empirical proof is already embedded in standard fiction art making and consumption practices. Concerning audiences, it

It is possible, however, that proponents of narrow aesthetic cognitivism are not persuaded by the empirical findings I cite. One objection could be that psychological studies of the type that Friend et al. refer to and that are included in the meta-analysis are simply not a good way of exploring the question that interests us—whether art fiction changes beliefs. The objection broadly claims that what the cited experiments measure is not belief but something else. Therefore, it is irrelevant for the debate either way. If we think about beliefs as cognitive states which are relatively stable over time, then short-term studies deploying questionnaires only immediately after the exposure to art fiction to gage beliefs stemming from that text are lacking. Such studies are not really measuring belief but some kind of acceptance or subconscious recall. What is required is a long-term study where the questionnaires are repeated after a prolonged period of several weeks or more.

In another variant of the objection what is actually studied is not beliefs derived from fiction, but beliefs formed on the basis of who presented them. When participants are provided with information in experimental conditions because the information is provided by experimenters whom participants intuitively trust to convey truths, it does not matter whether information is labelled fiction or nonfiction. What matters is the implicit trust in the presenter of information. The conclusion of both variants is the same—short-term studies do not capture (relevant) beliefs, only long-term ones do.

Tackling the second variant first, if participants generally got their beliefs from the fact that experimenters provided implicitly trustworthy information, then all the experiments in which the control group engages unrelated or no work and the test group is presented with art fiction would need to show effects because questionnaires would necessarily include references to statements in the art fiction provided by the experimenters. Yet clearly there are such studies which do not find an effect. Moreover, as mentioned earlier there are studies which present one and the same text to the control and test group but label it as nonfiction and fiction, respectively (Prentice and Gerrig 1999). According to the

is uncontroversial that at least a significant subset understands the stories they are consuming. If our everyday conversations with fellow moviegoers, TV show watchers, or novels readers is not proof enough, then psychological literature I have already cited should, as it assumes this basic comprehension in its questionnaires. Turning to art fiction makers, it is unclear how, unless using something like ChatGPT or having a ghost writer, a writer of historical fiction like Mantel could produce her work without extensive historical research. And if that is not sufficient, I can point to Mantel's own words who admits that it took her five years to research the book (Alter 2019). Similar observation about learning through making art fiction can be made for filmmakers like Nolan and actresses like Streep and their subjects (Ebiri 2023; NPR staff 2011).

second variant the result of this experiment should be no difference between the two groups (both groups are getting relevant information from the trustworthy experimenters), but the experiment reports a significant difference. Most importantly, there is also the matter of long-term studies which invalidates both variants of the objection.

While there is only a handful of such long-term studies, they do exist (see Strange and Leung 1999; Brodie et al. 2001; Marsh et al. 2003; Howell 2011; Schneider-Mayerson et al. 2020). Crucially, all but Marsh et al. (2003) report disappearance of the effect while even Marsh et al. report significant decrease in the effect size. In other words, even if we accept that short-term studies do not actually measure beliefs, long-term studies evince that art fiction does not establish imparting or change of beliefs construed as temporally stable cognitive states.¹⁷

7. Future directions for narrow aesthetic cognitivism

While the above invalidates current defenses of narrow aesthetic cognitivism that does not mean that narrow aesthetic cognitivism does not hold. One obvious response is that the empirical findings about belief relate only to propositional belief and knowledge and even that only in part. There are other (sub)types of belief and knowledge that could be acquired including experiential belief and knowledge. This is certainly one avenue of research that should be pursued both experimentally and philosophically.

Defenders of narrow aesthetic cognitivism should also not neglect long-term studies of a somewhat different type. It is possible that beliefs and knowledge accrues over time through a repeated exposure to art fiction (sub)categories. Audiences might not pick up that the Roman Republic was run by the Senate after watching *Cleopatra* (Joseph L. Mankiewicz 1963) once or if they do the effect might not be long-term, but they might acquire a robust belief and knowledge of this fact after a more extensive engagement with the genre of Hollywood epics about Rome. In other words, there is much space for long-term studies investigating repeated exposure to certain types of art fiction.

¹⁷A somewhat different meta-analysis which looks at long term persuasion effects (attitudes, intentions, and beliefs) difference between narratives and non-narratives reveals that non-fictional narratives are more persuasive than non-narratives, but that fictional narratives are only as persuasive as non-narratives (Oschatz and Marker 2020). When it comes to beliefs as opposed to intentions and desires, narratives are no more persuasive than non-narratives regardless of whether narratives are fictional or not.

Focusing more on belief and knowledge of social relations which bridges propositional and how to knowledge, it has been recognized that it is difficult to evidence this type of belief and knowledge acquisition in adults because they might have already hit the ceiling having been embedded in a storytelling society from their birth (Boyd 2018). That is why we need to turn to children and adolescents to investigate in more detail whether art fiction imparts belief and knowledge of social relations.

However, proponents of narrow aesthetic cognitivism also need studies that address knowledge acquisition in addition to belief acquisition, because it is necessary to ascertain that, if belief acquisition is an empirical reality, it has an overall positive epistemic effect. In other words, narrow aesthetic cognitivists need to accept the fact that demonstrating belief acquisition might also lead to concede that, while art fiction can certainly have positive epistemic outcomes, in reality it may bring more epistemic harm than good.

Acknowledgments

I wish to thank the three anonymous reviewers for EuJAP for providing useful and constructive comments on a previous version of this paper. This work has been supported in part by the Croatian Science Foundation (project AETNA, grant UIP-2020-02-1309).

REFERENCES

- Alter, Alexandra. 2009, November 13. "How to Write a Great Novel." *The Wall Street Journal*. Accessed February 2, 2024. https://www.wsj.com/articles/SB10001424052748703740004574513463106012106?mod=WSJ_hpp_sections_lifestyle.
- Aristotle. 1987. *Poetics*. Translated by Richard Janko. Indianapolis: Hackett.
- Boileau-Despréaux, Nicolas. 1674. *L'Art Poétique*. Paris: Denys Thierry.
- Boyd, Brian. 2018. "The Evolution of Stories: From Mimesis to Language, from Fact to Fiction." *Wiley Interdisciplinary Reviews: Cognitive Science* 9 (1): e1444. <https://doi.org/10.1002/wcs.1444>
- Borenstein, Michael, Larry V. Hedges, Julian P. T. Higgins, and Hannah R. Rothstein. 2009. *Introduction to Meta-Analysis*. Hoboken: John Wiley & Sons.


- Braddock, Kurt, and James Price Dillard. 2016. "Meta-analytic Evidence for the Persuasive Effect of Narratives on Beliefs, Attitudes, Intentions, and Behaviors." *Communication Monographs* 83 (4): 446-467.
- Campbell, Richmond. 2019. "Moral Epistemology." *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. <https://plato.stanford.edu/archives/win2019/entries/moral-epistemology>
- Carroll, Noël. 2013. "Memento and the Phenomenology of Comprehending Motion Picture Narration." In *Minerva's Night Out: Philosophy, Pop Culture, and Moving Images*, 203-220. Hoboken: Wiley-Blackwell.
- Christensen, Alexander P., Eileen R. Cardillo, and Anjan Chatterjee. 2023. "Can Art Promote Understanding? A Review of the Psychology and Neuroscience of Aesthetic Cognitivism." *Psychology of Aesthetics, Creativity, and the Arts*. Online publication. <https://doi.org/10.1037/aca0000541>
- Currie, Gregory. 2020. *Imagining and Knowing: The Shape of Fiction*. Oxford: Oxford University Press.
- Ebiri, Bilge, 2023, July 17. "An Action Movie About Scientists Talking." *Vulture*. Accessed February 2, 2024. <https://www.vulture.com/article/oppenheimer-christopher-nolan-cillian-murphy-behind-the-scenes.html>
- Friend, Stacie. 2014. "Believing in Stories." In *Aesthetics and the Sciences of Mind*, edited by Greg Currie, Matthew Kieran, Aaron Meskin, and Jon Robson, 227-248. Oxford: Oxford University Press.
- Gaut, Berys. 2005. "Art and Knowledge." In *The Oxford Handbook of Aesthetics*, edited by Jerrold Levinson, 436-450. Oxford: Oxford University Press.
- Gerrig, Richard J. and Deborah A. Prentice. 1991. "The Representation of Fictional Information." *Psychological Science* 2: 336-40.
- Gibson, John, Wolfgang Huemer, and Luca Poggi, eds. 2012. *A Sense of the World: Essays on Fiction, Narrative, and Knowledge*. London: Routledge.
- Gilbert, Daniel T. 1991. "How Mental Systems Believe." *American Psychologist* 46 (2): 107-19.
- Gottschall, Jonathan. 2008. *Literature, Science, and a New Humanities*. London: Springer.
- Graham, Gordon. 1996. "Aesthetic Cognitivism and the Literary Arts." *Journal of Aesthetic Education* 30 (1): 1-17. <https://doi.org/10.2307/3333229>

- Green, Melanie C., and Timothy C. Brock. 2000. "The Role of Transportation in the Persuasiveness of Public Narratives." *Journal of Personality and Social Psychology* 79: 701–721.
- Horace. 1990. *Epistles Book II and Ars Poetica*. Translated by Niall Rudd. Cambridge: Cambridge University Press.
- Hutto, Daniel D. 2012. *Folk Psychological Narratives: The Sociocultural Basis of Understanding Reasons*. Cambridge: MIT Press.
- Lamarque, Peter, and Stein Haugom Olsen. 1994. *Truth, Fiction, and Literature*. Oxford: Oxford University Press.
- Levinson, Jerrold, ed. 1998. *Aesthetics and Ethics*. Cambridge: Cambridge University Press.
- Marsh, Elizabeth, and Lisa Fazio. 2006. "Learning Errors from Fiction: Difficulties in Reducing Reliance on Fictional Stories." *Memory & Cognition* 34: 1140–1149.
- Marsh, Elizabeth J., Michelle L. Meade, and Henry L. Roediger III. 2003. "Learning Facts from Fiction." *Journal of Memory and Language* 49: 519–536.
- McGregor, Rafe. 2021a. *A Criminology of Narrative Fiction*. Bristol: Bristol University Press
- . 2021b. "Response to Frauley, Simecek, Slugan, and Whitecross." *Journal of Theoretical & Philosophical Criminology* 13: 148–157.
- NPR staff. 2011, December 19. "From Meryl to Margaret: Becoming the Iron Lady?". *NPR*. Accessed February 2, 2024. <https://www.npr.org/2011/12/19/143962896/from-meryl-to-margaret-becoming-the-iron-lady>
- Nussbaum, Martha C. 1990. *Love's Knowledge: Essays on Philosophy and Literature*. Oxford: Oxford University Press.
- Oschatz, Corinna, and Caroline Marker. 2020. "Long-term Persuasive Effects in Narrative Communication Research: A Meta-analysis." *Journal of Communication* 70 (4): 473–496.
- Plantinga, Carl. 2018. *Screen Stories: Emotion and the Ethics of Engagement*. Oxford: Oxford University Press.
- Plato. 1997. "The Republic". In *Plato, Complete Works*, edited by J. M. Cooper, 971–1223, Indianapolis: Hackett.
- Prentice, Deborah A., and Richard J. Gerrig. 1999. "Exploring the Boundary between Fiction and Reality." In *Dual-Process Theories in Social Psychology*, edited by Shelly Chaiken and Yaacov Trope, 529–46. New York: Guilford Press.
- Prentice, Deborah, Richard Gerrig, and Daniel Bailis. 1997. "What Readers Bring to the Processing of Fictional Texts." *Psychonomic Bulletin & Review* 4: 416–20.
- Putnam, Hilary. 1978. *Meaning and the Moral Sciences*. London: London: Routledge & Kegan Paul.

- Schiller, Friedrich. 1794. *Über die Ästhetische Erziehung des Menschen*. Stuttgart: Cotta'sche.
- Sharpe, Donald, and Sarena Poets. 2020. "Meta-analysis as a Response to the Replication Crisis." *Canadian Psychology/Psychologie Canadienne* 61 (4): 377-387.
- Shelley, Percy Bysshe. 1891. *A Defense of Poetry*. Boston: Ginn & Company.
- Sluga, Mario. 2019a. *Fiction and Imagination in Early Cinema: A Philosophical Approach to Film History*. London: Bloomsbury.
- . 2019b. *Noël Carroll and Film: A Philosophy of Art and Popular Culture*. London: Bloomsbury.
- . 2021a. "Fiction, Knowledge and Cinematic Realism." *Journal of Theoretical & Philosophical Criminology* 13: 99-110.
<http://www.jtpcrim.org/OCT2021/Mario.pdf>
- . 2021b. "Textualism, Extratextualism, and the Fiction/Nonfiction Distinction in Documentary Studies." *Studies in Documentary Film* 15 (2): 114-126
<https://doi.org/10.1080/17503280.2021.1923142>
- . 2023. "Fiction as Challenge for Text-Oriented Film Studies." *New Review of Film and Television* 21 (3): 427-450
<https://doi.org/10.1080/17400309.2022.2132072>
- . Forthcoming 2025. *Taking Fiction Film Seriously: A Philosophical Approach to Cinema Theory*. London: Bloomsbury.
- Spivak, Gayatri Chakravorty. 2012. *An Aesthetic Education in the Era of Globalization*. Cambridge: Harvard University Press.
- Stolnitz, Jerome. 1992. "On the Cognitive Triviality of Art." *The British Journal of Aesthetics* 32 (3): 191-200.
- Thomson-Jones, Katherine. 2005. "Inseparable Insight: Reconciling Cognitivism and Formalism in Aesthetics." *The Journal of Aesthetics and Art Criticism* 63 (4): 375-384.
<https://www.jstor.org/stable/3700513>
- Vidmar Jovanović, Iris. 2019a. "Cognitive and Ethical Values and Dimensions of Narrative Art." In *Narrative Art, Knowledge and Ethics*, edited by Iris Vidmar Jovanović, 17-85. Rijeka: University of Rijeka, Faculty of Humanities and Social Sciences.
- . 2019b. "Literature and Truth: Revisiting Stolnitz's Anti-cognitivism." *Croatian Journal of Philosophy* 19 (56): 351-370.
- Walsh, Dorothy. 1969. *Literature and Knowledge*. Middletown: Wesleyan University Press.
- Walton, Kendall L. 1990. *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Cambridge: Harvard University Press.

Tversky, Amos, and Daniel Kahneman. 1973. "Availability: A Heuristic for Judging Frequency and Probability." *Cognitive Psychology* 5 (2): 207-232.

REVISITING MCGEE'S COUNTEREXAMPLE TO MODUS PONENS

Nicolás Lo Guercio¹ and Mariela Rubin¹ 

¹UBA/IIF-SADAF-CONICET, Argentina

Original Scientific Article – Received: 09/04/2024 Accepted: 02/11/2024

ABSTRACT

In this paper, we provide a novel explanation of McGee's (1985) alleged counterexample to Modus Ponens for indicative conditionals. Our strategy is to show that pragmatic phenomena interfere with intuitions concerning the acceptability of the inference. More specifically, we argue that two confounding factors at play affect our intuitions concerning the acceptability of the inference, neither of which is related to validity. First, the indefinite determiner phrase "a Republican" is ambiguous, to wit, it may receive either a specific or a non-specific reading (and as we shall see, substituting a disjunction or a definite description for the indefinite is of no help). Second, the conclusion triggers an ignorance inference concerning its antecedent. In light of this, we shall argue, McGee's diagnosis must be rejected.

Keywords: modus ponens; McGee's counterexample; pragmatics; conditionals.

