



# The New Collection

Volume 15 | 2021  
ISSN 1757-2541

# The New Collection

2021

*The New Collection*  
New College, Holywell Street,  
Oxford, OX1 3BN  
Email: [new.collection@new.ox.ac.uk](mailto:new.collection@new.ox.ac.uk)  
Website: <http://mcr.new.ox.ac.uk/journal>

Editor-in-Chief: Sophie Nagler

MCR Editorial Committee: Matthew Courtney, Marco Grossi, Nico Han, Bradley Hoover, Andrés Ibáñez Núñez, Aeron Laffere, Joseph Milton, Jemma Paek, Olana Peters, and Joseph Pitaluga

SCR Review Committee: Prof Holly Bridge, Prof E. Victor Flynn, Dr Elizabeth Frazer, Prof Antony Galione, Dr Edward Howell, Dr Luke Lewis, Prof David Parrott, Dr Lavinia Picollo\*, Dr Tyrone Rees†, and Prof Richard Rutherford‡

Publisher: Sophie Nagler

Treasurer: Jordan Lian

Typesetter: Jamie Gravell

Cover Photo: Dr Michelle Castelletti

#### CONTACT DETAILS

All submissions, subscriptions and business enquiries should be addressed to the above contact details.

---

\*External Reviewer - Assistant Professor, National University of Singapore, and Lecturer of Philosophy, University College London.

†External Reviewer - Researcher, STFC Rutherford Appleton Laboratory.

‡External Reviewer - Philip Chappell Student and Tutor in Greek and Latin Literature, Christ Church, Oxford.





## NOTES ON CONTRIBUTORS

All contributors are members of the MCR of New College at the University of Oxford. Contact details are provided at the beginning of each article.

## GUIDELINES FOR SUBMISSIONS TO THE NEW COLLECTION

*The New Collection* is open to all disciplines, and review, research, and opinion articles are accepted for submission to the journal. All members of the MCR and those who have recently graduated are encouraged to submit.

Submissions should be sent to [new.collection@new.ox.ac.uk](mailto:new.collection@new.ox.ac.uk); no hard copy submissions are accepted. Please use whatever reference style is appropriate for your discipline. Authors will have the choice of whether or not they would like their submissions considered working papers whereby readers are asked not to cite the work without the author's permission. This is to help ensure that publication in *The New Collection* does not preclude publication elsewhere.

*The New Collection* uses a double peer-review process for peer-reviewed pieces; articles are first reviewed by the editorial team composed of members of the MCR, and subsequently by reviewers from the SCR. The aim of the review process is to ensure that all articles best demonstrate the broad spectrum of the research currently being undertaken at New College, are accessible to a general academic audience, and are academically rigorous.

## Contents

Editorial Welcome <i>Sophie Nagler</i>	iii
MCR President's Foreword <i>Nico Han</i>	v
Warden's Foreword <i>Miles Young</i>	vi
The Historical Approach of Sir John Elliott <i>Elliot Jordan</i>	1
The Discovery and Development of Risdiplam – a small molecule SMN2 splicing modulator for the treatment of Spinal Muscular Atrophy <i>Harry Woodcock</i>	18
From the Condemnation to the Revival of the Feminine: The Transmission and Reception of the Phaedra Myth in Tsvetaeva's Drama <i>Katherine New</i>	40
Critical Theory's Objectives: Unfulfilled in Praxis, Fulfilled in Critique <i>Makiko Miyazaki</i>	69
Is the Concept of Truth Part of Every Proposition? <i>Mariona E. Miyata-Sturm</i>	80
Reducing Networks Sizes <i>Paul Dubois</i>	96
Brain-Machine Interfaces and Neuroplasticity: Towards Accurate and Robust Two-Learner Systems	

*Peter Holderrieth* 113

PART II: *NEW PERSPECTIVES*

*COVID, what might the future hold? A conversation with Alain Townsend FRS,  
Professor of Molecular Immunology.  
Joe Milton, Prof. Alain Townsend* 133

Papers Published 2020–21  
*MCR Members* 139

## Editorial Welcome

Welcome to the fifteenth volume of *The New Collection*, the academic journal published annually by New College's Middle Common Room (MCR). It represents the academic breadth and quality of New College's graduate community through a cross-disciplinary selection of double-blind double peer-reviewed articles. This volume is entitled *New Perspectives*, and I hope that it can provide exactly that at a time of dynamic change in our college and beyond.

Our volume starts by analysing the approach of historian Sir John Elliott. From there, we jump into the medical realm by looking at novel treatment options for Spinal Muscular Atrophy using Risdiplam. Next, we investigate how gender is represented in donor traditions via a cross-disciplinary analysis of the reception of the ancient Phaedra myth in Tsvetaeva's drama. Jumping to politics, we get to enjoy some critical reflections on the objectives of critical theory in international relations scholarship. Our fifth piece gives us some philosophical insights on whether the concept of truth is part of every proposition. Moving on to mathematics, we investigate how to reduce the size of network structures before we finish of the peer reviewed section of this volume by going back to the (neuro-)sciences: in an interdisciplinary piece, we take a closer look at neuroplasticity and brain-machine interfaces.

Like icing on a good cake, we finish this volume with an interview with Alain Townsend FRS FRCP, Professor of Molecular Immunology, on *COVID, what might the future hold?* We, thus, close the circle to this volume's motto *New Perspectives*, hopefully heralding a new post-pandemic era, both for *The New Collection* and the entire New College community.

The pandemic has been a difficult time for many in the college community, including our editors, and I am proud that we are publishing *The New Collection* despite all odds. I would like to cordially thank the editorial team, in particular Matthew Courtney for his proofreading work, Jamie Gravell for the typesetting of this volume, and Joseph Milton for conducting the interview. A special thanks also goes to all contributors to this volume as well as the many engaged interviewers in New College's Middle and Senior Common Rooms, and beyond.



I hope that we can provide some *New Perspectives* and I am hopefully looking forward to a dynamic and fresh new start for *New College*, especially its Middle Common Room, as well as *The New Collection*.

On behalf of the editorial team,

Sophie Nagler  
Editor-in-Chief

## MCR President's Foreword

The academic year 2020-21 at Oxford will always be remembered as a special year. Scholars from around the world arrived at New College during the COVID-19 pandemic and most of us, including myself, spent extra effort coping with isolation and uncertainty, both personally and professionally. The scholars of New College, however, showed tremendous strength, commitment and passion even during this challenging time.

Amidst this, the Middle Common Room has provided a sense of security and comfort for our postgraduates, as well as a benchmark of excellence at the University of Oxford. It is in this context that the New Collection has given postgraduate students the opportunity to share ideas through carefully crafted, well-constructed essays. The essays address some of the key issues facing society and cover a wide variety of subjects ranging from critical theory in international relations to collaborative research on brain-machine interfaces and neuroplasticity.

I want to thank all those who have contributed to the creation of the New Collection – the authors of the MCR for producing and sharing such stimulating research papers; the editors who worked tirelessly to ensure the high standard of the papers; and the SCR members for providing their insights, expertise and time to the authors. I especially wish to thank the editor-in-chief, Sophie Nagler, for her dedication in seeing the New Collection come to fruition. Upon reading the articles in this issue, I felt extremely humbled and honoured to be part of this community, I hope that you, too, will enjoy reading this issue and rejoice at the achievement of our outstanding scholars and supportive peers at the MCR.

Nico Han  
MCR President

## Warden's Foreword

*The New Collection* once again challenges and delights.

It challenges because it contains academic contributions of the highest standards, which provoke an examination of topics as diverse as the historiography of Spain, the concept of truth, the meaning of critical theory, the treatment of spinal muscular atrophy, the use of the Phaedra myth within the dramatic writings of Marina Tsvetaeva, the optimum size of information networks and the exciting opportunities inherent in the interface of brains and machines.

And it delights because of the sheer range of the material. This is surely exactly what being a 'university' means.

This edition is a vivid testimony to the vitality and rigour of New College as a research institution where young academics are pushing the boundaries of knowledge. Thanks to the contributors, editors and reviewers who have brought it to fruition.

MILES YOUNG  
Warden of New College

# The Historical Approach of Sir John Elliott

*Elliot Jordan\**

*Faculty of History, University of Oxford*

Sir John Elliott's work has engaged with a broad swathe of historical topics and questions across the early modern world, and is immediately familiar to all students of history from first-year undergraduates to veteran fellows in a way only a select few modern historians are. This essay seeks to argue that Elliott's historiographical approach has been defined by two fundamental themes- political history, and 'transnational' history.

Elliott's use of political history, which throughout his work counterbalances the socioeconomic and long-term structuralist ideas used by mid-twentieth century Marxist and Annalist historians, has allowed him to maintain the importance of the agency of individual figures and intellectual or cultural trends in determining historical change, as will be examined below. This will be further evidenced with three key areas of Elliott's historiography: his early work on Catalonia demonstrates his scepticism to existing historiographical paradigms and his attention to fine political detail, while his engagement with the questions of Spanish Decline and the 'General Crisis' alike show his prioritisation of political history and the agency of key decision-makers over economic or structuralist trends. The second half of the essay will argue that Elliott's work has also been consistently defined by 'transnational' history, or 'history across borders': this is firstly the case in his influential theory of 'composite monarchy', examining how connections between centre and periphery defined Spanish power. Elliott's 'transnational' approach is also visible in his growing interest in Atlantic history. Finally, Elliott's work has impressively demonstrated the value of comparative history, with Elliott's international comparisons of people and polities allowing him to provide more balanced views of his subjects and integrate them with wider historiographical and intellectual questions. In this blend of historiograph-

---

\* [elliott.jordan@new.ox.ac.uk](mailto:elliott.jordan@new.ox.ac.uk)



ical approaches, which as this essay seeks to demonstrate is present throughout his publications, Elliott's work thus represents a powerful example of 'total history'.

Firstly, the most defining feature of Elliott's approach to historiography is his constant emphasis on the importance of political history in understanding the past. In this regard, he was moving against the main historiographical trend of his formative period as a historian, i.e., at Cambridge in the 1950s: prioritisation of long-term social and economic history, and accompanying minimisation of human or political agency. In the case of British universities at the time of his early career, 'Marxist and marxisant history...was sweeping everything before it'- indeed, Himmelfarb highlights that it is itself surprising that a nation resistant to Marx's politics had 'so many Marxist historians', whose fixation on historical socioeconomic processes left little room for the history of political thought or action.<sup>1</sup> The great contemporary European historiographical movement, the *Annales*, similarly emphasised the *Longue durée* over political 'crests of foam' on the 'tides of history'.<sup>2</sup> Indeed, the *Annales* context is especially important in examining Elliott's historiography as all three 'masterpieces' of the *Annales* are early modern-focused, and two centre on Spanish imperial theatres: Braudel's Mediterranean and Febvre's Franche-Comté.<sup>3</sup> The latter is cited by Elliott as using similar themes of centre and periphery that he would later use in his work on 'composite monarchy' and his recommendation of Febvre is credited by Parker as inspiring the latter's own ground-breaking study of the 'Spanish Road'.<sup>4</sup> Accordingly, Elliott has drawn heavily on economic and social history throughout his career, remaining consistently alert to great socioeconomic processes in the imperial trajectory of Spain.<sup>5</sup> More broadly, he praises the influence of 'Marxist and marxisant' historians in maintaining a 'big picture' and laments its erosion by the 'nit-picking revisionism' of subsequent generations of historians.<sup>6</sup> Meanwhile, Braudel permeates Elliott's work, and Elliott specifically describes his own study of Catalonia as 'doing...what Febvre had done for the sixteenth-century Franche-Comté'.<sup>7</sup>

<sup>1</sup>John H. Elliott, *History in the Making* (New Haven, 2012), p.9; Gertrude Himmelfarb, *New History and the Old* (Cambridge, 2004), p.88.

<sup>2</sup>Fernand Braudel, *Mediterranean*, trans. Sian Reynolds (Berkeley, 1995), p.21.

<sup>3</sup>George Huppert, 'The *Annales*', in Michael Bentley (ed.), *Companion to Historiography* (London, 1997), p.874

<sup>4</sup>Elliott, *History in the Making*, p.54; Geoffrey Parker, 'Reflections', in Tonio Andrade and William Reger (eds.), *The Limits of Empire* (Abingdon, 2016), p.318.

<sup>5</sup>e.g. John H. Elliott, *Imperial Spain* (Harmondsworth, 1963), Ch.8, pt.I.

<sup>6</sup>Professor Sir John Elliott: Interview, 07/03/2008, <https://archives.history.ac.uk/makinghistory/resources/interviews/Elliott-John.html>.

<sup>7</sup>*Solemne Investidura*, Universitat de Barcelona (Barcelona, 1994), p.25.

Yet despite the power such socioeconomic and structuralist themes lend to his work, Elliott sees himself as 'strongly resisting a determinist approach' and his recovery of political history, though present throughout his career, is especially visible in his original research interest, the ministry of Olivares.<sup>8</sup> Though he sees Olivares as someone who 'mirrored with almost uncanny accuracy...the Castile of his times', and criticises Maranon's previously definitive biography as insufficiently aware of the 'social realities of the age', Elliott argues that only by including the neglected study of decision-making and the agency of key decision-makers may one 'achieve a genuinely "total history"'.<sup>9</sup> He thus consistently balances socioeconomic trends with active political processes, demonstrating that early modern states, far from showing the futility of studying *l'histoire événementielle*, 'showed how political skill and the deployment of resources' could determine success or failure.<sup>10</sup> In this regard, Elliott answers the criticism of Braudel raised by Green and Troup- that his work, and to an even greater degree that of subsequent *Annales* historians, is 'curiously devoid of people'.<sup>11</sup> Moreover, Elliott's recovery of human agency is also visible in how he counterbalances structuralist analysis with constant attention to cultural and intellectual contexts to illuminate the mentalities of the period. This is the case on the grand scale, with Elliott explaining Spain's seventeenth-century upsurge in military engagement not with reference to increased transatlantic treasure flow, as Chaunu did, but in the intellectual climate of an urgent search for martial valour to combat cultural decline, and he also applies it on smaller scales, exploring the engagement of Quevedo and Cortes with contemporary intellectual and political discourse.<sup>12</sup> Moreover, though Elliott has shown an enduring interdisciplinary interest, something identified by Huppert as a key *Annales* strategy, his has been not with geography or sociology but art history, using visual court culture and artistic propaganda to illuminate the cultural background of early modern Spanish political mentalities.<sup>13</sup> Finally, in contrast to the 'rejection of narrative' seen by Munz as a pillar of Braudel's work, Elliott's monographs convincingly demonstrate that the study of

<sup>8</sup>John H. Elliott, *History in the Making* (New Haven, 2012), p.9.