1. Introduction

In his seminal paper, McGee (1985) questioned the validity of Modus Ponens (MP, hereinafter) for indicative conditionals by resorting to a number of puzzling examples that haunt the literature on conditionals to this day. In that work, McGee offers several examples where one feels compelled to accept the premises and yet is left with a sense of puzzlement or perplexity concerning the conclusion. Here is McGee's most often quoted example:

Opinion polls taken just before the 1980 election showed the Republican Ronald Reagan decisively ahead of the Democrat Jimmy Carter, with the other Republican in the race, John Anderson, a distant third. Those apprised of the poll results believed, with good reason:

- (1) If a Republican wins the election, then if it's not Reagan who wins it will be Anderson.
 - (2) A Republican will win the election.
- Yet they did not have reason to believe
- (3) If it's not Reagan who wins, it will be Anderson. (McGee 1985, 462, numbering added)

McGee's article prompted a number of responses. Some of them can be classified as pragmatic responses: in particular, there are those that contend that the examples are ambiguous (Lowe 1987; Fulda 2010), and those that explain the counterexamples away by looking at the *acceptability* of the conclusion in light of the premises plus some epistemic or contextual considerations (Bledin 2015; Sorensen 1988).

On the one hand, solutions that resort to ambiguity maintain that there are different conditionals involved in the argument, so that some of them are indicative and some of them are material. Lowe (1987) claims that the embedded conditional in the first premise is material while the one in the conclusion is indicative, hence the examples do not constitute instances of MP after all. Fulda (2010), in turn, maintains that while the main conditional in the first premise is indicative, the example is perfectly valid because the conditional in the conclusion is material. On the other hand, the epistemic-contextual solutions can be classified into two sub-groups: informational approaches and epistemic approaches. The former argue that it is not truth preservation what we should be looking at, but rather preservation of acceptance of bodies of information. An argument preserves acceptance of bodies of information insofar as, if it incorporates a body of information that incorporates all its premises, it

also incorporates its conclusion by virtue of its logical form (Bledin 2015). In this revised sense of logical consequence, MP is not threatened, because even if it preserves bodies of information, it fails to preserve justification, an epistemic principle (not a logical one). The latter explains away the alleged counterexamples by noting that there are true sentences that might not be assertible because of pragmatic considerations (as in Jackson's 1987 sense of assertibility), so even if the conclusion of McGee's example is not assertible, it still preserves truth (Sorensen 1988, 449-451).¹

Our goal in this article is to explain away McGee's alleged counterexamples to MP by identifying two (so far unrecognized) confounding factors in the argument. Thus, if we are on the right track, McGee's examples do not justify abandoning MP for indicative conditionals. We will first point to a source of *ambiguity* in the Reagan example. In this sense, our view resembles to some extent ambiguity solutions. However, unlike those views, we identify the indefinite determiner phrase "a Republican" as the culprit, not the conditional. Admittedly, it is possible to reformulate the argument without using indefinites, but as we shall see these reformulations suffer from similar problems. Secondly, we contend that the intuitions surrounding McGee's examples mingle with pragmatic considerations concerning the felicity conditions for asserting the conclusion. Thus, our view resembles to some extent Sorensen's strategy, although we identify a different source of pragmatic infelicity, namely the existence of an ignorance inference typically licensed by conditionals. Unlike Sorensen, we are able to explain away the counterexamples without committing to any semantics for the indicative conditional. Thus, by recognizing different sources of pragmatic infelicity we can explain why this counterexample seems so puzzling: every time we think we pinpointed the problem, we keep on feeling there is still something missing. We argue that this happens because there is not just one particular problem but at least two, and these two problems come from the pragmatic realm, not the semantic one.

It is worth emphasizing that in what follows we will not commit to any particular semantics for indicative conditionals. To be sure, whether MP for the indicative conditional is valid depends on one's semantic choices concerning the conditional as well as on the notion of logical consequence one adopts. However, both issues are controversial, so any

¹ There are plenty of responses that cannot be framed into this broad picture of pragmatic solutions. We only mention those that bear some similarities to our own proposal in order to better situate our response on the theoretical map.

response to McGee that crucially relies on views concerning these matters will be, *prima facie*, controversial too. Hopefully, our view circumvents these problems by showing that the intuitions that might cast doubts upon the acceptability of MP can be explained away by resorting to pragmatic considerations which are independent of those issues.

The structure of the article is as follows. In Section 2, we discuss the ambiguity of the English indefinite determiner “a”, namely its specific and non-specific interpretations. After these preliminaries, we show, in Section 3, that McGee’s Reagan-example is ambiguous in exactly this sense due to the presence of the indefinite determiner phrase “a Republican” in both premises. We discuss alternative plausible disambiguations of the argument, which we circumscribe to two: a fully non-specific interpretation and a mixed interpretation where the indefinite in the first premise is interpreted non-specifically whereas the indefinite in the second premise is interpreted specifically. We show that the fully non-specific interpretation results in an intuitively acceptable instance of MP, when looked under the lens of truth preservation as well as when thought of in terms of assertibility preservation. As to the mixed interpretation, we identify two concurrent problems. First, we suggest that if one considers the overall content of the premises, including the specific dimension of meaning introduced by the indefinite article in the second premise, the mixed interpretation is not an intuitively acceptable instance of MP. In order to be able to pin down the second problem, we take a brief detour, in Section 4, to discuss the phenomenon of ignorance inferences. Finally, in Section 5, we go back to the mixed interpretation and explain away its intuitive unacceptability by arguing that an ignorance inference (commonly triggered by conditionals) makes the assertion of the conclusion pragmatically infelicitous.

2. Some basic notions regarding indefinites

It has long been acknowledged in the philosophical and linguistic literature that the English indefinite article has at least two readings, a specific and a non-specific one (e.g. Fodor and Sag 1982; Wilson 1978; Chastain 1975; Strawson 2017; Von Heusinger 2011). Arguably, there are different kinds of specificity (Farkas and Brasoveanu 2020), but we will focus on two of them. The first one is called *epistemic specificity* (Farkas and Brasoveanu 2020; Kamp and Bende-Farkas 2019). Take the following example:

- (4) A student in Syntax 1 cheated on the exam.

Here, the indefinite article “a” allows for both a specific and a non-specific interpretation. According to the former, the speaker has a particular student in mind, which she can tell apart somehow from other students within the relevant domain set, and she seeks to express a proposition about that individual. According to the latter, in turn, the speaker is just claiming that there is some student in Syntax 1 that cheated on the exam, but she cannot tell who. Thus, epistemic specificity concerns the way in which the use of an indefinite is related to the informational state of the speaker who uses it. This notion has been explained in modal terms (Alonso-Ovalle and Menéndez-Benito 2010, 2013). Epistemically specific indefinites prompt an interpretation according to which all the possible worlds that are compatible with what the speaker believes are worlds where the same individual satisfies the relevant property, e.g. cheating on the Syntax 1 exam. By contrast, on the non-specific readings, the possible worlds compatible with what the speaker believes contain at least two worlds such that different individuals satisfy the relevant property in each of them, i.e. where different students cheated on the exam.

The indefinite “a” also behaves ambiguously when it interacts with operators like conditionals, quantifiers, or intensional verbs, among others. In such cases, ambiguity is usually explained in terms of scope, thus the name *scopal specificity*. Take the following examples:

- (5) John wants to marry a linguist.
- (6) If a friend of mine from Texas dies in a fire, I will inherit a fortune.

(5) can be understood either as stating that there is a specific linguist that John wants to marry, or in turn as conveying that John wants to marry some person or other that is a linguist. Arguably, these two interpretations are due to a syntactic ambiguity having to do with the scope of the existential quantifier relative to the intensional verb: the wide scope reading corresponds to the specific interpretation, while the narrow scope reading corresponds to the non-specific one. This kind of ambiguity may also arise for indefinites in the antecedent of an indicative conditional. For example, in which case will I inherit a fortune, according to (6)? On the specific, wide-scope reading I will inherit a fortune if a specific friend of mine from Texas dies; on the non-specific, narrow-scope reading, in turn, I will inherit a fortune if the set of friends of mine from Texas who die is non-empty.

Finally, let us clarify a final issue. There is a debate concerning whether the (non-)specific dimension of meaning observable at the level of the

overall content of speech acts has to be couched in semantic terms, that is, as a difference in truth conditions or, in turn, it has to be dealt with within pragmatics. Put differently, one may treat (non-)specificity as part of conventional meaning, or in turn as conveyed through a pragmatic inference. Luckily, as shall become apparent below, we need not settle this issue here in order to make our point: if we are on the right track, McGee's alleged counterexamples are undermined either way. It suffices for our purposes to explain away the intuitions prompted by McGee's example, to note that indefinites in fact admit both readings. With this in mind, let us turn to McGee's example.

3. Back to McGee's example

In this section, we proceed to disambiguate McGee's example, which contains the ambiguous expression "a Republican" in both premises. We shall argue that the first premise only admits a non-specific interpretation (the specific interpretation does not make sense), while the second premise may be interpreted either specifically or non-specifically. This leaves room for two possible disambiguations of the argument, a fully specific one and a mixed one. We discuss both interpretations in turn.

3.1 Disambiguating the first premise

As we saw in discussing (6), the indefinite article may give rise to scopal ambiguity when occurring in the antecedent of an indicative conditional. Thus, *prima facie* (1) has two possible interpretations, the non-specific, narrow scope interpretation, and the specific, wide scope interpretation, which we paraphrase informally below:

(7) Specific and non-specific interpretations of (1):

- a. *non-specific*: If some Republican or other wins the election, then if Reagan doesn't win the election, Anderson will.
- b. *specific*: If Reagan wins the election, then if Reagan doesn't win the election, Anderson will.

Recall that under its specific interpretation, "a" (combined with a nominal phrase) expresses a content about a particular individual, about which the speaker seeks to convey a proposition. Arguably, given the context provided by McGee this individual must be Reagan (the first candidate in the opinion polls, as well as a Republican). Hence, the specific interpretation of (1) conveys that there is a particular Republican, namely Reagan, such that, if he will win the election, then if Reagan will

not win, then Anderson will. Clearly, this is not how we intuitively understand the first premise. In fact, this interpretation barely makes sense, if it makes sense at all (we are considering a non-material reading of the conditional). For this reason, in what follows we shall focus on the non-specific reading of (1), according to which there is some Republican or other such that, if Reagan does not win the election, then Anderson Will.

We have paraphrased the non-specific reading by using the expression “some *noun phrase* or other” (from now on we will use NP instead of “noun phrase”) because it is an indefinite that preserves the same core existential meaning as “a” but, unlike the latter, which is unmarked and may receive either reading, it forces a non-specific interpretation. To see the point, consider the following cases (see Abenina-Adar 2020, 101-102) (we will use the symbol “#” to mean “This is an infelicitous assertion”):

- (8) context: We saw Kim buying *War and Peace*.
 - a. #Kim bought some book or other.
 - b. Kim bought a book.

- (9) context: We saw Kim buying a book, but we didn’t see which book it was.
 - a. Kim bought some book or other.
 - b. Kim bought a book.

As the examples make clear, while “a” can be felicitously used either in contexts that favor a specific or a non-specific interpretation (namely contexts where the existential witness is considered to be something that is not identifiable in any salient way or, by contrast, where the identity of the existential witness is obvious), “some NP or other” is infelicitous when used in contexts that reinforce a specific reading. We can further press the point by considering the behavior of both kinds of expressions in combination with certain continuation discourse-moves or responses that contradict a non-specific reading (see also Abenina-Adar 2020, 101-102):

- (10) A: Kim bought a book.
 - a. A: Namely, *War and Peace*.
 - b. A: It was *War and Peace*.
 - c. B: What was it?
 - d. A: ... guess which one?

- (11) A: Kim bought some book or other.
 a. A: #Namely, *War and Peace*.
 b. A: #It was *War and Peace*.
 c. B: #What was it?
 d. A: # ... guess which one?

We observe that continuation discourse-moves that favor a specific interpretation are infelicitous after “some NP or other”, which shows that, unlike “a”, such expression has only a non-specific reading. Bearing this in mind will be of help later when we assess the disambiguated versions of McGee’s argument.

3.2 Disambiguating the second premise

The second premise contains the ambiguous expression “a Republican” as well, hence it can also be interpreted either specifically or non-specifically:

- (12) **Specific and non-specific interpretations of (2):**
 a. non-specific interpretation of (2): Some Republican or other will win the election.
 b. specific interpretation of (2): Reagan will win the election.

In this case, the indefinite is not in the scope of any relevant operator, so it is standard to interpret the (non-)specific ambiguity in epistemic terms. Recall that epistemic specificity concerns the informational state of the speaker. So, on its specific interpretation, and given the context put forward by McGee, (2) conveys that all the epistemic possible worlds compatible with what the speaker believes are worlds where Reagan wins the election. By contrast, on its non-specific interpretation, (2) conveys that at least one epistemic possibility compatible with what the speaker believes is such that Reagan wins, and at least one epistemic possibility compatible with what the speaker believes is such that Anderson wins. We consider both alternatives below.

Before proceeding to disambiguate the argument, it is worth noting that although both interpretations of the second premise make sense (unlike what we saw for the first premise), the context provided by McGee, where the witness of the existential claim introduced by the indefinite is assumed to be known (we are told that Reagan leads the polls by a significant difference), strongly favors a specific interpretation, that is, one where all epistemic possibilities compatible with what the speaker believes are such that Reagan wins.

3.3 Disambiguating the argument

In subsection 3.1, we saw that (1) only makes sense if understood non-specifically. Once we have settled on this interpretation, there are two alternatives left: we may understand (2) either non-specifically, which generates a fully non-specific interpretation of the argument, or specifically, which generates a mixed interpretation. Let us see what is going on in each case.

3.3.1 The fully non-specific interpretation

We start by considering the fully non-specific interpretation of the argument:

- (13) a. If some Republican or other wins the election, then if Reagan doesn't win the election, Anderson will.
 b. Some Republican or other will win the election.
 c. Therefore, if Reagan does not win the election, Anderson will.

This version of the argument seems to be an intuitively acceptable instance of MP. It is worth noting that by “acceptable” we do not necessarily mean “valid”: since we have not committed to any semantics for the indicative conditional, we cannot make claims about validity. It suffices to note, however, that once it has been disambiguated in this way, the argument completely loses its intuitive appeal, i.e. it is hard to see how anyone would be moved to abandon MP for the indicative conditional based on an example like (13).

3.3.1.1 A comment on validity as assertibility preservation

There are reasons to believe, nevertheless, that even if we analyze validity not in terms of truth preservation but as something weaker—say, probability, credence, or assertibility² preservation, like McGee originally intended—this disambiguation might still be problematic in a different sense.³ Probability preservation amounts roughly to having a threshold on the probability of the conclusion with regards to the probabilities of each premise of a valid argument, given this probabilistic sense of validity. In short, probability is preserved when, if one assigns high probability to

² As it was mentioned in the introduction, we use the technical term “assertibility”, following Jackson (1987) and Adams (1975).

³ We thank an anonymous referee from *European Journal of Analytic Philosophy* for pressing us on this point.

each premise of a probabilistically valid argument (and one has a small number of premises), then the probability of the conclusion cannot be too low. To see how this concept of validity works in relation to MP, one has to have a hypothesis on how to calculate the probability of an indicative conditional. The most common hypothesis one can find in the literature is *the Stalnaker Thesis*, which states that the probability of an indicative conditional “If A then B”, $P(A \rightarrow B)$, equals the conditional probability of B given A, $P(B|A)$ (Stalnaker 1968).⁴ A big problem with Stalnaker’s Thesis, however, is that it trivializes if one allows the probability function to apply to nested conditionals, as Lewis (1976) (and several authors after him) showed.

This is where the concept of assertibility becomes relevant. If we restrict the language to sentences where conditionals can only appear as the main connective, then we can assign something similar to probabilities, namely a degree of assertibility to conditionals, such that the assertibility of a conditional “if A then B”, $As(A \rightarrow B)$, equals the probability of B given A (for the rest of the language assertibility equals probability). This hypothesis is known as *Adams’ Thesis* or plainly as *The Thesis* (Adams 1975) (AT hereinafter). Then, assertibility is preserved when, if one assigns a high assertibility degree to each premise of an assertably valid argument (and one has a small amount of premises), then the assertibility degree of the conclusion cannot be too low.

So, how would assertibility preservation work for MP, given AT? Well, if we are dealing with a MP without nested conditionals, we would have an argument of the form:

- (i) $A \rightarrow B$
- (ii) A
- (iii) Therefore, B

Suppose that the assertibility of (i) and (ii) is high enough, let us say $As(A \rightarrow B) = As(A) = 0.9$, then, $As(B)$ has to be at least 0.81. Now, if we were to have a way to translate (at least) right-nested conditionals such as the one in McGee’s example into a non-nested conditional, maybe we could check if, under that translation t , if $As(t(A \rightarrow (B \rightarrow C)))$ and $As(A)$ are high enough, then they preserve some degree of assertibility to the conclusion, $B \rightarrow C$. Recall, McGee’s argument was of the form:

- (iv) $A \rightarrow (B \rightarrow C)$

⁴ $P(A \rightarrow B) = P(B|A) = P(A \& B|A)$

- (v) A
- (vi) Therefore, $B \rightarrow C$

However, as Stern and Hartmann (2018) show, if we look at synchronic probabilities the answer is negative. If we assume the validity of the inference called Import, namely:

(IMP) $A \rightarrow (B \rightarrow C)$ implies $(A \& B) \rightarrow C$

or we pose a translation function t such that it imports nested conditionals as a conjunction of both antecedents such that $t(A \rightarrow (B \rightarrow C)) = (A \& B) \rightarrow C$, then we can show that the assertibility of the conclusion (vi) can be 0 even if the premisses (iv) and (v) have, let us say, an assertibility degree of 0.9. As the authors argue, the reason for this is that after some math, we can check that $As(B \rightarrow C) = P(C|B) = P(C|A \& B). P(A|B) + P(C|not A \& B). P(not A|B)$, and even if $P(C|A \& B)$ might be high (because we assumed it as a premise), we have no information regarding $P(A|B)$, $P(C|not A \& B)$ and $P(not A|B)$, which can be extremely low, or even 0. In this particular sense, one could argue that McGee's example, even under its fully non-specific interpretation, might still work as a legitimate counterexample to the validity of MP, where validity is understood as assertibility preservation.

There are several things to say in this regard. First of all, one can show that even if there are cases where the assertibility of the premisses of a MP with a right-nested conditional as its main premise is high whereas the assertibility of its conclusion is 0, the fully non-specific interpretation of McGee's alleged counterexample is not such a case. Keep in mind that we are considering the fully non-specific interpretation, where "a Republican" must be understood non-specifically as "some Republican or other". There is only one reason you may have to assert the second premise under such interpretation, namely that you think that both candidates have chances of winning, something that clashes with the contextual information stipulated in the example. Put differently, on a fully non-specific interpretation, asserting the second premise in the context of the example would be either infelicitous (if one treats (non-)specificity as pragmatically inferred) or simply false (if one treats (non-)specificity as part of conventional meaning).⁵ If we assign a high

⁵ Remember that assertibility mirrors probabilities for simple sentences and simple conditionals, so when two propositions are mutually exclusive the probability of a disjunction equals the sum of the probability of each disjunct, $P(A \text{ or } B) = P(A) + P(B)$ and so does $As(A \text{ or } B)$. In this particular context, we have an existential claim of the form $\exists x (Rx \text{ and } Wx)$, meaning there is some x that is a

assertibility degree of winning to each candidate, then that scenario is not a particular case where we can assign high assertibility to (iv) and (v) but low to (vi), because:

$$\begin{aligned} \text{As}(\text{Reagan loses} \rightarrow \text{Anderson wins}) &= \text{P}(\text{Anderson wins} | \text{Reagan loses}) \\ &= \text{P}(\text{Anderson wins} | \text{Some Rep or other wins \& Reagan loses}) \\ &+ \text{P}(\text{Anderson wins} | \text{both Rep lose \& Reagan loses}) \\ &+ \text{P}(\text{both Rep lose} | \text{Reagan loses}) \end{aligned}$$

And under the assumption that we think both candidates have good chances of winning, already the left side of the sum must have a high degree of assertibility.

Admittedly, someone might push and say that this argument runs for assertibility, but not for degrees of belief or credences. But note that in order to do the math on the assertibility of this right-nested conditional, we made several non-trivial assumptions about conditionals and assertibility. The same goes for probabilities and probabilities of conditionals. We assumed that it is possible to validate both AT and IMP (or some way to translate right-nested conditionals into a conjunction of both antecedents) without trivializing. But we know since Lewis (1976) that we cannot do that in a classical context. So, if we want to make that possible, we need to adopt some non-classical probability function such as de Finetti's (1936) or Egré, Rossi, and Sprenger (forthcoming)'s three-valued probability function. At the same time, there are other ways of handling this, as Stern and Hatmann (2018) show, using diachronic probabilities.

Summing up, under several non-trivial assumptions, MP with right-nested conditionals can be problematic with regards to assertibility preservation, but in the fully-non-specific disambiguation of McGee's argument this is not a problem. Of course, there is still another scenario where we assert the existential because of the high chances of one of the candidates. In Section 4 we will explain why we think this is also not a problematic case.

3.3.2 The mixed interpretation

If we disambiguate (2) specifically we arrive at the following:

Republican and will win, and because we only have two Republican candidates, we can see how the probability of the existential claim equals the probability of the disjunction.

- (14) a. If some Republican or other wins the election, then if Reagan doesn't win the election, Anderson will.
 b. Reagan will win the election.
 c. Therefore, if Reagan doesn't win the election, Anderson will.

In the light of (14), one may be tempted to claim that the argument, thus understood, is not an instance of MP, since (14b) does not match the antecedent of the main conditional in (14a). This would straightforwardly explain away the intuition that the argument is unacceptable without the need to posit an additional ambiguity in the conditional, as Fulda (2010) does.

Things are not so simple though, for the specific reading of the indefinite, which we paraphrased as “Reagan will win the election”, seems to logically entail the non-specific one, which we paraphrased as “some Republican or other will win the election”, by means of the existential introduction rule. Hence, it seems that there is a MP below the surface after all. But then, why do we have the intuition that the argument is unacceptable?

There are two alternative strategies for dealing with this objection. Recall that in Section 2 we have characterized these two readings by saying that epistemically specific uses prompt an interpretation according to which all the possible worlds that are compatible with what the speaker believes are such that the same individual satisfies the relevant property, while epistemically non-specific uses convey that the possible worlds compatible with what the speaker believes contain at least two worlds such that different individuals satisfy the relevant property in each of those worlds. In other words, the specific interpretation ascribes to the speaker an epistemic state that is incompatible with the epistemic state ascribed to the speaker by the non-specific one. Thus, when we consider the overall content of each use of the indefinite, that is, its existential core meaning plus the (non-)specific dimension, we conclude that the specific interpretation is incompatible with the non-specific one. What kind of incompatibility is at play depends again on one's preferred view concerning the ultimate source of the (non-)specific flavor of indefinites: if one treats specificity as part of conventional meaning, one may deny that the specific interpretation entails the non-specific one; if, by contrast, one assumes that (non-)specificity is a pragmatic phenomenon, then one must say in turn that the specific and the non-specific interpretations have different assertibility conditions. At any rate, this suffices to explain why

the argument, in its mixed interpretation, is intuitively unacceptable. Hence, the objection is blocked.

The second strategy places the burden of the explanation in the relation between the second premise (understood specifically) and the conclusion. To be able to develop this strategy, however, we need first to discuss the phenomenon of ignorance inferences. We address this issue in the next section.

Before moving on, however, a final objection needs to be addressed. One may argue that the previous discussion about indefinites is beside the point, since it is possible to restate McGee's argument without relying on the indefinite article. Two alternative restatements have been suggested to us:

- (15) a. If the winner of the election is conservative, then if it's not Reagan who wins it will be Anderson.
 - b. The winner of the election will be conservative.
 - c. If it's not Reagan who wins it will be Anderson.

- (16) a. If Reagan or Anderson wins, then if Reagan doesn't win, Anderson will.
 - b. Reagan or Anderson will win.
 - c. If Reagan doesn't win, Anderson will.

As we shall see, both reformulations suffer from problems. Below, we address the reformulation in (15). In the next section we consider (16) and argue that it is of no help either (again, paying attention to ignorance inferences will reveal itself crucial here).

The challenge posed by (15) can be met by noting that definite descriptions may give rise to a kind of ambiguity concerning the informational state of the speaker, similar to the one raised by specific and non-specific uses of indefinites. In order to see the point, consider the distinction between *role-type* descriptions and *particularized* descriptions (Rothschild 2007). The distinction depends on the way in which the common ground interacts with the content of the description. The former are descriptions where it is part of the common ground i) that the description is uniquely satisfied across a broad variety of different possible worlds and ii) the unique satisfier of the description varies amongst these possible worlds; the latter, by contrast, are descriptions where it is not common ground that a unique individual satisfies the content of the description in a wide range of possible worlds. As Rothschild (2007, 76) puts it, when a description counts as particularized

“we can only know that there is a single most salient individual satisfying the descriptive content (and thus the description picks some individual out) by having some sort of knowledge particular to the narrow conversational context”. By contrast, when a description is role-type, we may not be able to do that, but we still have knowledge that the uniqueness requirement associated with definite descriptions is satisfied.

Now, recall that (non)specificity concerns the informational state of the agent and may be understood in modal terms: on a specific reading, all the possible worlds that are compatible with the speaker’s belief-state are worlds where the same individual satisfies the relevant property; on a non-specific reading, there are at least two possible worlds compatible with the agent’s belief-state such that different individuals satisfy the relevant property in each of them. The ambiguity between role-type and personalized definite descriptions gives rise to a similar phenomenon, in the sense that definite descriptions may be compatible with the agent being in two different informational states: in personalized uses, all the worlds compatible with the agent’s belief-state are such that the same individual, namely the unique most salient individual that satisfies the content of the description, has the relevant property; by contrast, in role-type uses, although uniqueness is satisfied in all the worlds compatible with the speaker’s belief-state, the individual which satisfies the description may vary, i.e. there are at least two worlds compatible with the agent’s belief-state such that a different individual satisfies the content of the description in each of them.⁶

Going back to (15), note that the description “the winner of the election” in the second premise gives rise to a role-type/particularized ambiguity. On the one hand, it is most naturally understood as a role-type description, since it is common ground both that there is only one winner in presidential elections and that the winner may vary; on the other hand, however, the context provided by McGee to interpret the example stipulates that it is common ground that it is highly probable that Reagan will win the election, and so the description may be interpreted as a particularized description denoting the unique most salient individual in the context who will win the election, namely Reagan. Hence, the premise “The winner of the election will be conservative” gives rise to an ambiguity in the sense that it is compatible with the agent being in two different belief-states: on the particularized interpretation, the agent

⁶ Rothschild argues for this distinction appealing to differences in scope concerning metaphysical modality, but the distinction intuitively extends to epistemic modality as well (see Rothschild 2007, 77 and fn. 7).

believes, based upon accessible information in the narrow conversational context, that the unique most salient individual in the context will win the election (namely Reagan, in light of the contextual information surrounding McGee's example); on the role-type interpretation, by contrast, the agent has knowledge that a unique individual will win the election, but her belief-state may be compatible with different individuals winning the election in different epistemic possibilities. Now, since the first premise only makes sense if the description is interpreted as a role-type one, disambiguation leaves us with two interpretations of the argument, a fully role-type interpretation, and a mixed interpretation. Then, the same kind of arguments we offered for the case of indefinites can be rehearsed for this version of the argument.

4. Conditionals and ignorance

In this section, we lay some basic facts about ignorance inferences that will help us understand what is to come. Here are two examples of ignorance inferences (we read “+ >” as “it is conversationally implicated”):

- (17) If John is still alive, he has two sons. (Gazdar 1979)
- a. + > The speaker is ignorant of whether John is still alive.
 - b. + > The speaker is ignorant of whether John has two sons.

- (18) Mary was late or John was late.
- a. + > The speaker is ignorant of whether Mary was late.
 - a. + > The speaker is ignorant of whether John was late.

Very roughly put, ignorance inferences are licensed by the Maxim of Quantity: given an uttered sentence *S*, if *A* is a more informative alternative than *S* and it is relevant in the context, the Maxim of Quantity licenses the inference that *A* is not in the speaker's belief set. Since it is standard to assume that relevance is closed under negation (that is, if a sentence *A* is relevant, then its negation, $\neg A$, is also relevant), the Maxim of Quantity also licenses the inference that $\neg A$ is not in the speaker's belief set. In other words, the audience infers that the speaker ignores *A*.

Identifying the alternatives to a given sentence is a complex matter we need not discuss in detail here, but it is standard to assume that both the antecedent and the consequent of a conditional (as well as sentences in their Boolean closure) are alternatives to the conditional itself. It is

important to emphasize, however, that not all alternatives will be considered in order to calculate ignorance inferences, but only those that are relevant in the context. We follow Roberts (2012) and define relevance in terms of the Question Under Discussion (QUD). Roughly put, we assume that an alternative is relevant relative to a contextually salient QUD if it provides a total or a partial answer to it. As we said, it is also assumed that relevance is closed under negation, so that if a sentence A is relevant to a given QUD, then $\neg A$ is also relevant to the same QUD.

Let us see how ignorance inferences work in (17). Since the antecedent is an alternative to the conditional, and assuming for the sake of the example that the antecedent is relevant to the contextually salient QUD, the Maxim of Quantity licenses the inference that the proposition that John is still alive is not in the speaker's belief set. But since the negation of the antecedent is also an alternative to the conditional, and relevance is closed under negation, the Maxim of Quantity licenses the inference that the proposition that John is not alive is not in the speaker's belief set as well. In other words, the audience infer that the speaker ignores (in the sense specified above) whether John is still alive. An analogous explanation can be formulated in order to account for the ignorance inferences licensed by (18) (note that both disjuncts (and their negations) are alternatives to the disjunction itself and are relevant to the QUD).

Let us call the overall content of an utterance, that is, its literal content plus its implicated contents, the strengthened meaning of the utterance. It has been noted that an utterance may turn out to be infelicitous in certain contexts by virtue of its strengthened meaning. This phenomenon has been extensively discussed for scalar implicatures (see, e.g. Magri 2009; Fox 2007; Chierchia 2004; Schlenker 2012, among many others). To illustrate the point, consider the following sentence:

(19) #Some Italians come from a warm country. (Magri 2009)

Sentence (19) generates a sense of oddness or pragmatic infelicity. Arguably, here is why. In normal contexts, it is common ground that all Italians come from the same country, which is also a warm one. Now, (19)'s truth conditional content, namely *some Italians come from a warm country*, is perfectly compatible with this common ground (in fact, the common ground entails it). But it is known that sentences like (19) typically trigger a scalar inference, namely that *not all Italians come from a warm country*. So, the strengthened meaning of the utterance, that is, its literal meaning plus the scalar inference, amounts to *some, but not all, Italians come from a warm country*. Crucially, unlike its truth conditional meaning, the strengthened meaning of (19) is inconsistent with the

common ground. Arguably, this leads to pragmatic infelicity: the speaker implicates some content that contradicts what is commonly accepted by the participants of the conversation.