<sup>9</sup>John H. Elliott, 'Self-Perception and Decline', in *Spain and its World* (New Haven, 1989), p.259; John H. Elliott, *The Count-Duke of Olivares* (New Haven, 1986), p.x; *Ibid.*, p.xii

<sup>10</sup>John H. Elliott, *Europe Divided 1559-1598, 2<sup>nd</sup> ed.* (Oxford, 2000), p.xi.

<sup>11</sup>Anna Green and Kathleen Troup, *The Houses of History* (Manchester, 1999), p.90

<sup>12</sup>John H. Elliott, 'Foreign Policy and Domestic Crisis', in *Spain and its World*, pp.118-120.; John H. Elliott, 'Quevedo and the Count-Duke of Olivares', in *Ibid.*, pp.189-209.; John H. Elliott, 'The Mental World of Hernan Cortes', in *Ibid.*, pp.27-41.

<sup>13</sup>Huppert, '*Annales*', in Bentley, Companion, p.856.; e.g. John H. Elliott and Jonathan Brown, *A Palace for a King* (New Haven, 2003)

historical figures and events ‘demands a chronological narrative’.<sup>14</sup> Thus, Elliott sees his own 1958 invitation to the editorial board of *Past and Present* as ‘to free it of its overtly Marxist associations’, and his work fulfils this aim ideologically and methodologically.<sup>15</sup> But we may classify Elliott’s historiographical method not as *opposed* to the structural analysis of Marxists or the Annalists, but as *blending* their approaches with the older emphasis on political decision-making and cultural and intellectual life. Thus, in *Imperial Spain*, Elliott’s presentation of Castile’s economy rests as much on his analysis of Castilian cultural attitudes to wealth as his use of economic data.<sup>16</sup> His recovery of the importance of studying so-called ‘Great Men’ may not seem radical compared to other historiographical turns of the later twentieth century, but Elliott’s return to political history must nonetheless be considered one of the ‘challenges to historical norms that exploded in the 1960s’.<sup>17</sup> This has influenced subsequent historiography- he himself highlights that the development of independent Spanish historiography in the second half of the twentieth century has been able to use Anglo-American influences such as his own to add much needed political content to what would otherwise be a *marxisant* dominated approach, while among his own students, Parker’s assessment of Philip II draws on Elliott in judging that it was the king’s characteristics as a leader, rather than the structural problems of his dynastic inheritance, which limited Spanish power.<sup>18</sup>

Moreover, an early example of Elliott’s balanced but critical relationship with historiographical paradigms, and close attention to political detail, is in his work on the 1640 Revolt of the Catalans. Catalonia has featured prominently in his career, being the subject of both his first published historical monograph and his most recent. Early modern Catalonia, especially in its relationship with Madrid, is a topic of rich historiographical controversy: Catalan nationalist historians have consistently seen their past after a Medieval ‘Golden Age’ as one of rightful nationhood and ancient constitutional liberty suppressed by Castilian overlordship.<sup>19</sup> On the 1640 rebellion, Catalan historians have even claimed that Olivares deliberately provoked an uprising to justify ‘finishing once and for all with [Catalonia’s] liberties’.<sup>20</sup> However, by Elliott’s early career this was increasingly counterbalanced by Vicens Vives, who sought

<sup>14</sup>Peter Munz, ‘The Historical Narrative’, in Bentley, *Companion*, p.853.; Elliott, *Count-Duke of Olivares*, p.xii.

<sup>15</sup>Elliott, *History in the Making*, p.63.

<sup>16</sup>Elliott, *Imperial Spain*, pp.311-313.

<sup>17</sup>Jeremy D. Popkin, *From Herodotus to H-Net* (Oxford, 2015), p.129.

<sup>18</sup>Interview ‘An Englishman Abroad’, *Itineratio*, p.17.; Geoffrey Parker, *Imprudent King* (New Haven, 2014), pp.336-375.

<sup>19</sup>Yolaine Cultaiux, ‘Shaping of a Nation’, <https://www.lama.univ-savoie.fr/pagesmembres/vuillon/DIECKH1.pdf>, p.3.

<sup>20</sup>Victor Balguer, *Historia de Cataluna*, in John H. Elliott, *The Revolt of the Catalans* (Cambridge, 1963), p.ix.

to present a more sober, impartial view avoiding nationalist mythology.<sup>21</sup> Elliott is far from unsympathetic to Catalonia, recognising the importance of *patria* and nationalism as a unifying force for Catalans.<sup>22</sup> He highlights that Castile failed to use empire as a tool of integration, and describes 1626 as the year metropolitan attitudes to Catalonia switched from ‘years of neglect’ to ‘years of exploitation’.<sup>23</sup> Moreover, with his emphasis on the role of Olivares’ fiscal-military innovations in provoking rebellion, Elliott has agreed with the prevailing historical view of the revolt identified by Mitchell as the ‘political limits of “Castilianisation”’.<sup>24</sup> This has in turn fed into his influential theory of ‘composite monarchies’, as discussed below. Yet Elliott has enthusiastically supported the ‘demythifying’ aims of *Vicens Vives*: he argues that in contrast to romantic claims of a ‘free’ Catalonia and ‘enslaved’ Castile, the former may have been protected from political repression but existed in a ‘reality of social oppression’ by its own aristocracy.<sup>25</sup> Accordingly, Trevor Roper praises Elliott’s analysis of the revolt as an unprecedented placement of the subject ‘in its context...the structure and development of Spanish politics’.<sup>26</sup> Accordingly, Trevor Roper praises Elliott’s analysis of the revolt as an unprecedented placement of the subject ‘in its context...the structure and development of Spanish politics’.<sup>27</sup> In his engagement with Catalonia, Elliott thus shows the same historiographical scepticism as he does in his attitude to the Marxist and *Annaliste* schools, and demonstrates from his early career the fine attention to detailed political history that is visible throughout his work.

Meanwhile, Elliott’s recovery of the importance of political history is visible in his engagement with one of the most important questions of his field- that of the ‘Decline of Spain’. Spain’s seventeenth-century loss of vitality was something of which contemporaries were acutely aware, with the figure of the moralising ‘*arbitrista*’ already a satirical target by 1613.<sup>28</sup> Spanish Bourbon-era historians continued to return to *arbitrista* theories to explain decline while simultaneously, foreign Enlightenment thinkers, drawing on the ‘Black Legend’ and Protestant morality, criticised Spain’s intellectual and cultural backwardness.<sup>29</sup> However, over the twentieth century the question became one of economics- Hamilton explains it as the result of the Spanish failure to develop a robustly productive economy due to reliance on

<sup>21</sup>Elliott, *History in the Making*, pp.42-3.

<sup>22</sup>John H. Elliott, ‘King and *Patria* in the Hispanic world’, in *Spain, Europe and the Wider World*, p.185

<sup>23</sup>Elliott, *Revolt of the Catalans*, pp.1-48.; *Ibid.*, p.214

<sup>24</sup>Andrew Mitchell, ‘The Sacral Limits of Empire’, in Andrade and Reger, *Limits of Empire*, p.13

<sup>25</sup>Elliott, *Revolt of the Catalans*, p.16.

<sup>26</sup>John H. Elliott, ‘A Provincial Aristocracy’, in *Spain and its World* p.74.

<sup>27</sup>Hugh Trevor-Roper, *Review: The Revolt of the Catalans by J. H. Elliott*, *The Historical Journal*, Vol.7, No.1 (1964), p.171. Jean Vilar, *Literatura y Economía* (M

<sup>28</sup>Jean Vilar, *Literatura y Economía* (Madrid, 1973), p.13.

<sup>29</sup>Helen Rawlings, *Debate on the Decline of Spain* (Manchester, 2012), pp.7-9



transatlantic treasure flow, while the *Annales* 'long-run' socioeconomic analysis was applied to the decline of Spain by Vicens Vives' 'Barcelona school'.<sup>30</sup> . Initially, Elliott's attitude to Spanish decline appears to support the socioeconomic arguments- instead of seeing church and bureaucratic growth as causing economic backwardness, he reverses this, viewing the flocking of Spaniards to safe employment as priests or bureaucrats as symptomatic of deeper economic weakness, and references Spain's collapsing 'population...productivity and...overseas wealth'.<sup>31</sup> However, Elliott does more than repeat economic arguments- he criticises the 'overwhelmingly economic interpretation of Spain's decline', and revives the importance of cultural, intellectual and political factors: he argues that far from only being a symptom of decline, the *arbitristas* and their combined demands for reform and a return to martial vigour received 'legislative embodiment' in the government of Olivares, whose ensuing attempts at fiscal-military homogenisation in the Peninsula and strategic recovery overseas precipitated the disasters of 1640.<sup>32</sup> In this regard, Elliott's emphasis on the role of foreign policy in causing decline supports Kennedy's argument that Habsburg Spain represents an example of 'imperial overstretch'.<sup>33</sup> Thus, Elliott combines socioeconomic with political interpretations to view decline 'not simply as a problem in economic history, but as one in which domestic and international politics...cultural attitudes and the structure of society are all closely interwoven'.<sup>34</sup>

Furthermore, Elliott has also demonstrated the importance of attention to political history and his movement away from Marxist historiographical paradigms in his contribution to the debate over the 'General Crisis'. Encompassing an even wider geographical and chronological focus than that of Spanish decline, the question of whether the seventeenth century saw a 'General Crisis' has been an enduring one. Parker notes that Voltaire saw the period as one of 'usurpations almost from one end of the world to the other', and early twentieth century historians drew attention to the frequency of seventeenth century revolts- Mousnier summarised the period as 'mere trouble, agitation, chaos'.<sup>35</sup> Most influentially, Merriman highlighted that 'the two middle decades of the seventeenth century, the period when the Puritan Revolution is at its height, witness five other anti-monarchical rebellions'.<sup>36</sup> However, the modern debate was initiated and defined by Hobsbawm's contribution, which saw

<sup>30</sup>Rawlings, *Decline of Spain*, p.7

<sup>31</sup>John H. Elliott, 'Decline of Spain', in *Spain and its World* p.231.; *Ibid.*, p.222.

<sup>32</sup>*Ibid.*, p.220; Elliott, 'Self-Perception and Decline', p.259.

<sup>33</sup>Paul Kennedy, *The Rise and Fall of the Great Powers* (New York, 1987), p.48.

<sup>34</sup>Elliott, *Count-Duke of Olivares*, p.xii.

<sup>35</sup>Geoffrey Parker, *Global Crisis* (New Haven, 2013), p.1.; Ronald Mousnier, in Jonathan Dewald, 'Crisis, Chronology and the Shape of European Social History', *The American Historical Review*, Vol.113, No.4 (October, 2008), p.1032.

<sup>36</sup>Roger Merriman, *Six Contemporaneous Revolutions* (Oxford, 1938), p.viii.

the crisis in Marxist terms of class conflict and 'general transition from a feudal to a capitalist economy'. Though there was a degree of scepticism, Elliott stresses that most of the debate took place within these Marxist parameters.<sup>37</sup> Although Ogilvie sees supporters of the 'Crisis' as bifurcating into economic and political strands, with those in the former being divided among themselves, Elliott highlights that Mousnier, Trevor-Roper and Hobsbawm, 'three historians...of very different views and persuasions...have united in depicting the seventeenth century in such dramatic terms'.<sup>38</sup> Elliott's own relationship with the debate has two themes: he firstly expresses a degree of scepticism as to the uniqueness of the seventeenth century, by highlighting that despite the mid-sixteenth century seeing a similar rash of upheavals, no historian 'has grouped them together under the title "the general revolution of the 1560s", or has used them as evidence for a "general crisis of the sixteenth century"'.<sup>39</sup> More importantly, however, Elliott has also sought to move the debate outside Marxist or 'revolutionary' parameters by arguing that 'revolution' is itself an anachronistic concept for a period predating the eighteenth century, and that early modern political change occurred not due to lower-class violence, but 'inside the resilient framework of the aristocratic-monarchical state', and thus 'from within the political nation'. Moreover, Elliott argues that attempts to alter the *status quo* in the 'General Crisis' actually came from 'the ambitions of the state' against which revolts were borne out of 'determination of the dominant social groups to preserve their [constitutional-political] heritage'.<sup>40</sup> Accordingly, he summarises the Catalan rebellion as a "medieval" revolt against the attempted fiscal innovations of a centralising monarchy'.<sup>41</sup> Elliott's view of the 'General Crisis' thus supports that advanced by Steensgaard, who similarly argues that 'it was the governments that acted in a revolutionary manner...the six contemporaneous revolutions can only be seen as one if we rechristen them "the six contemporaneous reactions"'.<sup>42</sup> In his contribution to the General Crisis as in all of his other work, Elliott's historiographical approach is defined by constant attention to political, cultural and intellectual contexts, and an unwillingness to follow structuralist assumptions.

Moving from examining Elliott's use of political history, the second great theme of his historiographical approach is that of 'transnational' history. Though the term

<sup>37</sup>John H. Elliott, 'General Crisis in Retrospect', in *Spain, Europe and the Wider World*, p.55.

<sup>38</sup>Sheillagh Ogilvie, 'Germany and the Seventeenth-Century Crisis', in Geoffrey Parker and Lesley M. Smith (eds.), *The General Crisis of the Seventeenth Century*, 2nd (London, 1997), pp.57-9.; John H. Elliott, 'Revolution and Continuity', in *Spain and its World*, pp.92-3.

<sup>39</sup>*Ibid.*, p.95

<sup>40</sup>*Ibid.*, p.112.

<sup>41</sup>Elliott, *Revolt of the Catalans*, p.549.

<sup>42</sup>Niels Steensgaard, 'The Seventeenth-Century Crisis', in Parker and Smith, *General Crisis*, p.47.

is itself contested, all ideas of 'global', 'world' and 'transnational' history have in common the need to cross borders and move away from national history, and this is firstly visible in Elliott's theory of 'composite monarchy'. Influentially outlined in his 1992 article but present from the beginning of his work, this idea views the early modern world not in terms of simple nation-states but of polycentric powers in which each component owed allegiance to the monarch on its own unique terms. Thompson identifies continuity in Elliott's 'abiding interest in the themes of patria, national sentiment and community...apparent throughout his long catalogue', and we can accordingly view this theory as both showing Elliott's growing 'transnational' interests and as another manifestation of his continual interest in political culture.<sup>43</sup> Though Elliott credits the term 'composite monarchy' to Koenigsberger, he has convincingly argued that it is an ideal paradigm for understanding the state in the early modern world: 'its history needs to be assessed from this standpoint rather than from the society of unitary nation states that it was later to become'.<sup>44</sup> This has been especially well evidenced by his use of the concept to understand the inherently 'transnational' *Monarquía Hispanica*, as is visible in his original work on the Catalan Revolt, which he highlights as being triggered by the attempt by Olivares to change the terms under which the Catalan polity was integrated into the overall Spanish imperial patchwork. Meanwhile Parker and Kagan highlight the same 'dialogue...between centre and periphery' in Elliott's work on the General Crisis and Spanish America.<sup>45</sup> Elliott also argues convincingly for the validity of the concept in understanding the evolution of the Austrian and British monarchies, among others, and in doing so has influentially contributed to early modern historiography.<sup>46</sup> For example, Russell credited Elliott's early use of the concept for inspiring his conception of the breakdown of Stuart monarchy, Connolly summarises Ireland under Charles I as an example of 'the crisis of composite monarchy', and Elliott's own students have applied the same paradigms of centre-periphery relations in understanding early modern Granada and Valencia.<sup>47</sup> The composite monarchy idea has also proved popular in the historiography of empire, such as the works of Burbank and Cooper or Kumar, while Kennedy's earlier identification of Habsburg power as 'a congeries of territories...the only real

---

<sup>43</sup>I. A. A. Thompson, 'Castile, Spain and the Monarchy', in Richard L. Kagan and Geoffrey Parker (eds.), *Spain, Europe and the Atlantic World* (Cambridge, 1995), p.126.

<sup>44</sup>John H. Elliott, 'A Europe of Composite Monarchies', in *Spain, Europe and the Wider World* (New Haven, 2009), p.6.

<sup>45</sup>Kagan and Parker (eds.), *Spain, Europe and the Atlantic World*, frontmatter

<sup>46</sup>Elliott, 'Europe of Composite Monarchies', p.20.

<sup>47</sup>'Conrad Sebastian Robert Russell 1937-2004', in *Proceedings of the British Academy*, 138 p.340; S. J. Connolly, *Divided Kingdom: Ireland 1630-1800* (Oxford, 2008), pp.4-59; Kagan and Parker, *Spain, Europe and the Atlantic World*, Part II.

connecting link was the monarch himself' uses the same principal.<sup>48</sup> Even Cardim, Herzog, Ibanez and Sabatini, though emphasising the need to formulate new conceptions of Iberian colonial monarchy, acknowledge their 'debt' to Elliott's concept.<sup>49</sup> Thus, the first and possibly the most influential manifestation of Elliott's enduring use of transnational history is his use and popularisation of 'composite monarchy'.

Moreover, Elliott's use of transnational history is also visible in his emphasis on the importance of Atlantic and global history in studying early modern Spain. Elliott's study of the early modern Spanish world took place in the historiographical context of Chaunu's work, seen by Green and Troup as an effort to 'do for the Atlantic Ocean what Braudel had achieved for the Mediterranean sea', and Elliott identifies that Atlantic history grew in importance over the second half of the twentieth century in the context of a cold war 'Atlantic Alliance' and the notion that globalisation rendered national history insufficient.<sup>50</sup> Accordingly, Elliott has engaged with the early modern Atlantic as 'the perfect arena for the writing of...interconnected histories' across national borders and in acute understanding that though Hamilton and Chaunu had influentially studied the economic implications of early modern Atlantic expansion, the intellectual dimensions of it remained to be explored.<sup>51</sup> Elliott has thus investigated the political mentalities of the Atlantic, whether in his examination of the 'illusion and disillusionment' of Spanish governmental and intellectual thinkers with their conquests, or with his comparative study of Spanish and English attitudes to America.<sup>52</sup> In his arguments for the similarities between the Atlantic empires Elliott disagrees with historians such as Syme and Lang who emphasised the differences between Britain and Spain in the Atlantic- and where there were differences, Elliott highlights that this was often less a structuralist question and more one of human agency, showing the overlap between his political and transnational interests: 'behind the cultural values and the economic and social imperatives...lay a host of personal choices'.<sup>53</sup> Elliott has enthusiastically argued for the importance of Atlantic history, and his example of studying the oceanic connections underpinning Spain's empire has been followed by his students, such as Bakewell.<sup>54</sup> Moreover, Elliott's work has formed an important part of wider Atlantic historiography- *Empires*

<sup>48</sup>Jane Burbank and Frederick Cooper, *Empires in World History* (Princeton, 2010), p.465; Krishnan Kumar, *Visions of Empire* (Princeton, 2017), p.24; Kennedy, *Rise and Fall of the Great Powers*, p.52.

<sup>49</sup>Pedro Cardim et al., *Polycentric Monarchies* (Brighton, 2014), p.218.

<sup>50</sup>Green and Troup, *Houses of History*, p.91.; Elliott, *History in the Making*, p.200.

<sup>51</sup>*Ibid.*, p.204.

<sup>52</sup>John H. Elliott, 'Illusion and Disillusionment: Spain and the Indies', and 'Britain and Spain in America', in Spain, *Europe and the Wider World*, pp.131-172

<sup>53</sup>John H. Elliott, *Empires of the Atlantic World* (New Haven, 2006), pp.xiii-xx; *Ibid.*, p.411.

<sup>54</sup>e.g. John H. Elliott, 'Inaugural Lecture, 10 May 1991' (Oxford, 1991); Peter Bakewell, *Silver Mining and Society in Colonial Mexico* (Cambridge, 1971)



of the *Atlantic World* itself following the same techniques and concepts as his original Atlantic History lectures five decades earlier, is praised by Kagan as a crucial contribution to the field.<sup>55</sup> Thus, Parker argues that any '*scuola di Elliott*' would be centred on Atlantic history.<sup>56</sup> Beyond this, Elliott also constantly emphasises the importance of global connections and the importance of placing early modern Spain in an international context- hence, his inclusion of foreign policy in giving a detailed explanation of Spanish decline. Thus, Elliott's use of global historical perspectives through his enthusiastic support for Atlantic history and use of the international context in Spanish history forms another pillar of his transnational approach to historiography.

Finally, a running theme throughout Elliott's work is his use of comparative history, which he emphasises as, despite having a historiographical pedigree running from Bloch to Herodotus, something neglected by modern historians. Indeed, Elliott sees comparative history as an inherently transnational technique that like 'the tracing of connections can help break down the exceptionalism to which the writing of national history is all too liable to succumb' and emphasises this as especially important for the Hispanist, who 'must be capable of tracing the links between the peninsula and events in the wider world...to draw parallels and comparisons'.<sup>57</sup> In his evaluation of Olivares Elliott emphasises that many of the problems facing the Count-Duke were present elsewhere in Europe, a principle which he extends to the question of Spanish decline. Elliott also anchors his writing on Olivares in the context of early modern absolutism, highlighting that the propaganda used to portray Spanish power in the period can be used to understand the 'nature of contemporary governments and the difficulties they faced', and also compares early modern court cultures with references to their cultural and artistic dimensions.<sup>58</sup> In comparing Olivares' regime to Richelieu's administration in France, Elliott both allows for a more nuanced view of both men, and a more nuanced view of the assumption that France naturally replaced Spain as Europe's premier power: 'if Richelieu achieved his triumph by a hair's breadth, the margin by which Olivares was defeated was correspondingly close'.<sup>59</sup> Here, and elsewhere, Elliott has compared the phenomenon of 'Minister-Favourites' across Europe to provide a sophisticated response to Bérenger's

<sup>55</sup>Richard L. Kagan, 'People and Places in the Americas', in Nicholas Canny and Philip Morgan (eds.), *Oxford Handbook of the Atlantic World: 1450-1850* (Oxford, 2011), p.342

<sup>56</sup>Geoffrey Parker and Richard L. Kagan, 'Sir John Elliott: An Appreciation', in *Spain, Europe and the Atlantic World* p.13.

<sup>57</sup>Elliott, *Spain, Europe and the Wider World*, p.xvii.; John H. Elliott, *Discurso de investidura de doctor honoris causa*, Universidad Autónoma de Madrid (1984), p.24.

<sup>58</sup>John H. Elliott, 'Power and Propaganda', in *Spain and its World*, p.164; e.g. John H. Elliott, 'La Sociedad Cortesana en la Europa del Siglo XVII', in Jonathan Brown and Christopher Brown (eds.), *Velazquez, Rubens y Van Dyck* (Madrid, 1999), pp.15-32.

<sup>59</sup>John H. Elliott, *Richelieu and Olivares* (Cambridge, 1984), p.165.

earlier attempt to explain 'the problem of the *valido*'.<sup>60</sup> Elliott thus uses comparative perspectives in order to relate his historical foci to broader historiographical questions, such as in his comparison of British and Spanish empire building, while his juxtaposition of the Scottish and Catalan independence struggles provides a complex case study for the intellectual historiography of liberalism and nationalism.<sup>61</sup> Given the success with which he has used it, comparative history must surely also be part of any hypothetical 'school of Elliott', with Gil describing it as 'one of his clearest and most sustained endeavours'.<sup>62</sup> Perhaps because of the sheer scale of his work- an impressive thesis monograph and a timeless textbook within an astonishingly short time of completing his doctorate- or because of his broad and skilfully applied historical interests, Elliott has managed to avoid sustained criticism even from main historiographical rivals, and for some time has occupied a position as a grand maître comparable only to Braudel. One possible criticism one could level at Elliott is to question just how far the 'human' dimensions of his political historiography focus only on the agency of a tiny minority at the top of socioeconomic and political power. Elliott's decades of activity have seen many attempts to broaden the section of society given a voice in historical writing, such as the attempts by microhistory to preserve the agency of 'ordinary people' or of certain Marxists at 'rescuing' the working man from obscurity.<sup>63</sup> Recently there has been a vast expansion in the study of previously marginalised groups hitherto under-represented in historiography, one manifestation of which- studying the human victims of transatlantic slavery- has been praised by Elliott as bringing new energy to early modern study.<sup>64</sup> On one hand, it could be argued that Elliott's study of political history simply covers early modern 'great men'. However, the simple truth is that in the geopolitics and grand process of the sixteenth and seventeenth centuries which form the hard core of Elliott's historiographical interests, a tiny minority of ruling figures decided the fates of millions. As one contemporary playwright remarked, 'the same reason that makes a vicar go to law for a tithe-pig and undo his neighbours, makes them spoil a whole province and batter down goodly cities with the cannon'- and so it has necessarily been to their words and their thoughts that Elliott has consistently turned.<sup>65</sup>

<sup>60</sup>Jean Bérenger, 'Le probleme du ministériat au XVIIe siècle', in *Annales. Economies, Sociétés, Civilisations*. 29e année, N,1, 1974, p.166.

<sup>61</sup>John H. Elliott, *Scots and Catalans* (New Haven, 2018), pp.1-2.

<sup>62</sup>Xavier Gil, 'John Elliott en sus articulos', in Roberto Fernandez et al. (eds.), *John Elliott, El Oficio de Historador* (Lleida, 2001), p.152.

<sup>63</sup>Brad S. Gregory, 'Is Small Beautiful? Microhistory and the History of Everyday Life', in *History and Theory*, Feb.1999, Vol.38, No.1, pp.100-110.

<sup>64</sup>e.g. Joan Scott, 'Gender: A Useful Category of Historical Analysis', in *The American Historical Review*, Dec. 1986, Vol.91, No.5, pp. 1053-1075.; Elliott, *Empires of the Atlantic World*, p.xv.

<sup>65</sup>John Webster, *The Duchess of Malfi*, ed.Karen Britland (London, 2021), II.1, line 105.

In conclusion, this essay has argued that Elliott has shown a consistent historiographical approach, and it has been defined by two themes: an emphasis on the continued importance of political history and the study of key individuals for understanding the early modern world, and his use of 'transnational' approaches. Though Elliott's work demonstrates the value of the grand socioeconomic trends as championed by his Marxist and Annalist contemporaries, he consistently counterbalances them with close attention to the key political decision-makers of the period, and their cultural and intellectual worlds. It has been argued that this unwillingness to blindly follow historiographical trend and attention to political detail alike is exemplified by Elliott's work on early modern Catalonia, and his engagement with the 'Decline of Spain' has demonstrated the extent to which he is able to blend *Longue durée* and *événementielle* history. His engagement with the General Crisis has similarly shown scepticism to prevailing narratives, with his inverted view of the period as one of 'reaction' rather than 'revolution' possible only due to his continual use of political as well as socioeconomic analysis. Indeed, his study under the famously sceptical Butterfield probably influenced his willingness to challenge prevailing narratives, and Elliott's own summary of *Europe Divided* as 'written both under the influence of Braudel, and in reaction to him', can be taken as true for his writings as a whole.<sup>66</sup> In his use of transnational history, Elliott's work shows three key approaches- his influential idea of 'composite monarchy' has emphasised the importance of the links between centre and periphery in the early modern Spanish world, while his engagement with Atlantic history and the importance of the international context has given his work a consistently transnational focus. Finally, Elliott's use of comparative history also demonstrates the way his work includes, rather than excludes, different topics or methods of study, to produce a more well-rounded picture. This essay has sought to demonstrate that though the specific topics of his works vary, Elliott's approach has remained strikingly consistent. Thus, three works at the beginning, middle and more recent years of his career on superficially distinct subjects, *Revolt of the Catalans* (1963), *Olivares* (1986) and *Empires of the Atlantic World* (2006), all use close attention to political history and the importance of ties across borders between centre and periphery to produce rounded and 'total' historiographical pictures. Elliott has called for historians to not be 'afraid to ask big questions, or paint with a broad brush on a large canvas', and his own work, using a strikingly broad methodological inventory to produce an impressively consistent 'total' history, is the best evidence possible that he is right.<sup>67</sup> 🌟

---

<sup>66</sup>Elliott, *Europe Divided*, p.ix.