Now, it has been recently pointed out that this effect may arise for ignorance inferences as well (Singh, 2010; Fox, 2016; Meyer et al., 2013; Buccola and Haida, 2019). By way of illustration, consider again sentence (17), and compare it with (20):

- (17) If John is still alive, he has two sons.
 a. + > The speaker is ignorant of whether John is still alive.
 b. + > The speaker is ignorant of whether John has two sons.
- (20) # If I am married to an American, I have two sons.
 (Singh, 2010)
 a. + > The speaker is ignorant of whether she is married to an American.
 b. + > The speaker is ignorant of whether she has two sons.

Arguably, the difference between (17) and (20) can be explained in similar terms as example (19). Both conditionals trigger ignorance inferences, namely that the speaker ignores both the antecedent and the consequent. The reason why only (20) is infelicitous is that in that case the strengthened meaning of the conditional, which includes the ignorance inferences, is inconsistent with what normally constitutes common ground, namely that people know who they are married to, whether they have sons and how many. By contrast, we do not necessarily assume that speakers know whether other people are still alive or not, so without further specifications about the context we tend to judge (17) as felicitous.⁷ Thus, indicative conditionals trigger ignorance inferences which may lead sometimes, in particular when they contradict the common ground, to pragmatic infelicity.⁸ With this in mind, we are now in position to go back to where we left, that is, to the mixed interpretation of McGee's argument.

⁷ Again, we are assuming for the sake of the example that the antecedent and the consequent are relevant relative to the contextually salient QUD.

⁸ There is a debate about whether these facts about scalar and ignorance inferences force us to abandon a Gricean approach to implicatures and adopt an alternative view, according to which these implicatures are triggered by the grammar. This is orthogonal to our goals in this article, however. It suffices for the purposes of this work to call attention to the existence of the phenomenon.

5. Back to the objections

Consider the mixed version of the example again, repeated here for ease of exposition:

- (21) a. If some Republican or other wins the election, then if Reagan doesn't win the election, Anderson will.
 b. Reagan will win the election.
 c. Therefore, if Reagan doesn't win the election, Anderson will.

Arguably, the conclusion triggers an ignorance inference like the ones we have been discussing. First, as we saw above the antecedent of a conditional is an alternative to the conditional itself. Second, the contextually salient QUD in McGee's context seems to be "Who will win the election?". The antecedent of the conclusion "Reagan does not win" is relevant relative to this QUD, since it provides an answer to it. In addition, since relevance is closed under negation, the antecedent's negation "Reagan wins" is relevant too. As a result, the Maxim of Quantity licenses an ignorance inference, namely that the speaker neither believes nor disbelieves that Reagan will win the election:

- (22) If Reagan doesn't win the election, then Anderson will win.
 a. + > The speaker is ignorant of whether Reagan will win.

The crucial point here is that the strengthened meaning of the conclusion, which comprises its truth-conditional meaning plus the ignorance inference above mentioned, creates a problem when combined with a specific interpretation of the second premise (as in the mixed interpretation of the argument). To see why, consider the following: McGee tells us that "Those apprised with the poll results believed, with good reasons" the premises of the argument, in particular premise (2), that a Republican will win the election. If we interpret that premise *specifically* in the context provided by McGee, this means that the speaker believes that Reagan will win the election. Moreover, arguably in accepting this premise we include it in the common ground, that is, we accept it and take it for granted there on for the purposes of the conversation. However, this common ground directly contradicts the strengthened meaning of the conclusion, according to which the speaker neither believes nor disbelieves that Reagan will win the election. As a result, interpreting the second premise of the argument specifically puts one in a position where affirming the conclusion is infelicitous. In other

words, interpreting the second premise specifically pragmatically blocks the conclusion.

In this way we can explain why, when the second premise is interpreted specifically, McGee's argument is intuitively unacceptable. Moreover, we are able to offer an explanation in terms of our intuitions about the felicity conditions of certain speech acts, and not in terms of our intuitions about validity or by assuming a certain semantics for indicative conditionals.

Finally, let us address a final worry. In Section 3, we considered the objection that McGee's argument can be restated using disjunctions, without appealing to indefinites:

- (15) a. If Reagan or Anderson wins, then if Reagan doesn't win, Anderson will.
- b. Reagan or Anderson will win.
- c. If Reagan doesn't win, Anderson will.

Crucially, this version of the argument is not problematic (it does not provide a counterexample to MP) unless one is able to provide a context in which both premises are intuitively acceptable whereas the conclusion is not. The burden of proof is here on MP detractors, but to be sure, the new version of the argument is not intuitively appealing in the original context provided by McGee. As we saw, disjunctions license ignorance inferences. In particular, premise (15b) licenses the inference that the speaker is ignorant of (neither believes nor disbelieves) whether Reagan will win the election and whether Anderson will win the election. Now, in the original context provided by McGee it is common ground that Reagan is "decisively ahead" of Carter and Anderson on the polls, with Anderson a distant third. Uttering (15b) would be clearly infelicitous in such context since the strengthened meaning of the sentence (either *Reagan or Anderson will win, but I don't know who*) clashes with what is common ground. Hence, restating McGee's argument using disjunctions would undermine its intuitive appeal.

It may be argued that even in the original context of the example, the speaker has reasons to believe that Reagan will win the election, and this entails (15b) by the rule of the introduction of disjunction. Hence, (15b) is in fact acceptable in such a context. The answer to this worry is parallel to the one we provided for the case of the mixed interpretation of the original argument. Let us grant that the speaker accepts (15b) because she infers it from the implicit premise "Reagan will win the election". If this is so, we can offer an alternative pragmatic explanation for the resistance

one feels to accept the conclusion: the implicit premise “Reagan will win the election” clashes with the ignorance inference licensed by the antecedent of the conditional in the conclusion, namely that the speaker is ignorant of whether Reagan will win the election. Thus, the uneasiness caused by the example can be pinned on pragmatics and need not force us to abandon MP for indicative conditionals.

The same explanation can be used to understand the problems discussed in section 3.3.1.1 about assertibility preservation. Suppose that we accept AT as well as all the other non-trivial assumptions that allow us to show that the degree of assertibility of the conclusion can be drastically lower than the degree of assertibility of the premises of the argument. Nevertheless, in the scenario where the assertibility degree of “Some Republican or other will win” or “Either Reagan or Anderson will win” is high only because the assertibility degree of “Reagan will win” is high, asserting either the existential or the disjunction is pragmatically infelicitous because of the ignorance inferences these sentences trigger. As a matter of fact, this is at the core of Adams’ proposal: asserting a disjunction while knowing only one of its members has a high assertibility degree is misleading and “runs against standards of correct communication” (1965, 15).

Of course, one could argue that even if not assertible, it is reasonable to assign a high degree of belief to each premise and a low degree of belief to the conclusion. Yet, assertibilities are not probabilities, and in principle, if we adopt AT, the assertibility function speaks of how confident we are in asserting sentences not in believing them. It is true though, that we could move to some non-classical probability function like Egré, Rossi and Sprenger do, and then we could rephrase the argument in terms of degrees of belief and check that if we analyze MP with right-nested conditionals as premises in a synchronic way, then the degree of belief of the conclusion can drastically descend. Yet, we have to bear in mind that we are standing in a quicksand, since none of these assumptions are conventional.

6. Summing up

On these pages, we have argued that McGee’s most famous counterexample to MP can be explained away by noting the interference of two different pragmatic phenomena, without assuming anything about the semantics of indicative conditionals. These phenomena are related, in the sense that both involve the transmission of information concerning the epistemic state of a speaker that accepts both the premises and the

conclusion, yet they arise in parallel through different mechanisms. In particular, we pointed out that both premises are ambiguous. We then argue that once they are disambiguated, the example loses its intuitive appeal. On the one hand, the fully non-specific version is an intuitively acceptable instance of MP. On the other hand, we identified two problems for the mixed interpretation. First, if one considers the overall content of the premises, including pragmatically conveyed information concerning the (non-)specific dimension, it fails to be an instance of MP. The second, concurrent problem, was the conditional conclusion “If Reagan doesn’t win the election, Anderson will win” triggers the implicature that the speaker ignores whether Reagan will win the election, which contradicts the specific interpretation of the second premise. The fact that there are two concurring pragmatic phenomena at play interfering with the example is probably the reason why it is so hard to pin down the problem with this argument.

In summary, the point we would like to draw from this work is that whether MP is valid or not will be a decision the semanticist or the logician will have to make independently of these supposed counterexamples, since the reasons one would raise against believing their conclusions in light of believing their premises are not related to truth or belief preservation.

Acknowledgments

We are grateful to Natalia Buacar, Ramiro Caso, Eleonora Cresto, Lucas Rosenblatt, Damián Szmuc, Diego Tajer, and Paula Teijeiro for their attentive reading, to the audience from the WIP seminar of the BA Logic group in Buenos Aires, and to three anonymous reviewers for the *European Journal of Analytic Philosophy* for helpful comments and suggestions on previous versions of this article.

Funding information. This research was supported by the National council for scientific and technological investigations (CONICET), Argentina and PLEXUS (Grant Agreement no 101086295), a Marie Skłodowska-Curie action funded by the EU under the Horizon Europe Research and Innovation Programme.

REFERENCES

- Abenina-Adar, Maayan. 2020. *Expressing ignorance with determiner phrases*. PhD thesis, University of California.
<https://escholarship.org/uc/item/8964606g>
- Adams, Ernest W. 1975. *The Logic of Conditionals*. Dordrecht: D. Reidel.
- Alonso-Ovalle, Luis, and Paula Menéndez-Benito. 2010. “Modal indefinites.” *Natural Language Semantics*, 18 (1): 1–31.
<https://doi.org/10.1007/s11050-009-9048-4>
- Alonso-Ovalle, Luis, and Paula Menéndez-Benito. 2013. “Two Views on Epistemic Indefinites.” *Language and Linguistics Compass*, 7 (2): 105–122. <https://doi.org/10.1111/Inc3.12009>
- Bledin, Justin. 2015. “Modus Ponens Defended.” *The Journal of Philosophy*, 112 (2): 57–83.
<https://www.jstor.org/stable/43820889>
- Buccola, Brian, and Andreas Haida. 2019. “Obligatory Irrelevance and the Computation of Ignorance Inferences.” *Journal of Semantics*, 36 (4): 583–616. <https://doi.org/10.1093/jos/ffz013>
- Chastain, Charles. 1975. “Reference and Context.” In *Language, Mind, and Knowledge: Minnesota Studies in the Philosophy of Science, volume 7*, edited by Keith Gunderson, 194–269. University of Minnesota Press.
- Chierchia, Gennaro. 2004. “Scalar Implicatures, Plarity Phenomena, and the Syntax/Pragmatics Interface.” In *Structures and Beyond: The Cartography of Syntactic Structures, volume 3*, edited by Adriana Belletti, 39–103. Oxford: Oxford University Press.
- de Finetti, Bruno. 1936. “La logique de la probabilité.” In *Actes du congrès international de philosophie scientifique, volume 4*, 1–9. Paris: Hermann Éditeurs.
- Égré, Paul, Rossi, Lorenzo, and Jan Sprenger. Forthcoming. “Certain and Uncertain Inference with Indicative Conditionals.” *Australasian Journal of Philosophy*. <https://arxiv.org/abs/2207.08276>
- Farkas, Donka F., and Adrian Brasoveanu. 2020. “Kinds of (Non)-specificity.” In *The Blackwell Companion to Semantics*, edited by Daniel Gutzmann, Lisa Matthewson, Cecile Meier, Hotze Rullmann, and Thomas Ede Zimmerman, 1–26. Blackwell.
<https://doi.org/10.1002/9781118788516.sem037>
- Fodor, Janet D., and Ivan A. Sag. 1982. “Referential and Quantificational Indefinites.” *Linguistics and philosophy*, 5 (3): 355–398.
- Fox, Daniel. 2007. “Free Choice and the Theory of Scalar Implicatures.” In *Presupposition and implicature in compositional semantics*, edited by Uli Sauerland and Penka Staveva, 71–120. New York: Palgrave.

- Fox, Daniel. 2016. "On Why Ignorance Might Be Part of Literal Meaning. Commentary on Marie Christine Meyer." In *Talk given at the MIT Workshop on Exhaustivity*.
- Fulda, Joseph S. 2010. "Vann McGee's Counterexample to Modus Ponens: An Enthymeme." *Journal of Pragmatics*, 42 (1): 271–273.
- Gazdar, Gerald. 1979. *Pragmatics: Implicature, Presupposition and Logical Form*. New York: Academic Press.
- Gibbard, Allan. 1980. "Two Recent Theories of Conditionals." In *Ifs*, edited by William Leonard Harper, Robert Stalnaker, and Glenn Pearce, 211–247. Dordrecht: D. Riedel.
- Jackson, Frank. 1979. "On Assertion and Indicative Conditionals." In *Philosophical Review*, 88 (4): 565–589. <https://doi.org/10.2307/2184845>
- Kamp, Hans, and Agnes Bende-Farkas. 2019. "Epistemic Specificity from A Communication-Theoretic Perspective." *Journal of Semantics*, 36 (1): 1–51. <https://doi.org/10.1093/jos/ffy005>
- Lowe, Edward J. 1987. "Not a Counterexample to Modus Ponens." *Analysis*, 47 (1): 44–47. <https://doi.org/10.1093/analys/47.1.44>
- Lewis, David K. 1976. "Probabilities of Conditionals and Conditional Probabilities." *Philosophical Review*, 85: 297–315
- Magri, Giorgio. 2009. "A Theory of Individual-Level Predicates Based on Blind Mandatory Scalar Implicatures." *Natural language semantics*, 17: 245–297. <https://doi.org/10.1007/s11050-009-9042-x>
- McGee, Van. 1985. "A Counterexample to Modus Ponens." *The Journal of Philosophy*, 82 (9): 462–471.
- Meyer, Marie C. 2013. *Ignorance and Grammar*. PhD thesis, Massachusetts Institute of Technology. <http://hdl.handle.net/1721.1/84420>
- Roberts, Craige. 2012. "Information Structure: Towards an Integrated Formal Theory of Pragmatics." *Semantics and Pragmatics*, 5: 6–1. <https://doi.org/10.3765/sp.5.6>
- Rothschild, Daniel. 2007. "Presuppositions and Scope." *The Journal of Philosophy*, 104(2): 71–106. <https://www.jstor.org/stable/20620000>
- Schlenker, Philippe. 2012. "Maximize Presupposition and Gricean Reasoning." *Natural Language Semantics*, 20: 391–429. <https://doi.org/10.1007/s11050-012-9085-2>
- Singh, Raj. 2010. "Oddness and Ignorance Inferences." Handout presented at Modularity Reading Group at MIT. <https://carleton.ca/singhr/wp-content/uploads/singh-oddness-ignorance-handout.pdf>

- Sorensen, Roy. A. 1988. "Dogmatism, Junk Knowledge, and Conditionals." *The Philosophical Quarterly*, 38 (153): 433–454. <https://doi.org/10.2307/2219708>
- Stalnaker, Robert. 1968. "A Theory of Conditionals". In *Studies in Logical Theory* edited by Nicholas Rescher, 98-112. Oxford: Blackwell. https://doi.org/10.1007/978-94-009-9117-0_2
- Stern, Reuben, and Stephan Hartmann. 2018. "Two Sides of Modus Ponens." *Journal of Philosophy* 115 (11): 605-621. <https://www.jstor.org/stable/48568445>
- Strawson, Peter F. 2017. *Subject and Predicate in Logic and Grammar*. London: Routledge. <https://doi.org/10.4324/9781315242132>
- von Heusinger, Klaus. 2011. "Specificity." In *Semantics: Noun Phrases and Verb Phrases*, edited by Paul Portner, Klaus Heusinger, and Claudia Maienborn, 70-111. Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110589443-003>
- Wilson, George. 1978. "On Definite and Indefinite Descriptions." *The Philosophical Review*, 87 (1): 48–76. <https://doi.org/10.2307/2184347>

BOOK REVIEW

Amy Kind
IMAGINATION AND CREATIVE THINKING
Cambridge University Press, Element Series, 2022
ISBN 10: 1108977227
Paperback: £17.00

Ksenija Savčić¹ 

¹University of Rijeka, Faculty of Humanities and Social Sciences, Croatia

Amy Kind in her book “Imagination and Creative Thinking” provides a comprehensive overview of the current discussion of imagination and creativity, defending the skill-based framework for understanding both concepts. The book consists of an introduction, four chapters, and a conclusion.