<sup>67</sup>Elliott, 'General Crisis in Retrospect', p.73.

## References

Bakewell, Peter, *Silver Mining and Society in Colonial Mexico: Zacatecas, 1546-1700* (Cambridge, 1971).

Balguer, *Historia de Catalunya*, in Elliott, John H., *The Revolt of the Catalans* (Cambridge, 1963).

Bérenger, Jean, 'Le problème du ministeriat au XVIIe siècle', in *Annales. Economies, Societes, Civilisations*. 29e année, N,1, 1974.

Braudel, Fernand, *The Mediterranean and the Mediterranean World in the Age of Philip II*, Vol.1., trans. Reynolds, Sian (Berkeley, 1995).

Burbank, Jane and Cooper, Frederick, *Empires in World History: Power and the Politics of Difference* (Princeton, 2010).

Cardim, Pedro, Herzog, Tamar, Ruiz Ibanez, Joe Jacier and Sabitini, Gaetano, *Polycentric Monarchies* (Brighton, 2014).

'Conrad Sebastian Robert Russell 1937-2004', in *Proceedings of the British Academy*, 138.

Connolley, *Divided Kingdom: Ireland 1630-1800* (Oxford, 2008).

Cultaix, Yolaine, 'Shaping of a Nation: Catalan History and Historicity in Post-Franco Spain', accessed at <https://www.lama.univ-savoie.fr/pagesmembres/vuillon/DIECKH1.pdf> 24/11/2020

Elliott, John H., 'The Decline of Spain' in Elliott, J. H., *Spain and its World 1500-1700: Selected Essays* (New Haven, 1989).

Elliott, John H., *The Revolt of the Catalans* (Cambridge, 1963).

Elliott, John H., *Imperial Spain 1469-1716* (Harmondsworth, 1963).

Elliott, John H., 'Revolution and Continuity in Early Modern Europe', in Elliott, John H., *Spain and its World 1500-1700: Selected Essays* (New Haven, 1989).

Elliott, John H., 'Power and Propaganda in the Spain of Philip IV', in Elliott, John H., *Spain and its World 1500-1700: Selected Essays* (New Haven, 1989).

Elliott, John H., *Richelieu and Olivares* (Cambridge, 1984).

Elliott, John H., *The Count-Duke of Olivares: The Statesman in an Age of Decline* (New Haven, 1989).

Elliott, John H., 'La Sociedad Cortesana en la Europa del Siglo XVII', in Brown, Jonathan and Brown, Christopher (eds.), *Velazquez, Rubuens y Van Dyck* (Madrid, 1999).

Elliott, John H., 'Self-Perception and Decline in Early Seventeenth-Century Spain', in Elliott, John H., *Spain and its World 1500-1700: Selected Essays* (New Haven, 1989).

Elliott, John H., 'Foreign Policy and Domestic Crisis: Spain, 1598-1659', in Elliott, J. H., *Spain and its World 1500-1700: Selected Essays* (New Haven, 1989).

Elliott, John H., 'Quevedo and the Count-Duke of Olivares', in Elliott, J. H., *Spain and its World 1500-1700: Selected Essays* (New Haven, 1989).

Elliott, John H., 'Illusion and Disillusionment: Spain and the Indies' in Elliott, John H., *Spain, Europe and the Wider World 1500-1800* (New Haven, 2009).

Elliott, John h., 'Britain and Spain in America', in Elliott, John H., *Spain, Europe and the Wider World 1500-1800* (New Haven, 2009).

Elliott, John H., 'A Provincial Aristocracy: The Catalan Ruling Class in the Sixteenth and Seventeenth Centuries', in Elliott, J. H., *Spain and its World 1500-1700: Selected Essays* (New Haven, 1989).

Elliott, John H., 'The Mental World of Hernan Cortes', in Elliott, J. H., *Spain and its World 1500-1700: Selected Essays* (New Haven, 1989).

Elliott, John H., 'National and Comparative History: An Inaugural Lecture delivered before the University of Oxford on 10th May 1991' (Oxford, 1991).

Elliott, John H., *Europe Divided 1559-1598*, 2nd ed. (Oxford, 2000)

Elliott, John H., and Brown, Jonathan, *A Palace for a King: The Buen Retiro and the Court of Philip IV* (New Haven, 2003).

Elliott, John H., 'King and Patria in the Hispanic World', in Elliott, John H., *Spain, Europe and the Wider World 1500-1800* (New Haven, 2009).

Elliott, John H., *Empires of the Atlantic World: Britain and Spain in America, 1492-1830* (New Haven, 2006).

Elliott, John H., 'The General Crisis in Retrospect: A Debate without End', in Elliott, John H., *Spain, Europe and the Wider World 1500-1800* (New Haven, 2009).

Elliott, John H., 'A Europe of Composite Monarchies', in Elliott, John H., *Spain, Europe and the Wider World 1500-1800* (New Haven, 2009).

Elliott, John H., *History in the Making* (New Haven, 2012).

Gil, Xavier, 'John Elliott en sus artículos', in Fernandez, Roberto, Passola, Antonio and Vilalta, Maria Hose (eds.), *John Elliott, El Oficio de Historador*.

Green, Anna and Troup, Kathleen, *The Houses of History: A Critical Reader in Twentieth Century History and Theory* (Manchester, 1999).

Gregory, Brad S., 'Is Small Beautiful? Microhistory and the History of Everyday Life', in *History and Theory*, Feb. 1999, Vol.38, No.1.

Himmelfarb, Gertrude, *The New History and the Old: Critical Essays and Reappraisals*, Rev. ed. (Cambridge, 2004).

Hobsbawm, E. J., 'The Crisis of the Seventeenth Century', in Aston, Trevor (ed.), *Crisis in Europe 1560-1660* (Crawley, 1970).

Huppert, George, 'The *Annales* Experiment', in Bentley, Michael (ed.), *Companion to Historiography* (London, 1997).

Interview 'An Englishman Abroad: Sir John Elliott and the Hispanic World', *Itineratio*: <https://ezproxy-prd.bodleian.ox.ac.uk:2117/core/services/aop-cambridge-core/content/view/92246FB7680423B99766A45C2AE9A315/S0165115300006768a.pdf/an->

englishman-abroad-sir-john-elliott-and-the-hispanic-world.pdf, Accessed 11/01/2021.

Kagan, Richard L., 'People and Places in the Americas: A Comparative Approach', in Canny, Nicholas and Morgan, Philip (eds.), *The Oxford Handbook of the Atlantic World* (Oxford, 2012).

Kagan, Richard L. and Parker, Geoffrey, 'Sir John Elliott: An appreciation', in Kagan, Richard L. and Parker, Geoffrey (eds.), *Spain, Europe and the Atlantic World: Essays in honour of John H. Elliott* (Cambridge, 1995).

Kennedy, Paul, *The Rise and Fall of the Great Powers* (New York, 1987).

Kumar, Krishnan, *Visions of Empire: How Five Imperial Regimes shaped the World* (Princeton, 2017).

Mitchell, Andrew, 'The Sacral Limits of Empire', in Andrade, Tonio and Reger, William (eds.), *The Limits of Empire: European Imperial Formations in Early Modern World History- Essays in Honor of Geoffrey Parker* (Abingdon, 2016).

Mousnier, Ronald, in Dewald, Jonathan ; Crisis, Chronology and the Shape of European Social History', in *The American Historical Review*, Vol.113, No.4 (October, 2008).

Munz, Peter, 'The Historical Narrative', in Bentley, Michael (ed.), *Companion to Historiography* (London, 1997).

Ogilvie, Sheillagh, 'Germany and the Seventeenth-Century Crisis', in Parker, Geoffrey and Smith, Lesley M. (eds.), *The General Crisis of the Seventeenth Century*, 2nd ed. (London, 1997).

Parker, Geoffrey, *Global Crisis: War, Climate Change and Catastrophe in the Seventeenth Century* (New Haven, 2013).

Parker, Geoffrey, 'Reflections on a Misspent Past', in Andrade, Tonio and Reger, William (eds.), *The Limits of Empire: European Imperial Formations in Early Modern World History-Essays in Honor of Geoffrey Parker* (Abingdon, 2016).

Kagan, Richard L. and Parker, Geoffrey (eds.), *Spain, Europe and the Atlantic World: Essays in honour of John H. Elliott* (Cambridge, 1995).

Parker, Geoffrey, *Imprudent King: A New Life of Philip II* (New Haven, 2014).

Popkin, Jeremy P., *From Herodotus to H-Net: The Story of Historiography* (Oxford, 2015).

'Professor Sir John Elliott: Interview', 07/03/2008' accessed at <https://archives.history.ac.uk/makinghistory/resources/interviews/ElliottJohn.html> on 24/11/2020.

Rawlings, Helen, *The Debate on the Decline of Spain* (Manchester, 2012).

Scott, Joan, 'Gender: A Useful Category of Historical Analysis', in *The American Historical Review*, Dec. 1986, Vol.91, No.5.

'Solemne Investidura de Doctor Honoris Causa al professor Sir John Elliott', Universitat de Barcelona (Barcelona, 1994).

Steensgaard, Niels, 'The Seventeenth-Century Crisis', in Parker, Geoffrey and Smith, Lesley M. (eds.), *The General Crisis of the Seventeenth Century*, 2nd ed. (London, 1997).

Thompson, I. A. A., 'Castile, Spain and the Monarchy', in Kagan, Richard L. and Parker, Geoffrey (eds.), *Spain, Europe and the Atlantic World: Essays in honour of John H. Elliott* (Cambridge, 1995).

Vilar, Jean Berrogain, *Literatura y Economica: La Figura Satirica del arbitrista en el Siglo de Oro* (Madrid, 1973).

Webster, John, *The Duchess of Malfi*, ed. Britland, Karen (London, 2021).



# The Discovery and Development of Risdiplam – a small molecule SMN2 splicing modulator for the treatment of Spinal Muscular Atrophy

*Harry Woodcock\**

*Department for Continuing Education, University of Oxford*

## **Abstract**

Risdiplam received FDA approval in 2020 for the treatment of spinal muscular atrophy in adults and children over the age of two months. The approval of Risdiplam validates small molecule RNA splice modulation as a new therapeutic modality and is a remarkable story of drug development.

## **Disease Background and Target Rationale**

Spinal muscular atrophy (SMA) is the most common childhood genetic cause of death (Crawford et al. 1996), with an incidence of 1 in 10,000 live births. SMA is caused by homozygous disruption of the survival motor neuron 1 (SMN1) gene. Patients are characterised by degeneration of  $\alpha$ -motor neurons in the anterior horn of the spinal cord (Tisdale et al. 2015) (Fig. 1), leading to progressive muscular atrophy and childhood mortality in 50% of cases.

The human genome contains two genes capable of generating SMN protein. Duplication of the large SMN1-containing portion of chromosome 5q results in the highly homologous SMN2 gene (Stabley et al.2015). These genes are almost identical, with the exception of 5 single nucleotide variances (SNVs). Importantly, the C to T SNV in exon 7 ablates an exonic splicing enhancer (ESE) and creates an exonic splicing silencer (ESS) (Cartegni et al.2002, Kashima et al.2003). This SNV results in exon 7 skipping in 80-90% of SMN2 transcripts, generating a truncated SMN isoform known as  $\Delta 7$ SMN. Exon skipping in SMN2 is not entirely faithful, and a small

---

\*harry.woodcock@new.ox.ac.uk

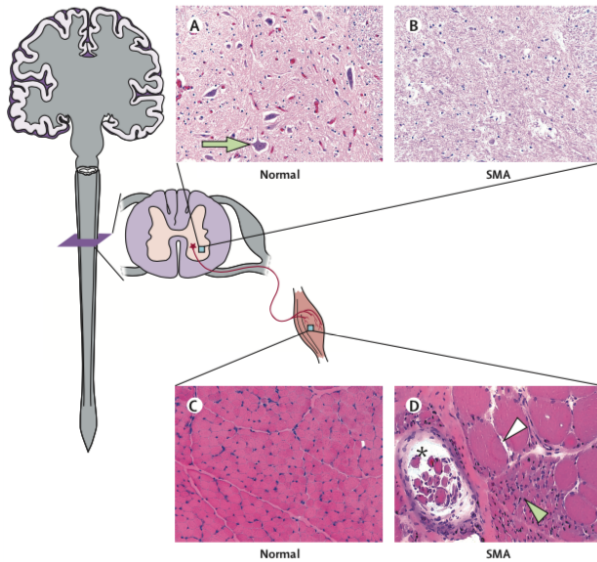


Figure 1: Histopathology of SMA. (Lunn et al. 2008)

amount of full length, functional SMN protein is generated through the SMN2 gene. In SMA patients, for whom SMN1 is disrupted, this low level of functional full-length protein generated by the SMN2 gene prevents lethality (Fig 2).

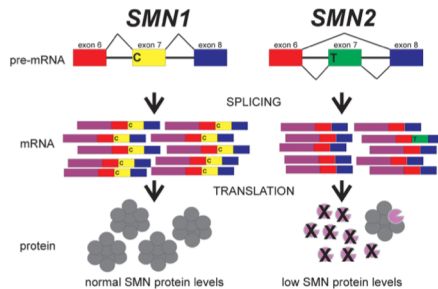


Figure 2: Schematic depicting the impact of C to T transition in exon 7 of SMN2 on splicing (Butchbach et al. 2016)

The linkage of the chromosomal region 5q11.2-q13.3 to SMA in 1990 (Brzustowicz et al.1990, Melki et al.1990), and resulting identification and characterisation of the SMN1 and SMN2 genes (Lefebvre et al.1995) has enabled a deeper under-

standing of the molecular pathogenesis of SMA. SMN protein is expressed ubiquitously, and it is not understood why reduced expression preferentially impacts motor neurons. There has been much research into the role of SMN protein to understand the molecular pathogenesis. It is established that SMN forms a multiprotein complex (Pellizzoni et al.2007) required for the assembly of small nuclear ribonucleoproteins (snRNPs) of the major (U2-dependent) and minor (U12-dependent) spliceosomes (Fischer et al.2011, Li et al.2014), highlighting a key role for SMN protein in the processing of intron-containing mRNAs. This is supported through the correlation of spliceosomal snRNP dysfunction with disease severity in mouse SMA models (Gabanella et al.2007). Secondary to the role of SMN in major and minor spliceosomal snRNPs, SMN also facilitates the assembly of the U7 snRNP. This complex is involved in the 3'-end processing of histone mRNAs. As expected, SMN deficiency leads to histone mRNA processing interruption in SMA (Tisdale et al.2013).

Mouse SMA models have been informative in elucidating the selective vulnerability of motor neurons to SMA. SMA mice motor neurons have particularly low levels of full length, functional SMN2 transcript, and a higher level of  $\Delta 7$ SMN protein and appear selectively vulnerable to snRNP reduction (Ruggiu et al.2012). There is also activation of a negative feedback loop, in which reduced snRNPs further decrease exon 7 splicing in SMA (Jodelka et al.2010, Ruggiu et al.2012). These findings provide insight into the potential mechanisms for the selective vulnerability of motor neurons in SMA.

SMA can be divided into type 1-4 based on disease severity (Table 1), however the boundaries between these categories are arbitrary and there is a spectrum of clinical severity (Markowitz et al.2004).

Type	Age of onset	Requires respiratory support at birth	Able to sit	Able to stand	Able to walk	Life expectancy	Predicted SMN2 copy number
0	Prenatal	Yes	No	No	No	<6 months	1
1	<6 months	No	No	No	No	<2 years	2
2	6-18 months	No	Yes	No	No	10-40 years	3
3	> 18 months	No	Yes	Yes	Assisted	Adult	3-4
4	>5 years	No	Yes	Yes	Yes	Adult	>4

Table 1: Clinical classification of SMA types (Butchbach et al.2016)

Importantly for the prognosis of SMA patients, multiple copies of SMN2 can be present in the genome as a result of gene conversion events (Stabley et al.2020). Additional SMN2 copies increase levels of functional SMN protein and correlate inversely with disease severity. SMA carriers demonstrate an inverse correlation between SMN1 and SMN2 copy numbers, consistent with gene conversion (Fig 3).

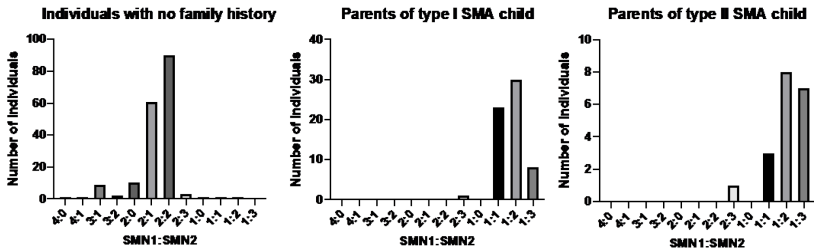


Figure 3: SMN1:SMN2 copy number ratio of parents of SMA patients or individuals with no family history (adapted from Ogino et al.2003).

SMN2 copy number is the most established prognostic biomarker for SMA. SMA symptoms can be masked entirely, through having 5 copies of the SMN2 gene (Prior et al.2004). In these patients, the low level of functional SMN protein generated by SMN2 is increased through high copy number to levels sufficient to prevent pathogenesis. This remarkable finding highlights the therapeutic potential of increasing functional SMN protein levels through SMN2 splice modulation.

All SMA patients carry 1 gene copy of SMN2 (Rao et al.2018), making this an attractive target. Antisense oligonucleotides (ASOs) are well established as a modality capable of splice modulation. Nusinersen is an ASO approved in 2016 for the treatment of SMA in patients of any age. It is injected intrathecally and requires repeated dosing. In addition to its invasive administration, it is particularly expensive, at \$750,000 in the first year, reducing to \$350,000 for subsequent years (Wirth et al.2021). Risdiplam, approved in 2020, is the first approved small molecule splice modulator. Orally administered and highly selective for SMN2, Risdiplam costs \$100,000 - \$340,000 per year based on patient weight (Wirth et al.2021).

### The Discovery of Risdiplam

The first SMN2 splicing modulators were identified by Naryshkin et al. in their seminal Science paper (Naryshkin et al.2014). The group developed a high throughput screen (HTS) in HEK293H cells expressing a SMN2 minigene. This minigene contains exons 6, 7 and 8 and introns 6 and 7 of SMN2 fused to a firefly luciferase reporter. Luminescence is only observed when full-length SMN2 minigene is translated. In the  $\Delta 7$  gene product, the SMN2 translation initiation codon is not in frame with the luciferase coding sequence resulting in no luminescence. A number of compound scaffolds were identified, which increased exon 7 inclusion. Three of these

chemical series were identified as orally available. Hit compounds SMN-C1, SMN-C2 and SMN-C3 (Fig 4C) increased full length SMN2 mRNA levels with a reduction of  $\Delta 7$ SMN2 mRNA in a primary cell assay utilising SMA type 1 patient fibroblasts (Fig 4A,B) (Naryshkin et al. 2014). Lead compounds were then investigated in a motor neuron cell model to provide a more disease relevant cell type to assess their efficacy. Motor neurons generated from SMA type 1 and 2 patient induced pluripotent stem cells (iPSCs) exhibited increased SMN protein expression when treated with SMN-C3 (Naryshkin et al.2014). At this stage, hits were confirmed as SMN2 splicing modulators with demonstrated activity in relevant cell types modelling a spectrum of SMA disease severities. Importantly, these compounds must be selective for SMN2 over other pre-mRNAs. RNA-seq was employed to understand SMN2 selectivity over other pre-mRNAs. 11,714 expressed genes were assessed for compound treatment effects. Twelve genes were identified, where a greater than two-fold shift in transcript levels was observed (Fig. 5A). Compound effects on selected gene families are presented in Fig. 5B (Naryshkin et al..

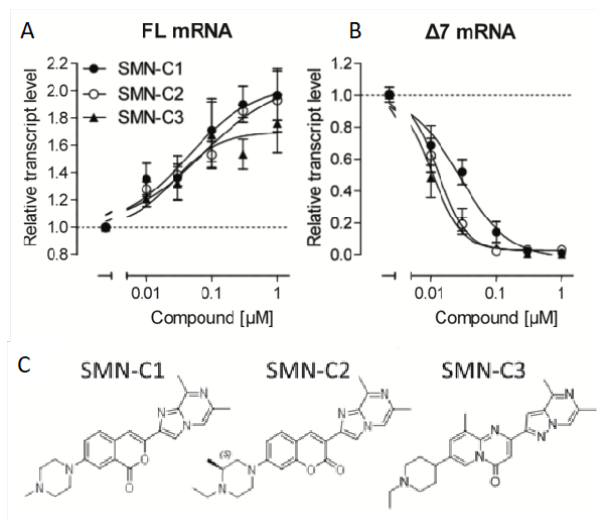


Figure 4: A, B RT-PCR of SMN2 and  $\Delta 7$ SMN2 transcript in response to 24-hour compound treatment in SMA type I patient fibroblasts. (A) SMN transcript level increase is dose dependent (B)  $\Delta 7$ SMN2 transcript decrease is dose dependent (C) Chemical structures of SMN-C compounds (Naryshkin et al.2014)

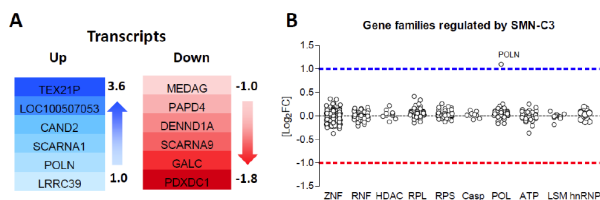


Figure 5: Transcriptional profiling of SMN-C3. (A) Significant transcriptional changes in SMA type I patient fibroblasts treated with SMN-C3. A threshold of 2-fold over DMSO treated is applied. SMN2 displayed no significant change in total mRNA abundance (B) Transcriptional changes to selected gene families. Zinc finger proteins (ZNF), ring finger proteins (RNF), histone deacetylases (HDAC), ribosomal proteins from large subunit (RPL), ribosomal proteins from small subunit (RPS), caspases (Casp), DNA and RNA polymerases (POL), ATPases (ATP), Like Sm proteins involved in the splicing machinery (LSM) and Heterogeneous nuclear ribonucleoproteins involved in splicing regulation (hnRNP). (Naryshkin et al.2014.)

Lead compounds (SMN-C class) progressed to an *in vivo* mouse model to assess efficacy. Liquid chromatography-tandem mass spectrometry (LC-MS) was employed to demonstrate excellent brain penetration for all lead SMN compounds in the mild C/C SMA mouse model (Naryshkin et al.2014). SMN2 full length mRNA increased from a baseline of 40% to 90% within four hours of compound dosing (Naryshkin et al.2014). Compounds progressed to the more aggressive  $\Delta 7$  SMA transgenic mouse model. Ordinarily, these mice have a median survival of 18 days. SMN-C3 treatment increased survival in a dose-dependent manner. In the two highest dose groups, 90% of the 32 animals survived beyond study completion at postnatal day 65 (Fig 6) (Naryshkin et al.2014). The SMN C compounds 1-3 discovered by Naryshkin et al. have demonstrated selectivity for SMN2, efficacy and brain penetrance. This series of compounds proceeded to lead optimisation, with a strategy to deliver compounds that were not P-glycoprotein (P-gp) substrates and thus not pumped out of cells, with no potential for genotoxicity and a suitable Pharmacokinetic (PK) profile. This lead optimisation resulted in the discovery of RG7800 (Fig 9). Importantly, further characterisation was required to understand the mechanism of action of the SMN C class of compounds. This was documented in a series of experiments by Sivaramakrishnan et al. summarised below.

To understand the mechanism of action of lead compounds, pharmacological agents were employed to block biological processes from DNA transcription to protein translation. These agents were able to confirm that SMN-C compounds act on pre-mRNA and not transcription or translation (Sivaramakrishnan et al.2017). We

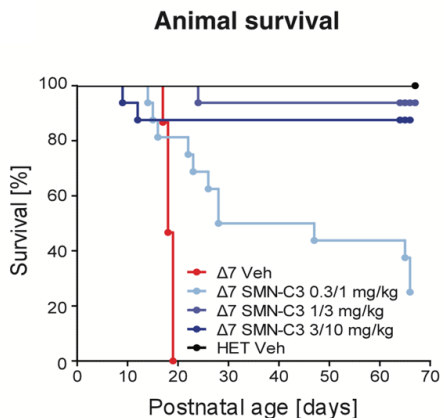


Figure 6: Kaplan-Meier survival curves.  $\Delta 7$  SMA mice treated with SMN-C3 vs untreated  $\Delta 7$  SMA mice. (Naryshkin et al.2014).

can predict a similar mechanism to that established for NVS-SM1, an alternative SMN2 splice modulator which shares a pharmacophore with SMN-C3. NVS-SM1 has been demonstrated to increase the binding affinity of U1 spliceosomal RNA (U1 snRNP) for the 5' splice site (5'ss) of exon 7 (Palacino J et al. 2015). As predicted, nuclear magnetic resonance confirmed that SMN-C3 binds to the 5' ss of SMN2, albeit in a different location to NVS-SM1 (Sivaramakrishnan et al.2017). Whilst NVS-SM1 is a useful tool and shares similarities in structure with SMN-C3, it is far less selective for SMN2 and causes profound perturbations across the transcriptome (Fig 7). In order to explore the high selectivity of SMN-C class compounds, Sivaramakrishnan et al. performed RNA-sequencing on SMA type I fibroblasts in the presence of SMN-C3. SMN-C3 exhibited strong splice modulation of two genes, SMN2 and STRN3. A small region of homology was identified between these two gene sequences, with STRN3 containing a weak 5' ss and a purine-rich region with high similarity to the exonic splicing enhancer 2 (ESE2) motif, present in SMN2. NVS-SM1 does not recognise SMN2 ESE2, supporting the hypothesis that ESE2 binding provides SMN-C compounds with their superior specificity to SMN2 (Sivaramakrishnan et al. 2017). Disruption of both ESE2 and 5'ss renders SMN-C5 ineffective as a SMN2 splice modulator. SMN-C5 retains some splice modulating activity when one of ESE2 or 5'ss is disrupted (Fig 8 B, C). These data support the presence of two binding sites driving increased selectivity for SMN2. Disruption of 5'ss causes a greater reduction in efficacy, suggesting that 5'ss binding is the main driver of potency.

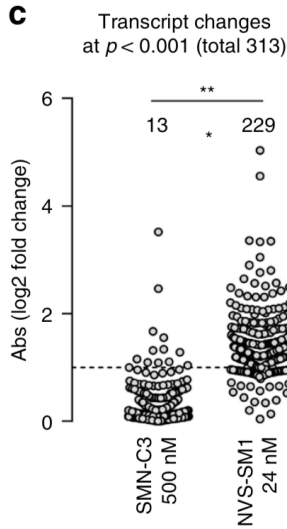


Figure 7: RNA-Seq analysis in type 1 SMA fibroblasts. Transcriptome-wide gene expression changes for SMN-C3 and NVS-SM1 compounds (Sivaramakrishnan et al.2017).

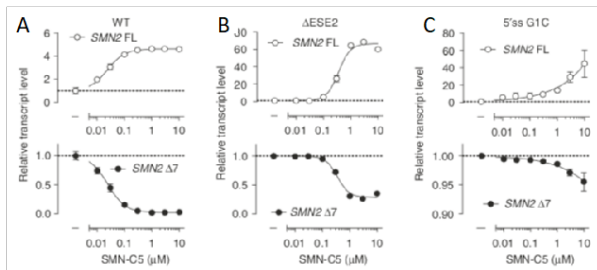


Figure 8: RT-PCR data demonstrating SMN-C5 impact on SMN2 transcript levels (B)  $\Delta$ ESE2 Deletion in ESE2 region. (C) 5'ss G1C mutation. (Sivaramakrishnan et al.2017).