In the Introduction, Kind states her intention to explore the concepts of imagination and creativity. Her aim is to clarify both concepts and their relation. Kind challenges the view of these abilities as inborn, arguing instead that they are best understood as skills that can be developed through practice.

In Chapter 2, Kind provides a concise yet comprehensive overview of the literature on imagination, defining it as a speculative mental activity. She presents three vignettes to illustrate the diverse uses of imagination: (1) two children involved in the pirate-themed pretend play, (2) a man deciding on the shade of a new wall color, and (3) a woman trying to make sense of her coworker’s disagreement. The first two vignettes represent visual imagination, while the third exemplifies emotional or experiential imagination. Kind argues that these vignettes also help distinguish between different aims of imagination, specifically between instructive and transcendent uses. Instructive uses enable the imaginer to learn about the world, whereas transcendent uses allow escape from the world as it is. Thus, the first vignette illustrates transcendent imagination, while the second and third vignettes exemplify instructive imagination.

Kind adopts the division of imagination into two broad types: propositional and sensory. Propositional imagination is best understood in comparison with belief and is contrasted with other propositional attitudes, such as desires, hopes, and intentions, through the direction of fit. Kind explains the direction of fit as follows: “For beliefs, we try to fit our mind to the world; for desires, we try to fit the world to our mind” (6). The main difference between propositional imagination and belief is belief’s orientation toward truth. The second type of imagination—sensory imagination—is understood by analogy with perception, with the primary distinction being the degree of vividness of the experience.

Kind argues that both propositional and sensory imagination share common features: will-dependence (i.e., the fact that imagining is subject to our will) and world-sensitivity (i.e., unlike perception, imagination does not track changes in the world and does not adjust to them). Thus, perception tracks changes in the world and adapts to them, while imagination doesn’t. Kind argues that both types share another common feature, mental imagery. She advocates for an imagery-based account of imagination, arguing that, despite some disagreement, this account “can be motivated in large part by contrasting imagining with other speculative activities such as supposing and conceiving” (12).

In Chapter 3, Kind discusses the concept of creativity. She notes the widespread consensus that originality (or novelty, which she treats as equivalent) and value are necessary conditions for creativity, though there is disagreement on their sufficiency. Kind provides an overview of two taxonomies of creativity, which she views as complementary rather than competing.

The first taxonomy is illustrated through three vignettes: (1) an inspiring speaker, (2) a particularly creative approach to solving a math problem, and (3) a creative chess move selected by a computer program. These vignettes represent three distinct types of creativity: person-creativity, process-creativity, and product-creativity. Kind points out that there is debate over whether one of these forms is more fundamental than the others.

The second taxonomy, presented by Margaret Boden (2004), distinguishes between psychological and historical creativity, endorsing novelty, value, and surprisingness as necessary criteria for creativity. Kind examines the value requirement, tracing its origins to Kant’s assertion that creativity must be exemplary. She acknowledges several challenges associated with this requirement, including its demanding nature, lack of clarity, and the implication that negative creations cannot

be deemed creative. In response, Kind advocates for a more nuanced approach, endorsing Livingstone's (2018) distinction between instrumental and intrinsic value. The author criticizes Boden's surprisingness requirement, arguing that the distinction between surprisingness and novelty must be clarified before this requirement can be met. She supports her skepticism with examples, such as children finding basic laws of physics surprising despite their being well-known to adults, and a scientist who, after extensive work on a project, sees the creative output as inevitable rather than surprising. Based on these examples, Kind concludes that the surprisingness requirement is not widely accepted. The chapter concludes with a discussion of intentional agency as an additional criterion for creativity. Kind argues for its acceptance on two grounds: first, describing something as creative implies an evaluative component that "might presuppose agency" (31); and second, "an agency requirement proves useful in explaining the connection between imagination and creativity" (31).

In Chapter 4, Kind examines the relationship between imagination and creativity, advancing her central claim that both concepts are best understood within a skill-based framework. She argues that although "imagination" and "creativity" are often used interchangeably (32), conflating the two would be a mistake. Kind asserts that while imagination is a mental activity, creativity is not confined to mental processes, as it is also a characteristic of persons and products. Nonetheless, she acknowledges that there is "something importantly right about the idea that creativity and imagination are tightly connected" (32) and identifies three possible relationships between them: imagination is (1) necessary; (2) sufficient; or (3) both necessary and sufficient for creativity. However, she observes that many instances of imagination do not lead to creative outcomes, and the second vignette from the previous chapter illustrates creativity without the use of imagination. Despite this, Kind argues that imagination should be regarded as a primary, and possibly the primary means of achieving creativity. She supports this claim by referencing Dustin Stokes (2014), who asserts that creativity requires cognitive manipulation—a function well-served by imagination. Kind notes that Stokes' perspective aligns with Boden's tripartite distinction of novelty, where cognitive manipulation is central. Kind concludes by suggesting that the concept of creativity can be applied broadly across the domains of science and art by utilizing the distinction between instructive and transcendent imagination, as discussed in Chapter 2. Thus, she argues that imagination typically undergirds creativity.

Kind's central claim is that both imagination and creativity are best understood within the skill-based framework, and that both can be developed and improved with adequate practice. She asserts that while there is no universal consensus on the definition of skill in philosophy, there is an agreement about at least three features of any skilled activity: (1) it can be evaluated; (2) it is practiced intentionally; (3) it can be improved via practice—all of which apply to imagination. Based on these criteria, she concludes: "If imagination undergirds creativity, and imagination can be trained, then that gives us reason to think that creativity can be trained as well" (45).

The final chapter explores the challenges surrounding the possibility of imagination and creativity in machines. Kind examines artworks generated by machines and explores the extent to which machines can engage in imaginative and creative processes. She focuses on two examples: Cohen's AARON and a Generative Adversarial Network (GAN) created by the Obvious Collective. Kind argues that the outputs of Cohen's AARON could be considered product-creative, as AARON's drawings meet the criteria of novelty and value for creativity. However, when shifting the focus from product-creativity to process-creativity and introducing requirements like surprisingness and agency, the same conclusion cannot be drawn. Even if AARON were considered an intentional agent, it does not follow that it exercises intentional agency in its drawings, as it merely follows a set of rules without a mechanism for individual judgment. Kind then discusses the case of GANs and Bellamy's portrait, noting that GANs present a more significant challenge because they operate differently from traditional computers. GANs consist of two components—the generator and the discriminator—which engage in an evaluative process that AARON lacks. However, Kind argues that, given the substantial human involvement in creating the algorithm, selecting the type of artwork, assembling the dataset, and ultimately choosing which output to present to the public, GANs fail to meet the criteria for agency. She addresses Turing and Chen's concept of human-centric bias, the tendency to view humans as superior to machines even when results are indistinguishable. Ultimately, Kind contends that "the question of whether a machine can be creative can be seen to reduce to the question of whether a machine can have imagination" (55). Since machines lack agency, they cannot be attributed with imagination and, consequently, cannot achieve process-creativity.

In conclusion, Kind leaves the question of a machine's potential for creativity open. She outlines the increasing pressure brought forward by ever more impressive machine achievements to expand our understanding of imagination and creativity beyond the human context.

The book presents an eloquent and elaborate discussion of imagination and creativity, successfully providing a clear distinction between them. However, in what follows, I will offer a critical reflection on one key aspect of Kind's argument—specifically, her skill-based framework for understanding imagination and creativity, which appears to rest on the following two conditionals:

1. "If creativity is a skill, and imagination typically undergirds it, then one might naturally expect that imagination would be a skill as well." (41)
2. "If imagination undergirds creativity, and imagination can be trained, then that gives us reason to think that creativity can be trained as well." (45)

Individually, each conditional raises concerns.

Regarding the first conditional, Kind argues that imagination and creativity "typically go together" (56), despite her argument that imagination is neither necessary nor sufficient for creativity. This raises concerns about the cogency of the assertion. She addresses the issue by noting that "the problem with these claims is not their attempt to link imagination and creativity but rather their attempt to cast the link as stronger than it in fact is" (36). Kind draws on Stokes's account to establish imagination as the most suitable candidate for the role of cognitive manipulation, as it involves "voluntarily thinking about the contents of some conceptual space in non-truth bound ways" (Stokes 2014, 171, as cited in Kind 2022, 36). However, significant aspects of Stokes's account of creativity that may not support Kind's proposal seem not to be present in her discussion.

Stokes (2014) argues that imagination is indeed central to creativity, but he maintains that his account of creativity is a minimal one, emphasizing the capacity of individuals for frequent creative thought qua novel thought. By acknowledging the mundanity of minimal creativity, Stokes effectively addresses potential criticisms related to the triviality objection. The triviality objection states that imagination may undergird a wide range of cognitive processes—such as belief formation, concept acquisition, inference, deduction, and abduction—since, according to some accounts, it is fundamental to all cognition due to its role in grasping meaning (see McGinn 2004). Stokes counters this by clarifying that his account "does not require rich imagination in its explanans" (Stokes 2014). As a result, his account does not suggest a positive correlation between imagination and creativity, where a more developed

imagination would necessarily lead to an increased generation of valuable new ideas.

However, Kind asserts precisely this kind of correlation. She explores methods for training imagination, such as imaginative matching, where one person describes an object in detail, and the other tries to visualize it accurately before comparing it to the actual image. Kind asserts that

[T]he kinds of imagination-stretching exercises just discussed seem perfectly suited toward enabling novelty, and depending on how they are structured, they could be well suited for achieving value as well. (45)

However, it remains unclear how these exercises would necessarily lead to new and valuable thoughts rather than merely fostering excessive daydreaming.

Regarding the second conditional, subchapter 4.2, “Imagination and Creativity as Skills” presents arguments supporting the claim that imagination is a skill that can be developed through practice and exposure to literature. In that part, Kind argues that imagination meets the three criteria for skill. However, despite the subchapter’s title, creativity is scarcely mentioned in this part of the text and is not evaluated against the same three criteria, but rather assumed to be a skill. This is problematic for the following reason.

The argument, as presented through the two conditionals, might appear circular. It assumes that creativity is a skill to justify the claim that imagination is also a skill. The reasoning then reverses to suggest that if imagination can be trained, so can creativity. However, the claim that imagination can be trained relies on the assertion that imagination is a skill, which, as we have seen, rests on the insufficiently justified premise that creativity is a skill. In this respect, the argument appears to beg the question.

Additionally, Stokes’s account of minimal creativity, which does not establish a positive correlation between imagination and creativity, may not fully support Kind’s argument. It is also worth noting that Stokes (2008) provides an account of what he terms “rich creativity”, explicitly stating that “agency and novelty are not sufficient for creativity” (Stokes 2008). Since he does not endorse the value requirement, he introduces a modal condition as a third criterion for creativity: “an F is creative only if F could not, relative to the cognitive profile of the agent in question, have been done or performed (by A) before the time it actually was” (Stokes

2008, 26). He revisits this modal requirement in Stokes (2014), emphasizing that “imagination plausibly plays a central role in the formation of modal beliefs” (Stokes 2014, 180). Stokes’s modal condition implies that achieving rich creativity necessitates the acquisition of new knowledge or skills, which differs significantly from the proposal to reach it by training imagination through exercises and reading.

Lastly, addressing the triviality objection would have been a valuable addition to Kind’s discussion. Without it, it remains unclear why the development of imagination would specifically enhance creativity and whether this implies that all other cognitive processes—merely in virtue of being undergirded by imagination—should also be understood within a skill-based framework.

Although I have provided some critical remarks, they are not meant as definitive judgments, but rather a starting point for the further discussion. Overall, I think the book provides a wide and insightful overview of the current discussion in philosophy of creativity. It is an engaging and eloquent read, one that I would recommend to anyone interested in philosophy of mind, aesthetics, and artificial intelligence.

References

- Boden, Margaret Ann. 2004. *The Creative Mind: Myths and Mechanisms*. 2nd ed. London: Routledge.
- Kind, Amy. 2022. *Imagination and Creative Thinking*. Cambridge: Cambridge University Press.
- Stokes, Dustin. 2008. “A Metaphysics of Creativity.” In *New Waves in Aesthetics*, edited by Kathleen Stock and Katherine Thomson-Jones, 105–24. Palgrave-Macmillan.
- . 2014. “The Role of Imagination in Creativity.” In *The Philosophy of Creativity*, edited by Elliot Samuel Paul and Scott Barry Kaufman. Oxford: Oxford University Press.



BOOK REVIEW

Holly Lawford-Smith
GENDER-CRITICAL FEMINISM
Oxford University Press, 2022
ISBN: 9780198863885
Hardcover: \$47.99

Eugen Golubić¹ 

¹University of Rijeka, Faculty of Humanities and Social Sciences, Croatia

Holly Lawford-Smith is a political philosopher with a particular interest in feminism. In her book, “Gender-Critical Feminism”, she offers a comprehensive examination of the gender-critical feminist perspective which prioritizes biological sex over gender identity. Her central goal is the abolition of the norms of femininity and to provide a theory of feminism that unapologetically advocates for the liberation of female people (60). In the introduction of the book, Lawford-Smith lays out the reasons for writing the book as well as the questions she wants to answer. She expressed concern that women’s issues have taken a secondary position, which she argues have diluted the focus on women’s issues by expanding the scope of feminism to include concerns beyond the direct oppression of women that is primarily based on sex. In her opinion, this was particularly evident during the observance of International Women’s Day as articulated by the International Women’s Day Melbourne Collective whose events are inclusive, welcoming cis and trans women, non-binary, and gender-diverse individuals from various backgrounds (2-3). Her objective is for feminism to reclaim its roots and act as a movement initiated by women for women, instead of turning into a movement that encompasses other global justice concerns. Throughout the book, the author critically examines the prevailing tenets of socially dominant feminism and attempts to validate her perspective by claiming that women are adult human females, thereby establishing the criteria for what qualifies as a woman and what does not.

The book includes ten chapters that are divided into two parts (chapters 2-6 form the first part, while chapters 7-9 constitute the second part). In the first part, Lawford-Smith aims to explain what gender-critical feminism is, while in the second part, she discusses the feasibility of

gender-critical feminism and its relationship with liberalism. The book starts with the introduction (chapter 1) and concludes with “A Gender-Critical Manifesto” (chapter 10).

After outlining the motivation for writing the book in the introduction, chapter 2, titled “Gender-Critical Feminism’s Radical Roots”, explains that gender-critical feminism originates from radical feminism, a movement associated with second-wave feminism, which began in the late 1960s. Following an examination of key figures in second-wave radical feminism alongside influential intellectuals like Mary Wollstonecraft, Simone de Beauvoir, and Gerda Lerner to establish the historical context of the movement, she describes the sustaining mechanisms of female oppression, understood as oppression based on sex differences, and various solutions that are offered.

In chapter 3, titled “Gender-Critical Feminism”, Lawford-Smith outlines the foundations of gender-critical feminist theory. Central to this perspective is the view that biological sex plays a pivotal role. In this chapter, the author argues that biological sex is both a standard scientific term and a key factor in understanding women’s oppression. She asserts that it is integral to the concept of gender, as it shapes the social meanings attributed to it (47). Additionally, she maintains that recognizing relevant differences between men and women, such as physical distinctions, is crucial for protecting women in contexts like the workplace and sports (48).

In chapter 4, titled “The Sex Industry”, she critiques prostitution as exploitative and argues that it reinforces the notion that women’s primary purpose is to fulfill men’s sexual desires. She explores the reasons why most women engage in sex work, questions what and who can be bought or sold, and examines possible policy models. Ultimately, she advocates for the Nordic model, which criminalizes buyers and facilitators of prostitution while decriminalizing prostituted persons. She argues that this approach protects women from trafficking and exploitation by targeting the demand for prostitution within the sex industry.

In chapter 5, titled “Trans/Gender”, the author critiques gender identity approaches to defining what it means to be a woman, which suggest that being of a particular gender involves self-identification rather than biological factors. Gender-critical feminism, by contrast, grounds the concept of womanhood in biological sex and understands feminism as a movement focused on addressing the oppression of females. As such, gender-critical feminists reject identity-based approaches to sex, as these remove the essential reference to biological sex when defining

womanhood. In this chapter, the author also addresses the question of whether gender-critical feminism is “trans-exclusive”. She argues that gender-critical feminism is not “anti-trans”. Instead, she frames it as a movement centered on the liberation of females, which includes groups such as trans men and female nonbinary individuals (115). In chapter 6, Lawford-Smith continues explaining why she thinks that gender-critical feminism is unjustifiably vilified, focusing on the fundamental moral disagreement and political propaganda from various feminist and progressive circles.

In the second part of the book, Lawford-Smith poses the so-called “hard questions”, since it addresses complex and controversial matters for gender-critical feminism. Among these she includes questions such as whether gender-critical feminism is intersectional, whether it is feasible, and whether it is liberal. In chapter 7, she argues against the idea that feminism should be intersectional, although acknowledging the interplay between gender and other forms of discrimination like race, socioeconomic status, physical and mental ability, etc. (145). In chapter 8, she discusses the feasibility of the gender-critical feminist movement, where she generally endorses government intervention, especially in contexts where she thinks women’s rights and protections are threatened, including sports, workplace discrimination, and violence (174). Chapter 9 discusses liberalism and its roots in the works of John Locke and John Stuart Mill, and the chapter focuses on highlighting the similarities and differences between the gender-critical feminist movement and liberal ideology.

In the final, chapter 10, titled “A Gender-Critical Manifesto”, she summarizes her claims, emphasizing the importance of ending of male violence against women, protecting women’s health and bodily autonomy, and supporting women’s freedom of thought and their access to and full participation in public life.

One of the advantages of this book is the precise articulation of the author’s views. She was able to distinguish gender-critical feminism from other forms of feminism while attempting to find common ground with other feminists without deviating from its original claims. Moreover, the book offers connections to a wide range of philosophical fields, including epistemology, philosophy of language, philosophy of mind, ethics, and political philosophy, making it potentially relevant to philosophers with diverse interests.

However, as might be expected, while Lawford-Smith highlights specific challenges faced by women and many of her calls to end oppression

against women are well-founded, the book also presents several controversial claims. The book faced petitions and protests before its publication, as the author's views can be interpreted as aligning with those often labelled as TERF (trans-exclusionary radical feminist), which are frequently associated with transphobia (118).¹ These issues seem to stem from one of the central claims of the book that grounds feminism in the idea of biological sex differences between men and women. This framing of feminist theory raises important questions: can gender-critical feminism effectively address the contemporary challenges affecting women, as well as the diverse conceptualizations of womanhood across disciplines and social practices? To what extent does this approach engage with broader concerns about inclusivity and intersectionality within contemporary feminist theory, while recognizing that issues affecting women as women may also intersect with broader social issues impacting other groups?

There is also room for disagreement on issues in the book that have provoked less controversy, such as those related to prostitution. While it is widely acknowledged that significantly more women and even children than men are coerced into sex work, and such coercion is undeniably morally wrong, one might still question why an informed and consenting adult woman cannot freely offer sexual services and require the state to regulate such transactions. After all, despite the Nordic model's decriminalization of sex workers, it remains associated with well-known problems linked to criminalizing aspects of prostitution. These issues include pushing sex work underground, which can exacerbate stigma, discourage sex workers from seeking help, limit access to healthcare, and reduce the reporting of violence.

In conclusion, I think that the book is thought-provoking and invites further discussion across several philosophical and ethical fields of inquiry, including their intersection with biology and the social sciences. While many progressive voices view it as controversial, it must be acknowledged that achieving universal agreement on complex issues is virtually impossible. Disagreement is inevitable, and fostering open dialogue is essential in an open society. For this reason, and despite the controversies surrounding it, I would recommend this book to anyone interested in exploring the different strands of contemporary feminist theory, as well as the challenges it faces and aims to resolve.

¹ For media coverage of the controversy surrounding the publication of the book, along with commentaries provided by philosophers, see the following posts on *Daily Nous*: "OUP Responds to Letter Regarding *Gender-Critical Feminism*" and "OUP's Decision to Publish *Gender-Critical* Book Raises Concerns of Scholars and OUP Employees".



BOOK REVIEW

Sanja Dembić
***PHILOSOPHY OF MENTAL DISORDER: AN ABILITY-
BASED APPROACH***
Routledge, 2023
ISBN: 9781032435466
Hardcover: £108.00
eBook: £31.99

Luca Malatesti¹ 

¹ University of Rijeka, Faculty of Humanities and Social Sciences, Croatia

Sanja Dembić in *Philosophy of Mental Disorder: An Ability-Based Approach* integrates two general ideas. Methodologically, she adopts an explicationist method, which defines concepts to serve theoretical and practical objectives. This method underpins her defence of the Reability View, an inability-based account of mental disorder. According to this view, mental disorders are conditions marked by harmful inability to respond to reasons.

The explicationist method and inability-in-reason-responsiveness accounts of mental disorder have already been proposed in the philosophy of psychiatry, yet Dembić's account is innovative. Traditional a priori descriptive methods favoured by analytic philosophers have faced criticism for their inconclusiveness in defining mental disorder. This has led some to adopt conceptual explication's prescriptive method. Moreover, there are other proponents of the inability-in-reason-responsiveness account, notably George Graham. However, Dembić is the first to defend this type of account explicitly through the explicationist method. Her further contributions to the debate are detailed in the following chapter summaries.

In the first chapter, Dembić advances and defends her explicationist project. She identifies four goals for the concept of mental disorder, that she calls adequacy conditions. Theoretically, the proper applications of the concept should (1) distinguish pathological from healthy conditions and (2) societal deviations, and (3) mental from bodily disorders. Practically, the concept of mental disorder should (4) provide the affected

individuals a *pro tanto* reason to seek treatment and mitigate some of their social, moral, or legal obligations. A *pro tanto* reason provides initial justification that remains valid unless outweighed by stronger, overriding considerations. She adopts a conservative approach to current use of the concept of mental disorder, insofar the stated goals remain attainable. She critiques a priori conceptual analysis and rejects extrapolating empirically the concept from paradigmatic cases, as proposed by Dominic Murphy.

The second chapter considers Harm Views and Action Views of mental disorder. Regarding the Harm Views, Dembić considers those advanced by Danner Clouser, Charles Culver, and Bernard Gert; Lawrie Reznick; and Rachel Cooper, who agree that harm is necessary but insufficient for defining disorders. Clouser, Culver, and Gert add that disorders lack sustaining external causes; for instance, strangulation is not a disorder since removing the external cause resolves the harm. Reznick defines disorder as an abnormal mental condition harming species-typical individuals in standard circumstances and requiring medical intervention. Cooper describes disorder as a bad, unfortunate, and potentially treatable condition. Regarding the Action Views, Dembić considers the positions of William Fulford and Lennart Nordenfelt. These accounts ground the concept of mental disorder in agential failure. According to Fulford, a disorder (illness in his terminology) involves the experience of a failure of ordinary doing, acting as we intend to act, in the absence of obstruction or opposition. For Nordenfelt, an agent is disordered when they do not have the agential abilities to achieve their vital goals, these goals being those whose fulfilment is necessary and jointly sufficient for a minimal amount of welfare and happiness.

Dembić criticizes all the above positions but also highlights certain positive aspects. She attacks each of them for specific shortcomings but emphasizes that, collectively, they fail to meet the adequacy conditions (58-64). She argues that they cannot, in relevant cases, distinguish mental disorders from non-pathological conditions, bodily disorders, or deviations from social, political, or ethical rules. Additionally, they fail to explain satisfactorily why a mental disorder provides a *pro tanto* reason for treatment. However, she acknowledges that harm is undoubtedly a central element of the concept of disorder, though it needs further specification beyond what the theories considered so far have provided. Similarly, she recognizes that agential inabilities are promising concepts for explicating that of mental disorder.

In the third chapter, the discussion shifts to Biological Function Views, which link mental disorders to biological dysfunctions. Dembić considers

the influential proposals of Jerome Wakefield and Christopher Boorse. According to Wakefield, a mental disorder constitutes harmful dysfunction, with dysfunction understood in evolutionary terms. Boorse, while acknowledging the clinical and practical relevance of harm, identifies dysfunction itself as central to mental disorder. Diverging from Wakefield, Boorse adopts a goal-directed account of dysfunction. Specifically, he defines the function of a mechanism in an organism as determined by the overall goals of that type of organism, which are, in turn, established by a reference class of exemplars of that organism.

In her critique of Wakefield's and Boorse's positions, Dembić considers key aspects of the broader debate they have generated. Regarding Wakefield's account, amongst other shortcomings that she identifies (73-80), she maintains that this account makes psychiatry overly and impractically dependent on speculative or future evolutionary accounts of mental mechanisms. Moreover, this account imposes overly stringent explanatory requirements: mental disorder causes must be tied to hereditary traits established through evolution.

Regarding Boorse's account, she focuses her criticism, amongst other issues (84-87), on his defining function and dysfunction based on a reference class of individuals. If all members of a species suffered from a particular disease, they would paradoxically be classified as healthy. Furthermore, she shows the limitations of these proposals when they are tested in terms of the criteria of adequacy (87-90). Both accounts fail in recognizing certain clear cases of mental disorder. While both can account for the difference between mental disorders and departures from social, political, or ethical norms, Boorse is not able to give a satisfactory account of the practical normative goals, while the concept of harm in Wakefield's account has this capacity.

The heart of her framework, the Rehability View is presented in the fourth chapter. This account defines mental disorders as harmful inability to act on normative reasons. To give an idea of the complexity and precision of the definition, I quote here the full version of the account, that is meant to account for the clearest types of cases that should fall under the concept of mental disorder:

RHApsy An individual S has a mental disorder if and only if S does not have the ability to respond adequately to some of their available (apparent) reasons for (or against) some of their reason-sensitive attitudes or actions; in view of their mental constitution and their life circumstances (where the threshold of inability is determined by the degree at which

individuals in the relevant comparison class are, on average, harmed by their condition C in some respect X) and ii S is harmed by their condition C in some respect X, where X is some component of what makes S's life a non-instrumentally good one for S.

For S to be harmed by C in some respect X, it is sufficient that i C is non-instrumentally bad for S in some respect X ii C causes S to be sufficiently worse than before or that they would otherwise have been in respect X; or iii C prevents S from receiving a good in respect X, thereby leaving S in a non-instrumentally bad state in respect X, where X is some component of what makes S's life a non-instrumentally good one for S. (Dembić 2024, 126-127)

Dembić discusses all the main concepts that enter this complex definition. Here I briefly examine her views on the central concepts of ability, (apparent) reason, and harm.

Other authors have identified abilities as central to analysing mental disorders. Dembić deepens this type of account with insights from ontological investigations on abilities developed outside the philosophy of psychiatry. She proposes four characteristics of abilities. First, an ability to ϕ is a modal property of an agent, reflecting what they can or could do. Second, abilities can be agentive, involving actions, or non-agentive, where actions are not involved. For mental disorders, both types of inability are relevant. For example, according to Dembić, schizophrenia involves the non-agentive inability to avoid holding certain beliefs that they have reason not to endorse. Compulsive disorder, instead, involves an inability to resist acting on obsessive thoughts, although they have reason not to do so. Third, abilities exist in degrees. An individual's inability to respond to apparent reasons is significant when it is, compared to appropriately individuated peers, low enough to harm members of that reference group. Finally, abilities depend on specific facts. For example, a swimmer can swim, but this is compromised with a broken arm. Regarding mental disorders, relevant abilities or inabilities must be identified based on internal mental constitutions and external conditions, what Dembić terms "life circumstances", which the agent cannot easily change.

Central to Dembić's account are the inabilities to respond with appropriate attitudes and actions to normative reasons and *apparent* normative reasons (110-120). For Dembić, a normative reason is given by a fact or a true proposition that counts in favour of (or against)

responding with an action or attitude. An apparent normative reason would count in favour (or against) a certain action if it were true. For example, anxiety disorder arises from a harmful inability to avoid frequent fear despite beliefs providing (apparent) reasons against it (120-126).

Dembić addresses complex issues regarding the proper characterization of harm (113-119). Although she acknowledges the provisional nature of her proposal, she argues that psychiatric harm must consider both the individual's subjective perspective and objective criteria independent of it. Thus, she advocates a hybrid view. A central feature of her account is that the harmfulness of an inability to respond to reasons depends on the reference class. This requires identifying relevant classes, by means of age or other contextual factors, and determining the threshold where inability becomes psychiatrically harmful.

The fifth chapter defends the Reliability View. First, Dembić argues that this view meets the adequacy conditions outlined in the first chapter. For example, she tests its descriptive adequacy in distinguishing health from pathology by examining principal mental disorders classified so far. For each disorder type, she identifies the relevant agentive or non-agentive abilities to respond to reasons central to these disorders. Additionally, she demonstrates how her account avoids the difficulties faced by positions she criticized in the second and third chapter.

Moreover, she addresses two objections to her account (155-161). First, it could be argued that her account assumes the mental is constituted just by sensitivity to reasons, thus excluding mental states as our sensations. She states that she is not assuming this. Instead, she considers the concept of the mental that is relevant to psychiatry. For practical purposes, psychiatry focuses on conditions involving insensitivity to reasons, particularly in interventions like discursive therapy. Second, it could be argued that harm is not a necessary requirement for a mental disorder. She acknowledges the possibility of a purely theoretical concept of mental disorder based on inability to respond to (apparent) reasons. However, she argues that since her goal is to provide a relevant concept that offers *pro tanto* reasons for intervention, the concept of harm must be retained.

A case study on addictive disorders, in the last chapter, underscores the framework's practical value. Addiction, often framed as a failure of willpower or a biological defect, is reconceived as an inability to act on reasons. This reinterpretation demonstrates the Reliability View's capacity to inform both philosophical debates and clinical practice.

This book examines mental disorder, focusing on ability, reasons, harms, all concepts central to unresolved controversies in philosophy. For instance, Dembić endorses what appears a metaphysically realist account of normative reasons. Not all readers may find this endorsement convincing.

There may also be internal tensions between Dembić's conceptual work and psychiatric theory and practice, rooted in her methodological assumptions. Her explication of the concept of mental disorder does not appear to be maximally conservative. In fact, it is not in line with descriptive psychopathology. For example, she tests the adequacy of her proposed explication of mental disorder against currently classified conditions, by showing that they might involve inability to respond to reasons. However, existing diagnostic categories prioritize behavioural symptoms and inferred mental states and traits rather than inability. Bridging this gap requires empirical research, which could shift her project from conservation to revision.