## Characterisation of RG7800

RG7800 was progressed from lead compound, into a Phase 1 clinical trial in healthy adults (Eudract number 2013-004097-95) where it was well tolerated and there were no serious adverse events reported. These data supported progression to a Phase 2 trial in SMA patients (MOONFISH). In both trials, increased full-length SMN2 mRNA was reported (Kletz et al. 2019). The MOONFISH Phase I trial of RG7800 reported no serious adverse events. However, dosing was suspended due to retinal toxicity observed in a concurrent RG7800 toxicology assessment in cynomolgus monkeys (Sturm et al. 2019). This was believed to be an off-target effect, as monkeys do not possess the SMN2 gene. This finding resulted in further optimisation of RG7800 and the discovery of Risdiplam.

## Optimisation of Selectivity

The findings from the first-in-patient RG7800 study (MOONFISH) supported the proof of mechanism but concerning animal retinal toxicity findings highlighted the need for a molecule with increased therapeutic index. Improved selectivity was targeted to achieve this. The RNA-seq work published by Naryshkin et al. previously (Fig 5), identified a small number of perturbed genes with well characterised roles in cell death or cell cycle regulation. Ratni et al. focussed on the off-target activity of RG7800 on FOXM1 splicing in an attempt to improve the therapeutic index (Ratni et al. 2018). FOXM1 encodes forkhead box protein M1, a key protein in cell division, highly expressed in rapidly dividing cells. The identification of this gene as a secondary splice target of compounds in this pyridopyrimidinone series is supported by a clear correlation for SMN2 and FOXM1 splice modulation. Disturbance of FOXM1, is consistent with *in vitro* evidence of cell cycle arrest, micronucleus induction, and apoptosis. Additionally, germ cell degeneration observed in cynomolgus monkey tissue after treatment with RG7800 is suggested to be due to interruption of FOXM1 splicing.

## Optimisation of RG7800

RG7800 has a large volume of distribution ( $V_{ss}$ ) and is basic with a high  $pK_a$ , consistent with phospholipidosis observed in rat (Ratni et al.2018). Reduction of basicity and  $V_{ss}$  became key targets for optimisation to improve PK. Oral dosing of RG7800 in rodents and monkeys identified a metabolite with 10-fold increased potency for SMN2 and FOXM1 splice modulation. This metabolite is a strong p-Gp substrate, with an efflux ratio (ER) of 18.7 in human and 20.0 in mouse cells (Ratni et al.2018) resulting in this metabolite being pumped out of the CNS and restricted to peripheral

tissues. Peripheral restriction causes lack of efficacy due to low CNS penetrance, but retention of undesirable FOXM1 splice modulation in peripheral tissues. To improve the therapeutic window, avoiding the formation of peripherally restricted active metabolites was a key path. Suppressing the N-dealkylation would block the formation of N-dealkylated active metabolites. It had previously been demonstrated for this pyridopyrimidinone series that lipophilicity (logD) is a strong predictor of p-Gp efflux (Ratni et al.2018), maintaining a logD between 1.6 and 3.0 to achieve a p-Gp ER <2.5 was key to preventing high clearance. A balance was required between basicity and lipophilicity while maintaining high potency. A virtual compound library was generated through modifications to the left-hand-side amine moiety with compound logD between 1.6 and 3, and pKa between 6.5 and 8 (Fig 9A). Shortlisted compounds were assessed *in vitro*. Risdiplam was selected as the most promising compound due to its very high potency, strongly reduced basicity, lipophilicity of 2.5 and lack of active metabolites (Fig. 9, Table 2). Risdiplam has a favourable PK profile over RG7800 with  $V_{ss}$  significantly reduced (Table 1). Risdiplam is compliant with Lipinski's rule of 5, with a molecular weight of 401 g/mol, XLogP of 3.52, 5 hydrogen bond acceptors and 1 hydrogen bond donor. The main metabolite of Risdiplam is inactive for SMN2 or FOXM1 splice modulation due to removal of basicity (Ratni et al.2018). The improved therapeutic window for Risdiplam over RG7800 made Risdiplam a safer compound to proceed with.

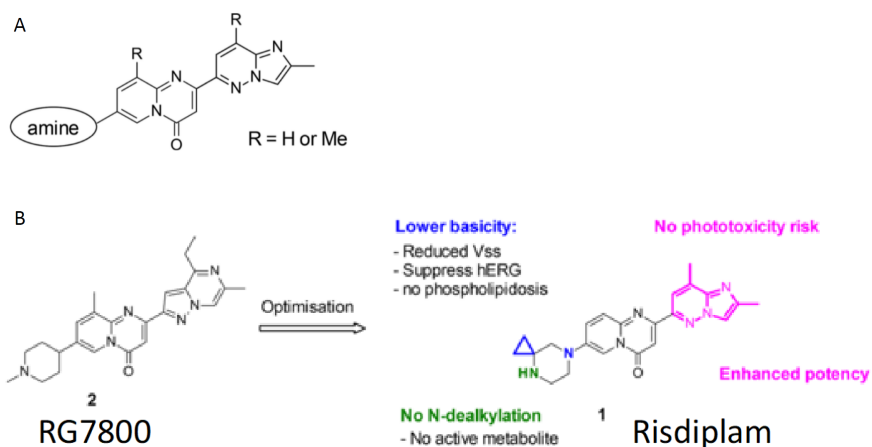


Figure 9: A) Virtual compound library focused on selecting a left-hand-side basic amine moiety with the lowest  $pK_a$  (B) Optimisation of RG7800 leading to the discovery of Risdiplam (Ratni et al.2018)

compd	rat			
	Cl <sup>a</sup> (mL min <sup>-1</sup> kg <sup>-1</sup> )	V <sub>ss</sub> <sup>a</sup> (L/kg)	T <sub>1/2</sub> <sup>b</sup> (h)	F <sup>b</sup> (%)
1	8.9	3.1	6.4	~100
2	25	29	19	~100

compd	cynomolgus monkey			
	Cl <sup>c</sup> (mL min <sup>-1</sup> kg <sup>-1</sup> )	V <sub>ss</sub> <sup>c</sup> (L/kg)	T <sub>1/2</sub> <sup>d</sup> (h)	F <sup>d</sup> (%)
1	5.7	2.0	5.4	43
2	5	20	42	52

Table 2: *In vivo* single dose PK profile for compounds 1 (Risdiplam) and 2 (RG7800)  
<sup>a</sup>For compound 1: (a) iv, 1.9 mg/kg; (b) po, 5.5 mg/kg; (c) iv, 0.1 mg/kg; (d) po, 0.5 mg/kg. For compound 2: (a) iv, 2 mg/kg; (b) po, 5 mg/kg; (c) iv, 0.3 mg/kg; (d) po, 1.3 mg/kg. (Ratni et al. 2018).

compd	SMN2 splicing EC <sub>1.5x</sub> [nM]	FOXM1 splicing EC <sub>50</sub> [nM]	SMN protein EC <sub>1.5x</sub> [nM]	log D	h P-gp ER	mpKa	f <sub>up</sub> [%] h/m
1	4	67	29	2.5	2.2	6.8	11/10
2	23	247	87	2.3	2.0	10.9	13/2

Table 3: *In vitro* characterisation for compound 1 (Risdiplam) and compound 2 (RG7800) (Ratni et al.2018).

### Characterisation of Risdiplam

Risdiplam was identified through structure-based drug design. In 2015, the PK of Risdiplam in healthy adults were characterised in the entry-into-human (EIH) study (NCT02633709). Maximum exposure was set before study commencement, to limit individual subject exposure. A target exposure was identified to assess potency of Risdiplam (Sturm et al.2019). This was the first example of Bayesian adaptive design principles being used in combination with emergent Pharmacodynamic (PD) data to inform dose escalation in an EIH study. This novel approach enabled assignment of more patients to more informative pharmacological doses than would have been possible with traditional methods. Importantly, the Bayesian adaptive design prioritised subject safety, as dose escalation was only performed after assessment of

safety, tolerability and available PK/PD data for the previous dose. A starting dose of 0.6mg Risdiplam was selected as this was predicted to be well tolerated in humans. The predicted therapeutic dose determined by PK/PD modelling was 2-22mg (Sturm et al.2019).

## Absorption

Risdiplam displayed linear PK over the 0.6-18.0mg dosing range (Fig 10A) suggesting constant clearance across the concentration range. Both  $C_{max}$  and area under the curve (AUC) increased in a dose-dependent manner (Fig 10A). Following oral administration, time taken to reach maximum plasma concentration ( $T_{max}$ ) was between 1 and 4 hours (Fig 10B) (Sturm et al.2018). An exploratory fasted vs fed comparison suggested that food has no effect on the  $C_{max}$  or AUC of Risdiplam. The half-life ( $T_{1/2}$ ) of Risdiplam was 40.1-68.7 hours for doses between 2.0 and 18.0mg (Sturm et al.2019). Following once daily oral-administration of Risdiplam in healthy adults, 3 fold accumulation of  $C_{max}$  and AUC<sub>0-24h</sub> was observed. Risdiplam reaches steady state exposure 7-14 days after once-daily administration (FDA prescribing information).

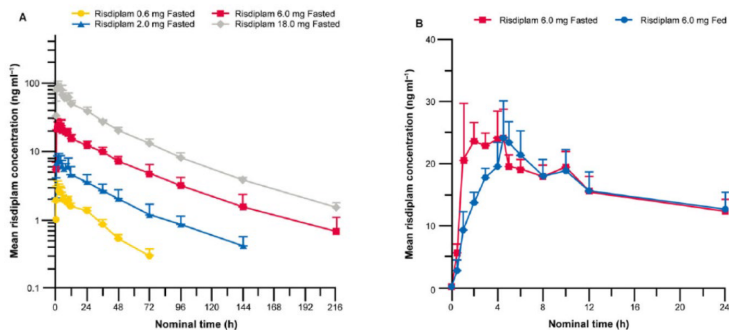


Figure 10: Mean plasma concentration vs time following single dose. (Sturm et al.2019).

## Distribution

Preclinical data provides evidence for systemic distribution and brain penetration in a range of species. Single or repeat intraperitoneal (IP) administration of Risdiplam for up to 39 weeks resulted in similar total drug levels in plasma, brain and muscle of rats, mice and monkeys (Poirier et al.2018) (Fig 11).

Human PK data from the FIREFISH and SUNFISH trials (described later) report an apparent  $V_{ss}$  of 6.3L/kg, central volume of distribution ( $V_c$ ) of 98L and peripheral volume of distribution ( $V_t$ ) of 93L, supporting systemic distribution. Risdiplam has a free fraction of 11%, predominantly binding to serum albumin (FDA prescribing information).

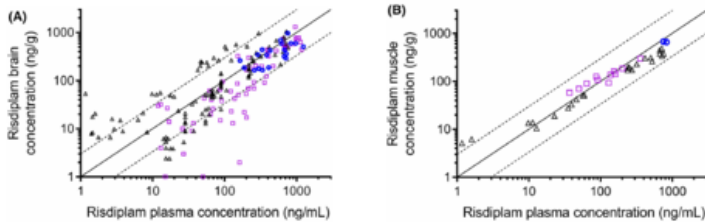


Figure 11: Risdiplam plasma vs tissue distribution in Mice (open squares), Rats (open triangles) and monkeys (open circles). Poirier et al.2018.

### Metabolism

Itraconazole, a strong cytochrome protein 3A (CYP3A) inhibitor was co-administered with Risdiplam to investigate the effect on PK. There was no clinically relevant effect on the metabolism of Risdiplam in the presence of itraconazole (Fig 12), suggesting a non-CYP mediated primary route of metabolism (Sturm et al.2018). Risdiplam is primarily metabolised by flavin monooxygenase 1 and 3 (FMO1 and FMO3). Small molecules primary metabolised by FMO are less likely to encounter drug-drug interactions (DDIs) and related potentially harmful side effects. Parent drug was the major component found in plasma, making up 83% of drug material in circulation. The pharmacologically inactive M1 metabolite was identified as the major circulating metabolite (Sturm et al.2018).

### Elimination

The apparent clearance (CL/F) of Risdiplam for a 14.9 kg patient is 2.1 L/h. The  $T_{1/2}$  of Risdiplam was 50 hours in healthy adults, with faeces being the primary route of excretion accounting for 53% of the dose, followed by urine, accounting for 28% (FDA Prescribing information).

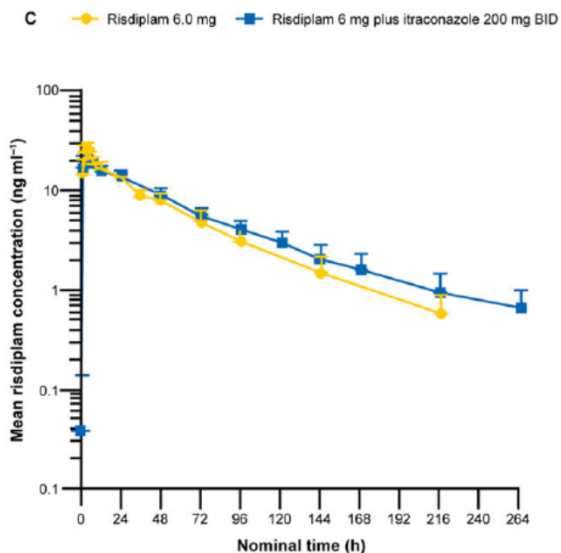


Figure 12: Plasma concentration vs time profile of Risdiplam following single oral doses of 6mg (Sturm et al.2018)

### Pharmacodynamics

Healthy patients dosed with Risdiplam displayed increased levels of SMN2 full length transcript and concomitant reduction in  $\Delta 7$ SMN transcript levels (Sturm et al.2018) (Fig 13). In SMA patient trials, Risdiplam resulted in a >2-fold increase in SMN protein levels within 4 weeks of treatment initiation (FDA prescribing information).