Her treatment of harm introduces further challenges. Referencing harm to a class of individuals raises two significant issues. First, her concept may inadequately safeguard individual perspective and rights. For example, individuals with comparable inability to respond to reasons may evaluate the associated harm differently and still be ascribed the same mental disorder. In this case, they should not have the same *pro tanto* reason for being treated and excused. Second, this framework risks criticisms like those Dembić has directed at Boorse's use of class of reference to fix biological dysfunction. Variations in the class of reference would determine too revisionary variations of the disorder status of certain conditions.

Finally, she incorporates reason-responsiveness into her explication, raising the question of whether this embeds a non-revisable assumption about the understanding of the notion of the mental relevant for mental disorder. Dembić asserts compatibility with most views of the mind-body problem, excluding eliminativism, but this claim requires further exploration. In the philosophy of psychiatry, the mind-body relation also involves determining the appropriate level of description or explanation for identifying standards relevant to disorder status. Even if her arguments against using biological dysfunctions are compelling, other standards for mental disorder might go beyond normatively individuated reason-responsiveness. For example, advances in computational accounts, such as the Bayesian brain framework in psychiatry, might suggest alternative empirically grounded views. Integrating the conceptual

elaboration of disorder standards with empirical research should remain open to revision. However, Dembić seems to treat reason-unresponsiveness as a conceptual and fixed criterion for determining when a condition qualifies as a mental disorder.

Despite these tensions, *Philosophy of Mental Disorder* is a remarkable book for its clarity, depth, and innovation in understanding mental disorders. It seamlessly bridges abstract theory, often drawing on areas of philosophy not typically applied to the philosophy of psychiatry, with practical concerns. By employing the method of explication, it makes an invaluable contribution to conceptual debates on the nature and treatment of mental health. This ground-breaking book is essential reading for philosophers, clinicians, and anyone invested in the conceptual foundations of mental health.

Acknowledgements

Many thanks to Sanja Dembić, Miguel Núñez de Prado Gordillo, and Marko Jurjako for reading and commenting on previous versions. This work was supported by the project PACS, funded by the University of Rijeka under contract UNIRI-ISKUSNI-HUMAN-23-14.



ABSTRACTS (SAŽECI)**Artistic Freedom Realised**

Murray Smith

University of Kent, United Kingdom

ABSTRACT

The fewer the constraints we encounter in an artistic endeavour, the greater our artistic freedom; as technology advances and presents us with more numerous options and ever-greater creative flexibility, so our artistic freedom burgeons. Such is the folk wisdom on the phenomenon. Is the wisdom wise? Many artists and art theorists think not; for Stravinsky, “[t]he more constraints one imposes, the more one frees one’s self of the chains that shackle the spirit”. I consider the orthodox view and this alternative view through an exploration of the constraints arising from physical, psychological, and technological factors, as well as the heterogeneous array of routine practices we label “conventions”. To deepen the analysis, I turn to the “problem-solving” perspective on artistic creativity, examining the interplay between invention and convention in art, drawing on a variety of examples from film and rock music. On the theory advanced by Jon Elster, the problems with which artists engage arise from a mix of chosen, imposed, and invented constraints; on the theory of David Bordwell, those problems can be solved through the replication, revision, synthesis, or rejection of existing solutions. I conclude that neither the folk theory, nor the alternative theory, are correct; rather, there is a “sweet zone” of artistic creativity poised between a disabling surfeit of options, and a stifling sparsity of them. I argue that Stravinsky’s counter-theory of artistic creativity, though not literally true, acts as a “felicitous falsehood”—an epistemically valuable overcorrection.

Keywords: artistic freedom; artistic creativity; shot/reverse shot; film noir; neo-noir; Peter Gabriel; Nick Cave.

Ostvarena umjetnička sloboda

Murray Smith

University of Kent, United Kingdom

SAŽETAK

Što je manje ograničenja s kojima se susrećemo u umjetničkom stvaranju, to je veća naša umjetnička sloboda; kako tehnologija napreduje i pruža nam sve više opcija i sve veću kreativnu fleksibilnost, tako naša umjetnička sloboda cvjeta. Tako glasi narodna mudrost o ovom fenomenu. No je li ta mudrost doista mudra? Mnogi umjetnici i teoretičari umjetnosti smatraju da nije; za Stravinskog, “[š]to više ograničenja namećemo, to se više oslobađamo lanaca koji sputavaju duh”. U radu razmatram ortodoksno gledište i ovu alternativnu perspektivu kroz analizu ograničenja koja proizlaze iz fizičkih, psiholoških i tehnoloških čimbenika, kao i iz heterogene skupine rutinskih praksi koje nazivamo “konvencijama”. Kako bih produbio analizu, oslanjam se na perspektivu “rješavanja problema” u umjetničkoj kreativnosti, ispitujući međudjelovanje između izuma i konvencija u umjetnosti, koristeći različite primjere iz filma i rock glazbe. Prema teoriji Jona Elstera, problemi s kojima se umjetnici suočavaju proizlaze iz mješavine izabranih, nametnutih i izmišljenih ograničenja; prema teoriji Davida Bordwella, ti se problemi mogu riješiti replikacijom, revizijom, sintezom ili odbacivanjem postojećih rješenja. Zaključujem da ni pučka teorija, ni alternativna teorija nisu točne; umjesto toga, postoji “idealna zona” umjetničke kreativnosti smještena između onemogućujućeg viška opcija i sputavajuće oskudice istih. Tvrdim da Stravinskijeva protu-teorija umjetničke kreativnosti, iako nije doslovno istinita, djeluje kao “sretna neistina”—epistemološki vrijedna pretjerana korekcija.

Ključne riječi: umjetnička sloboda; umjetnička kreativnost; kadar/obrnuti kadar; film noir; neo-noir; Peter Gabriel; Nick Cave.

Aesthetic Education: A Perceptual-Cognitive Model

Ted Nannicelli

The University of Queensland, Australia

ABSTRACT

Here is a puzzle about aesthetic education. In a variety of contexts, we commit significant time, energy, and resources to aesthetic education. We teach (and in many cases publicly subsidize) university courses and degrees that have aesthetic education as their primary aim; we also invest public resources into museums, including enrichment programs that are also designed to afford aesthetic education. It would seem that if our commitment to aesthetic education is rational, then aesthetic appreciation is something that can be done better or worse. However, we also, in a variety of contexts (oddly enough, some of them being the same sorts of contexts that are designed to abet aesthetic education), act as if it is true that there is no disputing taste. We may try to persuade students to come around to particular judgments, but we do not penalize students for judging one way or another.

The aim of this paper is to dissolve the apparent puzzle of aesthetic education by clarifying its aims and advancing a conception of it that deemphasizes the role of taste. I claim that, pace “the default view of aesthetic education” (as I shall call it), the primary purpose of aesthetic education is not to educate taste. It is, rather, to facilitate the development of certain perceptual-cognitive capacities so as to enhance aesthetic experience and improve aesthetic appreciation. Thus, I call the view of aesthetic education advanced here “a perceptual-cognitive model”.

Keywords: aesthetic education; aesthetic cognition; aesthetic normativity; aesthetic appreciation; taste.

Estetsko obrazovanje: perceptivno-kognitivni model

Ted Nannicelli

The University of Queensland, Australia

SAŽETAK

Evo zagonetke o estetskom obrazovanju. U raznim kontekstima posvećujemo značajno vrijeme, energiju i resurse estetskom obrazovanju. Podučavamo (i u mnogim slučajevima javno financiramo) sveučilišne kolegije i diplomske programe kojima je estetsko obrazovanje primarni cilj; također ulažemo javne resurse u muzeje, uključujući programe obogaćivanja koji su također osmišljeni kako bi omogućili estetsko obrazovanje. Čini se da, ako je naša predanost estetskom obrazovanju racionalna, tada je estetsko vrednovanje nešto što se može raditi bolje ili lošije. Međutim, također u raznim kontekstima (što je zanimljivo, neki od njih su isti oni konteksti koji su osmišljeni kako bi poticali estetsku edukaciju) ponašamo se kao da je istina da se o ukusu ne raspravlja. Možemo pokušati uvjeriti studente da prihvate određene sudove, ali ih ne penaliziramo ako prosuđuju na ovaj ili onaj način.

Cilj ovog rada je razriješiti prividnu zagonetku estetskog obrazovanja razjašnjavanjem njegovih ciljeva i razvijanjem pojma koji smanjuje naglasak na ulogu ukusa. Tvrdim da, unatoč “standardnom shvaćanju estetskog obrazovanja” (kako ću ga nazvati), primarna svrha estetskog obrazovanja nije edukacija ukusa. Umjesto toga, njegova je svrha omogućiti razvoj određenih perceptivno-kognitivnih sposobnosti kako bi se unaprijedilo estetsko iskustvo i poboljšalo estetsko vrednovanje. Stoga shvaćanje estetskog obrazovanja izneseno u ovom radu nazivam “perceptivno-kognitivnim modelom”.

Ključne riječi: estetsko obrazovanje; estetska spoznaja; estetska normativnost; estetsko vrednovanje; ukus.

Ethics Education from Suffering on Screen? Tragic Visions in Arrival

James MacAllister

The University of Edinburgh, United Kingdom

ABSTRACT

In this paper I argue that tragic films can have significant potential for ethics education when they prompt audiences to sympathise with

suffering on screen. I first summarise two accounts of the relationship between tragic art, moral education and aesthetic value (those provided by Rorty and Lamarque). I then discuss problems with these accounts and explain how a new criterion of aesthetic value might help to resolve them. I thereafter argue that tragic films have potential to ethically educate audiences in a way that enhances the aesthetic value of the films in at least three directions: by deepening moral understanding, by deepening understanding of the nature of human being and ethical purpose and by deepening understanding of ethical theory. I conclude by showing how Denis Villeneuve's film, *Arrival*, screens a tragic story with ethics education potential in each of the aforementioned senses.

Keywords: screen suffering; tragic film; ethics education through film; educational ethicism; ethics in Villeneuve's *Arrival*.

Etička edukacija kroz patnju na ekranu? Tragične vizije u filmu *Arrival*

James MacAllister
The University of Edinburgh, United Kingdom

SAŽETAK

U ovom radu tvrdim da tragični filmovi mogu imati značajan potencijal za etičku edukaciju kada potaknu publiku na suosjećanje s patnjom prikazanom na ekranu. Prvo sažimam dva pristupa odnosu između tragične umjetnosti, moralne edukacije i estetske vrijednosti (pristupe koje nude Rorty i Lamarque). Zatim raspravljam o problemima tih pristupa i objašnjavam kako novi kriterij estetske vrijednosti može pomoći u njihovom rješavanju. Potom iznosim argument da tragični filmovi mogu etički educirati publiku na način koji obogaćuje estetsku vrijednost filma u najmanje tri smjera: produbljivanjem moralnog razumijevanja, produbljivanjem razumijevanja ljudske prirode i etičke svrhe te produbljivanjem razumijevanja etičke teorije. Na kraju pokazujem kako film Denisa Villeneuvea, *Arrival*, prikazuje tragičnu priču s potencijalom za etičku edukaciju u svakom od navedenih aspekata.

Ključne riječi: prikaz patnje; tragični film; etička edukacija kroz film; edukativni eticizam; etika u Villeneuveovu filmu *Arrival*.

The Summit of Safe Horror: Defending Most Horror Films

Cara Rei Cummings-Coughlin
Morgan State University, USA

ABSTRACT

Many people regularly watch horror films. While it seems clear that sporadically watching horror films will not make us bad people, if it is the main type of media that we consume, then are we still safe? I will defend most horror films from Di Muzio (2006), who worries that we are harming our moral character by watching them. Most horror films (e.g., Candyman, Get Out, and Scream) fall into what I call the summit of safe horror (SoSH), the inverse of the uncanny valley effect, wherein almost-but-not-quite-human robots elicit discomfort from viewers rather than empathy. In the SoSH, violence elicits excitement rather than pity for the victims because the violence is mitigated by, among other things, comic relief and foolish choices by the characters. These narrative features allow most horror films to be intense enough to cause excitement and terror yet not so intense as to cause a negative moral attitude to form in our soul, because we feel what Aristotle would consider the appropriate amount of fear. Torture porn, a subset of horror films lacking plot and focusing solely on gore (e.g., the Saw sequels), falls outside of the SoSH because it lacks these narrative features, making the violence depicted too intense to be entertaining. These films outside the SoSH will not necessarily cause an inappropriate amount of fear but are simply the only ones that could possibly do so. Caution: spoilers ahead!

Keywords: comic relief; courage; fear; horror film; torture porn; uncanny valley.

Vrhunac sigurnog horora: obrana većine horor filmova

Cara Rei Cummings-Coughlin
Morgan State University, USA

SAŽETAK

Mnogi ljudi redovito gledaju horor filmove. Iako je očito da povremeno gledanje horor filmova neće negativno utjecati na naš moralni karakter, postavlja se pitanje: jesmo li i dalje “sigurni” ako su horor filmovi glavni oblik medija koji konzumiramo? U ovom radu branim većinu horor filmova od kritike Di Muzia (2006), koji tvrdi da njihovo gledanje šteti našem moralnom karakteru. Većina horor filmova (npr. Candyman, Get

Out i Scream) spada u ono što nazivam vrhuncem sigurnog horora (SoSH), obrnutim učinkom “doline nelagode” u kojoj gotovo, ali ne posve ljudski roboti izazivaju nelagodu umjesto empatije. U okviru SoSH-a, nasilje izaziva uzbuđenje, a ne sažaljenje prema žrtvama, jer je nasilje ublaženo, među ostalim, komičnim predahom i nesmotrenim odlukama likova. Ove narativne značajke omogućuju većini horor filmova da budu dovoljno intenzivni za izazivanje uzbuđenja i straha, ali ne toliko intenzivni da bi negativno utjecali na naš moralni stav, budući da osjećamo ono što bi Aristotel smatrao primjerenim stupnjem straha. Tzv. “torture porn” (npr. nastavci *Saw*), koji predstavlja podskup horor filmova bez zapleta koji se fokusiraju isključivo na krvoproliće, ne spada u SoSH jer mu nedostaju ove narativne značajke, zbog čega prikazano nasilje postaje previše intenzivno da bi bilo zabavno. Iako ovi filmovi izvan SoSH-a neće nužno izazvati neprikladan stupanj straha, oni su jedini koji bi to potencijalno mogli učiniti. Upozorenje: slijede spojleri!

Ključne riječi: komični predah; hrabrost; strah; horor film; mučenje kao zabava; dolina nelagode.

Cringe Overhang: The Perlocutionary Effects of Cringe Comedy

Alexander Sparrow
Falmouth University, United Kingdom

ABSTRACT

Cringe comedy can make people so uncomfortable, the cringe continues even after the comedy has stopped. This paper explains this effect, which I call “cringe overhang”. If audiences weakly connect to the characters, they laugh. If audiences strongly connect, they have a negative emotional response—say, struggling to watch, or wanting to leave the room.

Firstly, I argue cringe comedy jokes are illocutionary acts designed to provoke laughter through second-hand embarrassment (Austin 1975). Secondly, I acknowledge that these jokes don’t always produce the desired perlocutionary effect of laughter—sometimes the joke is unable to cut through the embarrassment, merely leaving the viewer in a state of discomfort. Thirdly, drawing on the benign violation theory of McGraw and Warren (2010), I explain that the surplus of embarrassment is due to maximising the violation in the comedy while adding comparatively little benign.

Finally, I argue that cringe comedy’s funniness is reliant on its lack of social psychological distancing. By leaving no room between the viewer and the character, embarrassment is maximised, the comedy is less

benign (i.e. a stronger violation) and more polarising as a result. This explains i) why cringe comedy produces a comedic “overhang” in some viewers, where they continue to cringe even after the comedy has stopped, and ii) why cringe comedy produces a laughter response in some audiences, and stress responses in others.