### Retinal Toxicity Controversy

Due to the toxicity observed in the related compound RG7800, Risdiplam was assessed for retinal toxicity in cynomolgus monkeys with 39 weeks of daily treatment. Monkeys were dosed at 3 levels. Retinal toxicity consisting of peripheral photoreceptor degeneration and microcystoid macular degeneration (MMD) was observed 5-6 months after daily treatment with Risdiplam. Photoreceptor degeneration was observed at the mid and high doses of Risdiplam (Fig 16). The mid-dose corresponds to >2-fold of the mean exposure achieved at the proposed therapeutic dose. MMD was observed in monkeys treated with the high dose only and was reversible (Sergott et al.2020).

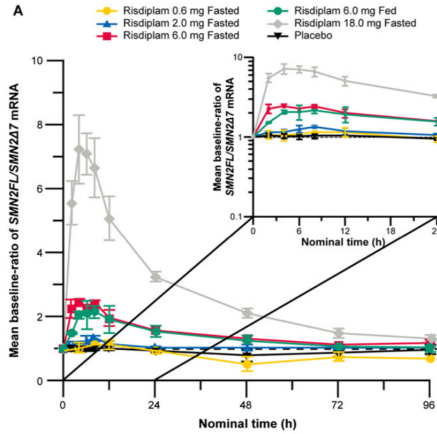


Figure 13: RT-qPCR assessment of baseline ratio of SMN2FL/ $\Delta$ 7SMN (Sturm et al.2018)

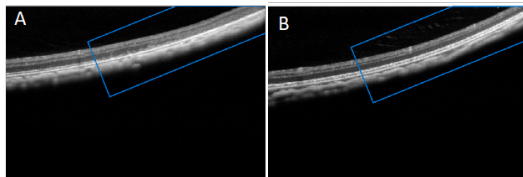


Figure 14: SD-OCT scan of cynomolgus monkey photoreceptors. (A) Retinal degeneration of peripheral retina, loss of layers and thinning. (B) Unaffected control (Sergott et al. 2020).

The proposed mechanism of retinal toxicity is due to the high melanin-binding capacity of Risdiplam, resulting in its retention in pigmented retinal cells (Ratni et al.2018). The high Risdiplam concentration in retinal cells may impair lysosomal/ autophagosomal pathways, impacting recycling of photoreceptors (Ratni et al.2018). This phototoxicity appears to be species specific, with no evidence in pigmented rats (Sergott et al.2020). Preclinical data suggests that a therapeutic effect is achieved with 2-fold increase in SMN protein levels. This level was reached at exposures corresponding to the no-observed-adverse-effect level (NOAEL) for retinal toxicity in monkeys (Sergott et al.2020). Risdiplam proceeded to clinical development with comprehensive ophthalmologic assessment.

### Clinical Trials

Risdiplam is currently under evaluation in four phase 2/3 clinical studies, summarised in Fig 14. These studies are designed to assess safety, tolerability and efficacy of Risdiplam in SMA patients.

#### FIREFISH – NCT02913482

FIREFISH is an open-label study to assess safety and efficacy in type 1 SMA infants. Part 1 of the study was dose-finding to select a safe and efficacious dose for part 2. Part 1 was divided into two cohorts, a high-dose tested in 17 infants and a low-dose tested in 4 infants. The high-dose targeted an exposure at steady-state of  $AUC_{0-24h, ss} = 2000 \text{ ng h / mL}$ . The low-dose targeted an exposure at steady-state of  $AUC_{0-24h, ss} = 700 \text{ ng h / mL}$ .

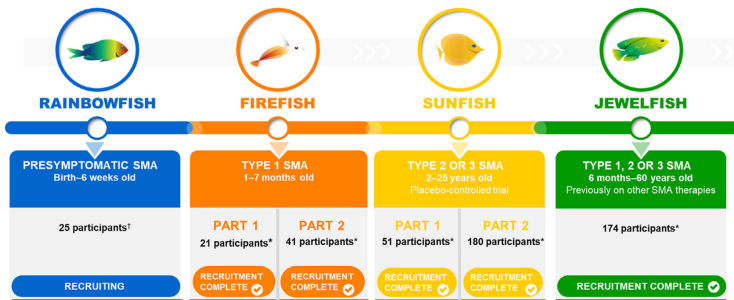


Figure 15: Risdiplam clinical development program overview (Sergott et al.2020).

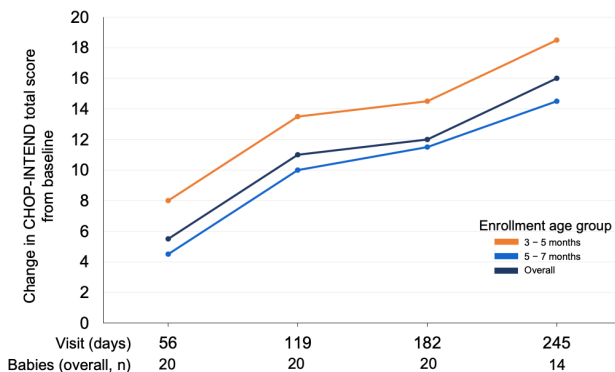
### Results : Part 1



No Risdiplam-related ophthalmologic complications were observed. 86% of all infants demonstrated improved movement scores on the CHOP-INTEND (Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders) scale (Fig 15). 90.5% of infants were alive and not requiring permanent ventilation at 12 months of treatment. For contrast, the prognosis for untreated SMA type 1 is death or permanent ventilation by 2 years of age. The primary outcome of part 1 was to establish a recommended dose, this dose was established at 0.2 mg / kg. This dose was safe, well-tolerated and efficacious, achieving the desired exposure at steady-state of  $AUC_{0-24h, ss} = 2000 \text{ ng h / mL}$ . (Baranello et al.2021).

### Results : Part 2

Part 2 initiated with starting doses of 0.04 mg / kg, 0.08 mg / kg or 0.2 mg / kg. Within a few months of treatment initiation, all subjects were adjusted to 0.2 mg / kg, adjusted to 0.25 mg / kg when patients reach 2 years of age (clinicaltrials.gov). At 12 months, 29% of infants were able to sit for five seconds unassisted. Ordinarily, SMA type 1 patients are not able to sit unsupported. 90% of infants had a CHOP-INTEND score increase of >4 points with a median increase of 20 points (clinicaltrials.gov).



Intent-to-treat patients from FIREFISH Part 1 (N=21). Day 245 (8 months) is the latest time point at which the majority of babies have follow-up data available. Data cut-off: 7 September 2018.

Figure 16: Change in CHOP-INTEND total score from baseline. (Seabrook et al.2019)

### Summary

The discovery and development of Risdiplam is a seminal moment in drug discovery, as the first approved RNA targeting small molecule. This landmark achieve-

ment is even more impressive when we consider the potentially derailing phototoxicity observed in cynomolgus monkeys. The thorough safety assessment and compound optimisation to achieve a high therapeutic index, enabled progression of clinical trials. This inspirational story validates the entirely new modality of splice switching for small molecule drug discovery. 🌟

## References

- Aumann S., Donner S., Fischer J., Müller F. (2019) *Optical Coherence Tomography (OCT): Principle and Technical Realization*. In: Bille J. (eds) *High Resolution Imaging in Microscopy and Ophthalmology*. Springer, Cham. [https://doi.org/10.1007/978-3-030-16638-0\\_3](https://doi.org/10.1007/978-3-030-16638-0_3).
- Baranello G, Darras B. T. et al. (2021) Risdiplam in Type 1 Spinal Muscular Atrophy. *The New England Journal of Medicine* 384(10):915-923.
- Berard C, Payan C. et al. (2005) A motor function measure for neuromuscular diseases. Construction and validation study. *Neuromuscular Disorders* 15(7): 463-470.
- Brzustowicz L. M, Lehner T, Castilla L. H, et al. (1990) Genetic mapping of chronic childhood-onset spinal muscular atrophy to chromosome 5q11.2-13.3. *Nature* 344 540-541.
- Butchbach M. E. R (2016) Copy Number Variations in the Survival Motor Neuron Genes: Implications for Spinal Muscular Atrophy and Other Neurodegenerative Diseases. *Frontiers in Molecular Bioscience* 3(7).
- Cartegni L, Krainer A. R. (2002) Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nature Genetics* 30 377-384.
- Clinicaltrials.gov - (<https://clinicaltrials.gov/ct2/show/record/NCT02913482>).
- Crawford T. L, Pardo C. A. The Neurobiology of Childhood Spinal Muscular Atrophy. (1996) *Neurobiology of Disease*, 3, 97-110.
- FDA prescribing information – Risdiplam ([https://www.accessdata.fda.gov/drugsatfda\\_docs/label/2020/213535s000lbl.pdf](https://www.accessdata.fda.gov/drugsatfda_docs/label/2020/213535s000lbl.pdf)) Accessed March 2021.
- Fischer U, Englbrecht C. et al. (2011) Biogenesis of spliceosomal small nuclear ribonucleoproteins. *Wiley Interdiscip Rev RNA* 2:718-731.
- Gabanella F, Butchback M. E. R et al. (2007) Ribonucleoprotein Assembly Defects Correlate with Spinal Muscular Atrophy Severity and Preferentially Affect a Subset

of Spliceosomal snRNPs. (2007) *PLoS One* 9: e921.

Jodelka F. M, Ebert A. D et al. (2010) A feedback loop regulates splicing of the spinal muscular atrophy-modifying gene, SMN2. *Human Molecular Genetics* 19(24): 4906-4917.

Kashima T, Manley J. L. (2003) A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. *Nature Genetics* 34 460-463.

Kletz H, Marquet A. et al.(2019) The oral splicing modifier RG7800 increases full length survival of motor neuron 2 mRNA and survival of motor neuron protein: Results from trials in healthy adults and patients with spinal muscular atrophy. *Neuromuscular Disorders* 29:21-29.

Kolb S. J, Battle D. J, et al.(2007) Molecular functions of the SMN complex. *J Child Neurol.* 22(8): 990-994.

Le T. T, McGovern V. L et al.(2011) Temporal requirement for high SMN expression in SMA mice. *Human Molecular Genetics* 20(18): 3578-3591.

Lefebvre S, Bürglen L et al.(1995) Identification and Characterization of a Spinal Muscular Atrophy-Determining Gene. *Cell* 80: 155-165.

Li D. K, Tisdale S. et al.(2014) SMN control of RNP assembly: from post-transcriptional gene regulation to motor neuron disease. *Semin Cell Dev Biol* 32:22-29.

Lunn M. R, Wang C. H. (2008) Spinal Muscular Atrophy. *The Lancet* 371(9630): 2120-2133

Lutz C. M, Kariya S. et al.(2011) Postsymptomatic restoration of SMN rescues the disease phenotype in a mouse model of severe spinal muscular atrophy. *The Journal of Clinical Investigation* 121(8): 3029-3041.

Markowitz J. A, Tinkle M. B, Fischbeck K. H, (2004) Spinal Muscular Atrophy in the Neonate. *JOGNN* 33(1): 12-20.

Melki J, Abdelhak S, Sheth P et al.(1990) Genes for chronic proximal spinal muscular atrophies map to chromosome 5q. *Nature* 344 767-78.

Naryshkin N. A, Weetall M et al.(2014) SMN2 splicing modifiers improve motor function and longevity in mice with spinal muscular atrophy. *Science* 345(6197): 688-693.

Ogino S, Gao S et al.(2003) Inverse correlation between SMN1 and SMN2 copy numbers: evidence for gene conversion from SMN2 to SMN1. *European Journal of Human Genetics* 11: 275-277.

Palacino J, Swalley S. E et al.(2015) SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice. *Nat. Chem. Biol.* 11: 511-517.

Pellizzoni L. (2007) Chaperoning ribonucleoprotein biogenesis in health and disease. *EMBO Rep.* 8: 340-345.

Poirier A, Weetall M et al.(2018) Risdiplam distributes and increases SMN protein in both the central nervous system and peripheral organs. *Pharmacol. Res. Perspet.* E00447.

Prior T. W, Swoboda K. J, Denman Scott H, Hejmanowski A. Q, (2004) Homozygous SMN1 Deletions in Unaffected Family Members and Modification of the Phenotype by SMN2. *Am J Med Genet A* 0 (3): 307-310.

Rao V. K, Kapp D. et al. (2018) Gene therapy for spinal muscular atrophy: an emerging treatment option for a devastating disease. *J Manag Care Spec Pharm.* 24(12-a Suppl):S3-S16.

Ratni H, Ebeling M et al. (2018) Discovery of Risdiplam, a Selective Survival of Motor Neuron-2 (SMN2) Gene Splicing Modifier for the Treatment of Spinal Muscular Atrophy (SMA). *Journal of Medicinal Chemistry* 61: 6501-6517.

Ruggiu M, McGovern V. L et al. (2012) A role for SMN exon 7 splicing in the selective vulnerability of motor neurons in spinal muscular atrophy. *Mol Cell Biol.* 32: 126-138.

Seabrook TJ, et al. P.064 FIREFISH Part 1: 1-year results on motor function in infants with Type 1 spinal muscular atrophy (SMA) receiving risdiplam (RG7916). *Canadian Journal of Neurological Sciences / Journal Canadien des Sciences Neurologiques* 2019;46(s1):S31-S31. doi: 10.1017/cjn.2019.164 [published Online First: 2019/06/05].

Sergott R. C, Amorelli G. M. et al. (2021) Risdiplam treatment has not led to retinal toxicity in patients with spinal muscular atrophy. *Annals of Clinical and Translational Neurology* 8(1): 54-65.

Sivaramakrishnan M, McCarthy K. D et al. (2017) Binding to SMN2 pre-mRNA-protein complex elicits specificity for small molecule splicing modifiers. *Nature Communications* 8(1476).

Stabley D. L, Harris A. W et al. (2015) SMN1 and SMN2 copy numbers in cell lines derived from patients with spinal muscular atrophy as measured by array digital PCR. *Molecular Genetics Genomic Medicine* 3(4): 248-257.

Stabley D. L, Holbrook J et al. (2021) Detection of SMN1 to SMN2 gene conversion events and partial SMN1 gene deletions using array digital PCR. *Neurogenetics*.