Keywords: speech acts; humour theory; benign violation theory; cringe comedy.

Krindž preostatak: perlokucijski učinci krindž komedije

Alexander Sparrow

Falmouth University, United Kingdom

SAŽETAK

Krindž komedija može učiniti ljude toliko nelagodnima da se osjećaj neugode nastavlja i nakon što komedija prestane. Ovaj rad objašnjava taj učinak, koji nazivam “krindž preostatak”. Ako se publika slabo poveže s likovima, smiju se. Ako se snažno povežu, imaju negativnu emocionalnu reakciju—na primjer, teško im je gledati ili žele napustiti prostoriju. Prvo, tvrdim da su šale u krindž komediji ilokucijski činovi osmišljeni da izazovu smijeh putem posredne neugode (Austin 1975). Drugo, priznajem da te šale ne proizvode uvijek željeni perlokucijski učinak smijeha—ponekad šala ne uspije prevladati neugodu, ostavljajući gledatelja samo u stanju nelagode. Treće, oslanjajući se na teoriju benigne povrede McGrawa i Warrena (2010), objašnjavam da višak neugode proizlazi iz maksimiziranja povrede u komediji uz relativno mali dodatak benignog. Na kraju, tvrdim da je zabavnost krindž komedije uvjetovana nedostatkom socijalno-psihološkog distanciranja. Budući da ne ostavlja prostor između gledatelja i lika, nelagoda se maksimizira, komedija postaje manje benigna (odnosno snažnija povreda) i polarizirajuća kao rezultat. Ovo objašnjava i) zašto krindž komedija kod nekih gledatelja proizvodi komičan “preostatak”, gdje se osjećaj neugode nastavlja i nakon što komedija završi, i ii) zašto krindž komedija kod nekih izaziva smijeh, a kod drugih stresne reakcije.

Ključne riječi: govorni činovi; teorija humora; teorija benigne povrede; krindž komedija.

What Counts as (Evidence of) Narrow Aesthetic Cognitivism

Mario Slugan

Queen Mary University of London, United Kingdom

ABSTRACT

In this paper I argue that the existing arguments for narrow aesthetic cognitivism are not valid. The reason is that the proponents of the view have mostly focused on theoretical debates rather than on empirical studies of the matter. I make my argument in five steps. First, I distinguish narrow from broad aesthetic cognitivism. Second, I specify the scope of the narrow claim and suggest that narrow aesthetic cognitivism is not about art in general but about narrative fictional works. Third, I point out that arguments in favor of the possibility of such works expressing truth do not discuss how truth is justified. Fourth, I concede that justification of truths in fictional and nonfictional works in the end rests on the experiences outside of the work but argue that it is an empirical matter whether belief has been acquired or not. Fifth, I demonstrate that the most recent analysis on belief-acquisition evinces that fiction is a not an effective vehicle for imparting beliefs. I conclude with a discussion of what would count as evidence in favor of aesthetic cognitivism.

Keywords: aesthetic cognitivism; empirical studies; knowledge; justification; fiction.

Što se računa kao (dokaz za) uski estetski kognitivizam

Mario Slugan

Queen Mary University of London, United Kingdom

SAŽETAK

U ovom radu tvrdim da postojeći argumenti za uski estetski kognitivizam nisu valjani. Razlog tome je što su se zagovornici ovog stajališta uglavnom usmjerili na teorijske rasprave, a ne na empirijske studije o toj temi. Svoj argument izlažem u pet koraka. Prvo, razlikujem uski od širokog estetskog kognitivizma. Drugo, preciziram opseg uske tvrdnje i sugeriram da uski estetski kognitivizam nije o umjetnosti općenito, već o narativnim fikcionalnim djelima. Treće, ističem da argumenti u prilog mogućnosti da takva djela izražavaju istinu ne raspravljaju o tome kako se istina opravdava. Četvrto, priznajem da opravdanje istina u

fikcionalnim i ne-fikcionalnim djelima na kraju počiva na iskustvima izvan samog djela, ali tvrdim da je empirijsko pitanje je li uvjerenje stečeno ili nije. Peto, pokazujem da najnovija analiza stjecanja vjerovanja dokazuje da fikcija nije učinkovit medij za prenošenje vjerovanja. Zaključujem raspravom o tome što bi se moglo smatrati dokazom u prilog estetskom kognitivizmu.

Ključne riječi: Estetski kognitivizam; empirijske studije; znanje; opravdanje; fikcija.

Revisiting McGee's Counterexample to Modus Ponens

Nicolás Lo Guercio
UBA/IIF-SADAF-CONICET, Argentina

Mariela Rubin
UBA/IIF-SADAF-CONICET, Argentina

ABSTRACT

In this paper, we provide a novel explanation of McGee's (1985) alleged counterexample to Modus Ponens for indicative conditionals. Our strategy is to show that pragmatic phenomena interfere with intuitions concerning the acceptability of the inference. More specifically, we argue that two confounding factors at play affect our intuitions concerning the acceptability of the inference, neither of which is related to validity. First, the indefinite determiner phrase "a Republican" is ambiguous, to wit, it may receive either a specific or a non-specific reading (and as we shall see, substituting a disjunction or a definite description for the indefinite is of no help). Second, the conclusion triggers an ignorance inference concerning its antecedent. In light of this, we shall argue, McGee's diagnosis must be rejected.

Keywords: modus ponens; McGee's counterexample; pragmatics; conditionals.

Ponovno razmatranje McGeejeva protuprimjera Modus Ponensu

Nicolás Lo Guercio
UBA/IIF-SADAF-CONICET, Argentina

Mariela Rubin
UBA/IIF-SADAF-CONICET, Argentina

SAŽETAK

U ovom radu nudimo novo objašnjenje navodnog protuprimjera Modus Ponensu za indikativne kondicionalne iskaze koji je iznio McGee (1985). Naša je strategija pokazati da pragmatički fenomeni utječu na intuicije vezane uz prihvatljivost zaključivanja. Konkretno, tvrdimo da dva zbunjujuća čimbenika utječu na naše intuicije o prihvatljivosti zaključivanja, a nijedan od njih nije povezan s valjanošću. Prvo, fraza s neodređenim člankom “neki republikanac” je dvoznačna; naime, može imati specifično ili nespecifično značenje (a, kako ćemo vidjeti, zamjena disjunkcijom ili određenim opisom ne pomaže). Drugo, zaključak aktivira inferenciju na temelju neznanja o svojem antecedensu. S obzirom na to, tvrdit ćemo da McGeejeva dijagnoza mora biti odbačena.

Ključne riječi: modus ponens; McGeejev protuprimjer; pragmatika; kondicionali.

Translated by Marko Jurjako (Rijeka) and Iva Martinić (Rijeka)

Proofread by Iva Martinić (Rijeka)

Papers in the online issue of EuJAP have been formatted by Iva Martinić and Marko Jurjako.

AUTHOR GUIDELINES

Publication ethics

EuJAP subscribes to the publication principles and ethical guidelines of the Committee on Publication Ethics (COPE).

Submitted manuscripts ought to:

- be unpublished, either completely or in their essential content, in English or other languages, and not under consideration for publication elsewhere;
- be approved by all co-Authors;
- contain citations and references to avoid plagiarism, self-plagiarism, and illegitimate duplication of texts, figures, etc. Moreover, Authors should obtain permission to use any third party images, figures and the like from the respective copyright holders. The pre-reviewing process includes screening for plagiarism and self-plagiarism by means of internet browsing and software Turnitin;
- Submissions should be made via EuJAP's Open Journal System at the following address: <https://hrcaak.srce.hr/ojs/index.php/eujap> (or, in the case of a special issue, they can be sent to the Guest Editors in a Word-compatible format);
- be prepared for blind refereeing: authors' names and their institutional affiliations should not appear on the manuscript. Moreover, "identifiers" in MS Word Properties should be removed;
- be accompanied by a separate file containing the title of the manuscript, a short abstract (not exceeding 300 words), keywords, academic affiliation and full address for correspondence including e-mail address, and, if needed, a disclosure of the Authors' potential conflict of interest that might affect the conclusions, interpretation, and evaluation of the relevant work under consideration;
- be in American or British English;
- be no longer than 9000 words, including references (for Original and Review Articles).

- be between 2000 and 5000 words, including footnotes and references (for Discussions and Critical notices)

We ask authors to submit only one manuscript at a time. A second submission by the same author is allowed only after a final decision has been made on their previously submitted manuscript.

Norms for publishing with AI

The Journal does not exclude the use of AI generated text. However, all authors (including reviewers and editors) take full responsibility for its factual accuracy and the proper acknowledgement of sources. In the acknowledgement section of your manuscript or the title page (depending on the submission/publication stage) or in other kind of reports you must identify the AI that was used, and the extent of the contribution. For instance, ChatGPT (version or the date when the AI was used).

The contribution level of the AI can be defined as follows:

- negligible – means the AI only made minor changes to the manuscript’s style or grammar (this includes using AI for copyediting and similar services);
- modest – means the AI made important suggestions but was not the primary driver of the research or had an essential role in writing the manuscript;
- substantial – means the AI made several crucial suggestions that shaped the research and the manuscript could not have been completed without it.

If the contribution of the AI is “negligible”, there is no requirement to mention its usage during the submission or review and publication processes. However, for any other level of contribution, it is expected that authors will report the extent of AI usage. In cases where the AI contribution is “substantial”, authors, reviewers, and editors should provide a comprehensive description of the AI usage and its contributions in a narrative format.

Initial submission

When first submitting a manuscript it is not required that the manuscript conforms to EuJAP’s style guidelines. Only after a manuscript has been accepted for publication we expect the authors to format the manuscript in accordance with EuJAP’s style guidelines.

Submitting revised manuscripts

When submitting a revised manuscript, please include also a separate document where it is explained how revisions were made in response to reviewers' comments.

Policy for submitted manuscripts

If the submitted manuscript is authored by more than one person, there should be a brief explanation in the title page of the contribution of each Author with respect to the conception and design of the argument, study, etc. and writing of the paper.

To preserve the anonymous status of the review process, we prefer (but do not require) that submitted versions of manuscripts are not deposited in open access article repositories.

Policy for accepted and published manuscripts

Accepted and published versions of the manuscript can be deposited in institutional or personal repositories without an embargo period. In case of published manuscripts, a link (with DOI) to the journal's web pages and/or HRCÁK should be added.

Malpractice statement

If the manuscript does not match the scope and aims of EuJAP, the Editors reserve the right to reject the manuscript without sending it out to external reviewers. Moreover, the Editors reserve the right to reject submissions that do not satisfy any of the previous conditions.

If, due to the authors' failure to inform the Editors, already published material will appear in EuJAP, the Editors will report the authors' unethical behaviour in the next issue and remove the publication from EuJAP web site and the repository HRCÁK.

In any case, the Editors and the publisher will not be held legally responsible should there be any claims for compensation following from copyright infringements by the authors.

For additional comments, please visit our web site and read our Publication ethics statement (<https://eujap.uniri.hr/publication-ethics/>). To get a sense of the review process and how the referee report ought to look like, the prospective Authors are directed to visit the *For Reviewers* page on our web site (<https://eujap.uniri.hr/instructions-for-reviewers/>).

Style

Accepted manuscripts should:

- follow the guidelines of the most recent Chicago Manual of Style
- contain footnotes and no endnotes
- contain references in accordance with the author-date Chicago style, here illustrated for the main common types of publications (T = in text citation, R = reference list entry)

Book

T: (Nozick 1981, 203)

R: Nozick, R. 1981. *Philosophical Explanations*. Cambridge: Harvard University Press.

Book with multiple authors

T: (Hirstein, Sifferd, and Fagan 2018, 100)

R: Hirstein, William, Katrina Sifferd, and Tyler Fagan. 2018. *Responsible Brains: Neuroscience, Law, and Human Culpability*. Cambridge, Massachusetts: The MIT Press.

Chapter or other part of a book

T: (Fumerton 2006, 77-9)

R: Fumerton, Richard. 2006. 'The Epistemic Role of Testimony: Internalist and Externalist Perspectives'. In *The Epistemology of Testimony*, edited by Jennifer Lackey and Ernest Sosa, 77–91. Oxford: Oxford University Press.

<https://doi.org/10.1093/acprof:oso/9780199276011.003.0004>.

Edited collections

T: (Lackey and Sosa 2006)

R: Lackey, Jennifer, and Ernest Sosa, eds. 2006. *The Epistemology of Testimony*. Oxford: Oxford University Press.

Article in a print journal

T: (Broome 1999, 414-9)

R: Broome, J. 1999. "Normative requirements." *Ratio* 12: 398-419.

Electronic books or journals

T: (Skorupski 2010)

R: Skorupski, John. 2010. "Sentimentalism: Its Scope and Limits." *Ethical Theory and Moral Practice* 13 (2): 125–36.
<https://doi.org/10.1007/s10677-009-9210-6>.

Article with multiple authors in a journal

T: (Churchland and Sejnowski 1990)

R: Churchland, Patricia S., and Terrence J. Sejnowski. 1990. "Neural Representation and Neural Computation." *Philosophical Perspectives* 4. <https://doi.org/10.2307/2214198>

T: (Dardashti, Thébault, and Eric Winsberg 2017)

R: "Dardashti, Radin, Karim P. Y. Thébault, and Eric Winsberg. 2017. Confirmation via Analogue Simulation: What Dumb Holes Could Tell Us about Gravity." *The British Journal for the Philosophy of Science* 68 (1): 55–89.

<https://doi.org/10.1093/bjps/axv010>

Website content

T: (Brandon 2008)

R: Brandon, R. 2008. Natural Selection. *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Accessed September 26, 2013.

<http://plato.stanford.edu/archives/fall2010/entries/natural-selection>

Forthcoming

For all types of publications followed should be the above guideline style with exception of placing 'forthcoming' instead of date of publication. For example, in case of a book:

T: (Recanati forthcoming)

R: Recanati, F. forthcoming. *Mental Files*. Oxford: Oxford University Press.

Unpublished material

T: (Gödel 1951)

R: Gödel, K. 1951. *Some basic theorems on the foundations of mathematics and their philosophical implications*. Unpublished manuscript, last modified August 3, 1951.

Final proofreading

Authors are responsible for correcting proofs.

Copyrights

The journal allows the author(s) to hold the copyright without restrictions. In the reprints, the original publication of the text in EuJAP must be acknowledged by mentioning the name of the journal, the year of the publication, the volume and the issue numbers and the article pages.

EuJAP subscribes to Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). Users can freely copy and redistribute the material in any medium or format, remix, transform, and build upon the material for any purpose. Users must give appropriate credit, provide a link to the license, and indicate if changes were made. Users may do so in any reasonable manner, but not in any way that suggests the licensor endorses them or their use. Nonetheless, users must distribute their contributions under the same license as the original.



Archiving rights

The papers published in EuJAP can be deposited and self-archived in the institutional and thematic repositories providing the link to the journal's web pages and HRČAK.

Subscriptions

A subscription comprises two issues. All prices include postage.

Annual subscription:

International:

individuals € 50

institutions € 100

Croatia:

individuals € 30

institutions € 60

Bank: Zagrebačka banka d.d. Zagreb

SWIFT: ZABHR 2X

IBAN: HR9123600001101536455

Only for subscribers from Croatia,

please add: “poziv na broj”: 0015-03368491

European Journal of Analytic Philosophy is published twice per year.

The articles published in the European Journal of Analytic Philosophy are indexed and abstracted in SCOPUS, SCImago, Web of Science (Emerging Sources), The Philosopher's Index, European Reference Index for the Humanities (ERIH PLUS), Dimensions, Directory of Open Access Journals (DOAJ), PhilPapers, and Portal of Scientific Journals of Croatia (HRČAK), ANVUR (Italy), Sherpa Romeo