Stuart A. (2021) The Nuts and Bolts of Fundus Autofluorescence Imaging. *American Academy of Ophthalmology* <https://www.aao.org/eyenet/article/nuts-bolts-of-fundus-autofluorescence-imaging>.

Sturm S, Günther A et al. (2019) A phase 1 healthy male volunteer single escalating dose study of the pharmacokinetics and pharmacodynamics of Risdiplam (RG7916, RO7034067), a SMN2 splicing modifier. *British Journal of Clinical Pharmacology* 85: 181-193.

Tisdale S, Lotti F et al. (2013) SMN is essential for the biogenesis of U7 snRNP and 3'-end formation of histone mRNAs. *Cell Reports* 5(5): 1187-1195.

Tisdale S, Pellizzoni L. (2015) Disease Mechanisms and Therapeutic Approaches in Spinal Muscular Atrophy. *The Journal of Neuroscience* 35(23): 8691-8700.

Wirth B. (2021) Spinal Muscular Atrophy: In the Challenge Lies a Solution. *Trends in Neurosciences* (Article in press).

# From the Condemnation to the Revival of the Feminine: The Transmission and Reception of the Phaedra Myth in Tsvetaeva's Drama

*Katherine New\**  
*Faculty of Medieval and Modern Languages*

*'... A binary division between mind  
and body spirit and matter.  
Women's bodies have been  
consigned to the dual domains  
of the maternal or the sexual,  
venerated on the one hand,  
damned on the other.'*<sup>1</sup>

## **Introduction**

The article is devoted to the analysis of the reception of the Phaedra myth in Tsvetaeva's drama through the representation of gender in donor traditions (French, Latin and Ancient Greek). Although the study of Tsvetaeva's Classical and European literary sources has been of interest to scholars in recent years, especially Roman Voitekhovits<sup>2</sup> and Zara Torlone<sup>3</sup>, the study of gender representation in Tsvetaeva's

---

\*katherine.new@new.ox.ac.uk

<sup>1</sup>A.M. Barker, J.M. Gheith ed., *A History of Women's writing in Russia*, Cambridge University Press (2002), p.9.

<sup>2</sup>Р.С. Войтехович Античные мотивы в творчестве Марины Цветаевой. Тарту 2007; Цветаева и античность. Москва - Тарту 2008; *eiusdem* Психея в творчестве М. Цветаевой: Эволюция образа и сюжета. Тарту, 2005. (Dissertationes Philologiae Slavicae Universitatis Tartuensis. 15.); *eiusdem* Ариадна Марины Цветаевой: Поэтика античности // Северный сборник: Proceedings of the №rFA Network in Russian Literature 1995–2000. Stockholm Studies in Russian Literature. 34. Stockholm, 2000. С. 232–242.

<sup>3</sup>Z. Torlone, "Voicing Passions: Phaedra of Euripides and Marina Tsvetaeva," ASEES, November 2008. Washington, DC; Z. Torlone, *Soul and Passion: Marina Tsvetaeva's Classical Plays. Introduction, Transla-*

Neoclassical drama has not attracted sufficient attention.<sup>4</sup> A new approach to the study of gender representations in *Fedra* will be suggested in this article by introducing three planes through which Tsvetaeva, and the traditions she draws upon, reveal the distribution and depiction of female and male roles in different cultures. The extent of condemnation or absolution allotted to female and male characters is analysed in this article on the extra-textual plane (on which the viewpoint of the author is manifested) and the intra-textual plane, revealed on two levels: the macrocosmic plane of divine motivation behind character's actions and the microcosmic plane of hero's psychology (the female body/mind set against the male). This new approach will ensure a more precise differentiation between the interrelation of the feminine body and mind, and the understanding of the psychology of male and female characters, elucidating the reception and transmission of the Phaedra myth.

Marina Tsvetaeva was the first woman to rework the Phaedra legend;<sup>5</sup> before her it existed entirely in the domain of male authors and was treated by Euripides, Apollodorus, Seneca, Racine, Swinburne, A'Annunzio, de Unamuno and O'Neill. The legend of Phaedra, wife of Theseus and queen of Athens, is one of the most well-known myths of the classical world. The first extant mention of Phaedra's name is made in the *Odyssey*, when the eponymous hero descends into the Underworld to question the prophet Tiresias (11.321-22).<sup>6</sup> The initial encounter with Phaedra is limited to three single words ("I saw... Phaedra", "Φαίδρην τε... ἴδον") of which one, the verb, is used to govern two other female names alongside hers ("ἴδον") and the other ('τε') is a polysyndetic particle: she is defined as nothing more than one of "the heroes' wives, whom [Odysseus] saw" (330). From this single reference it is hard to deduce whether Phaedra is referred to in the Homeric tradition as a famous woman in her own right or is reduced to a mere contextualising element, the named wife of a famous hero.

The schematic plot of Phaedra's fate is known from the mythological handbook ascribed to Apollodorus. In Apollodorus's *Epitome* Phaedra was the daughter of Minos and Pasiphae, sister of Ariadne, wife of Theseus, mother of two sons and step-mother of Hippolytus; she fell in love with her step-son; was rejected by him; accused him of rape and committed suicide (*Ep.* 1.18-19). In Apollodorus's account Phaedra is presented as a passive, one-dimensional character, whose name has survived primar-

---

*tion and Commentary* (co-authored with Maria S. Fox, published for educational purposes). Create Space (2011); Z. Torlone, "Muted Voices: Marina Tsvetaeva's Classical Heroines" In *Brill's Companion to Classical Reception and Modern World Poetry* (ed. Polina Tambakaki) 2019.

<sup>4</sup>The only exception is Catriona Kelly's discussion of gender and sexuality in Tsvetaeva's poetry in Kelly, C., *A History of Russian Women's Writing 1820-1992*, Oxford: Clarendon, 1998, Pp. 301-319.

<sup>5</sup>R. Lauriola and K. N. Demetriou ed. *Brill's Companion to the Reception of Euripides*, v.3, Brill, 2015, P.461.

<sup>6</sup>R. Lauriola, K. N. Demetriou, P.444.



ily because of her connection to two males, Theseus and Hippolytus. Nothing but her body, as defined in the epigraph by the ‘dual domains of the maternal or the sexual’ comes into play (she is a wife, mother and illicit lover); her “mind” so far carries no interest to the patriarchal society of ancient Greece.<sup>7</sup> Despite the static representation of Phaedra in mythology, her legacy, as is the case for many women of antiquity, is negative.<sup>8</sup> According to Seneca she is “a much, much greater evil than [Medea],”<sup>9</sup> she is the most nefarious of all women and the epitome of an evil step-mother.<sup>10</sup> Her crime, initially purely sexual, is converted into a grave moral flaw and is made to gradually expand into a failure not only of her body but also of her mind. This transformation of ‘body’ into ‘mind and body’ is implicitly invited by the absence of a predefined psychological dimension derived from mythology, in which Phaedra is acting but not suffering. Tsvetaeva, as each subsequent recipient of the myth, is thus left free to express her attitude to Phaedra and present her own rendition of the legend.

### Part I. Extra-textual plane

The study of gender representation in the Phaedra narrative is best begun by deploying the first of the three planes by which the viewpoint of the poet is manifested in other writings external to the textual evidence of the play. The extra-textual plane can shed light on Tsvetaeva’s representation of gender through epistolary accounts, drafts and letters to friends. In one of her letters to Yuri Ivask dated April 1933 Tsvetaeva divulges the source of her play: Gustav Schwab’s *Sagen des klassischen Altertums*.<sup>11</sup> Some scholars, taking her ‘confession’ at face value, amplify Tsvetaeva’s reliance on a ‘didactic, moralising and frequently bowdlerized version of the Greek myths’<sup>12</sup> presented by Gustav Schwab and, without attempting to investigate the veracity of Tsvetaeva’s assertion, use it to explain away ‘her more startling departures from the traditional stories of... Phaedra’.<sup>13</sup> This assumption has, however, been proved erroneous by Michael Makin, who, in presenting numerous excerpts from Tsvetaeva’s correspondence and diaries, demonstrates her intimate knowledge

<sup>7</sup>A. M. Barker, J.M. Gheith, P.9.

<sup>8</sup>The names of virtuous women such as Andromache, Penelope, Alcestis, Iphigenia, Polyxena, Hypermenestra etc. are few compared to those of wicked women such as Medea, Helen, Ariadne, Agave, Jocasta, Arachne, Clytemnestra, Niobe, Pasiphae, Sthenoboea etc.

<sup>9</sup>Colchide noverca maius haec, maius malum est.’ (697).

<sup>10</sup>R. Lauriola, K. N. Demetrious, P.445.

<sup>11</sup>M. Makin, *Marina Tsvetaeva: Poetics of Appropriation*, Clarendon Press Oxford, (1993), P.269.

<sup>12</sup>S. Karlinsky, *Marina Cvetaeva: Her Life and Her Art*, University of California Press, (1966), P.259.

<sup>13</sup>S. Karlinsky, P.259; M. Makin, P.269.

of European and Classical literature.<sup>14</sup> Tsvetaeva's acquaintance with European literature and her excellent knowledge of major European languages can be shown by referring to a few well-known facts of her biography. In a single year (1902) Marina Tsvetaeva, then ten years old, went to three separate schools, one in Italy (Nervi near Genoa), one in Switzerland (Lausanne) and, since her father was concerned for her German (both her spoken and written French were native<sup>15</sup>), one in Freiburg (Germany).<sup>16</sup> At the age of sixteen Marina Tsvetaeva travelled to Paris in order to attend a lecture course on Medieval French literature and attained such proficiency in the language as to be able, in her later years, to translate into French Pushkin, Mayakovsky and Russian folk poetry.<sup>17</sup> In addition to this, her father, Ivan Tsvetaev, was a well-known Classicist<sup>18</sup> and the author of a dissertation on the textological history of Tacitus' *De origine et situ Germanorum*, in which he compared different manuscript traditions.<sup>19</sup> He specialised in teaching Latin and Greek at university and founded the first museum of ancient Greek casts in Russia (which Marina Tsvetaeva called her younger brother).<sup>20</sup> As a Classical scholar specialising in Latin epigraphy, Tsvetaev had a formidable library which was available to his daughter.<sup>21</sup> As was stated by a well-known specialist in Tsvetaev's family, a close friend of Anastasia Tsvetaeva and

<sup>14</sup>To quote but a few: 'One poem *Pravda* has an epigram from Juvenal (I;46). The first scene of *Prikluchenie* is based on a scene from Apuleius' story of Psyche and Cupid. Tsvetaeva also quotes Ovid's version of the story of the Cumaean Sibyl in a letter to Bakhrakh in 1923. Later in that year in letters to Bakhrakh she claims: 'Сейчас лягу и буду читать Троянскую войну. Никого не могу читать, кроме греков. У меня огромный немецкий том: там всё.' and 'Я редко бываю в городе, только в библиотеке, где читаю древних.' and asks Bakhrakh to send her Nietzsche's 'The Birth of Tragedy' (Pis'ma Mariny Tsvetaevoi, ed. Bakhrakh, 2nd instalment, 322, 335, 337 (letters of August and September 1923) etc. M. Makin, P.269.

<sup>15</sup>За Марусю даже страшно: говорит, как взрослый француз, изящным, прямо литературным языком... пишет по-русски правильное и литературное пяти- и шестиклассников в гимназиях...': Письмо И. В. Цветаева к Л. А. Иловайской (26.06.1903)

<sup>16</sup>Каган Ю. М. И.В. Цветаев. Жизнь. Деятельность. Личность. Москва, 1987. С. 144-149.

<sup>17</sup>W. Coudenys, «Те последние вы можете исправлять с точки зрения стилистики»: четыре забытых французских перевода Марины Цветаевой, *Revue des études slaves* (1994) V. 66, pp. 401-411.

<sup>18</sup>«4 бюста вывезены мною в 1876 году из Мюнхена (евс, Афина, Диана Версальская и Ватикан, амур). Тогда я был всецело занят заготовлением атласа facsimiles осских надписей и писанием докторской диссертации. Я купил их на сбережения от уроков, которые я давал по греческому языку в доме графа Бобринского» in Каган Ю. М. (1987.) С.152.

<sup>19</sup>Каган Ю. М. И.В. Цветаев. Жизнь. Деятельность. Личность. Москва, 1987. С.24-25

<sup>20</sup>Каган's book on Ivan Tsvetaev is dedicated to his role as founder of the Museum of Fine Arts. (Каган Ю. М. И.В. Цветаев. Жизнь. Деятельность. Личность. Москва, 1987).

<sup>21</sup>Tsvetaeva describes her father's library in one of her poems: Как переполненные соты — / Ряд книжных полок. / — Тронул блик / Пергаментные переплеты / Старинных книг. / Цвет Греции и слава Рима, — / Неисчислимые тома! / Здесь — сколько б солнца

a Classicist herself, Judith Kagan, 'anybody studying Classics in Marina Tsvetaeva's poetry cannot discard the professional interests of her father'.<sup>22</sup>

Tsvetaeva's acquaintance with ancient sources is further confirmed by one of her own letters dated to 1923, in which she manifests not just an abstract awareness of Euripides' *Hippolytus* (Innokenty Annensky published a readily available translation of the play in 1902)<sup>23</sup> but an exact knowledge of his approach, i.e. the role given to the Nurse in revealing Phaedra's secret.<sup>24</sup> Tsvetaeva's knowledge of her literary predecessors is further suggested by her awareness of the shift of attention from the male character in Euripides' play, manifested by his title *Hippolytus*, to the female, as initiated by Seneca in his tragedy *Phaedra*. Tsvetaeva must have known Seneca's play since it is the only tragedy of antiquity to begin, like her *Fedra*, with a Choral song about hunting.<sup>25</sup> It is difficult to determine whether Tsvetaeva's title goes back to Seneca's tragedy or to its French adaptation by Racine, but it is known that she very much admired Sarah Bernhardt, the actress who performed the role of Phaedra in Racine's play *Phèdre*.<sup>26</sup> However, the most likely explanation of Tsvetaeva's choice of title is her interest in the fate of the female character, despised and condemned by the earlier Classical tradition.

Consequently, the question arises why, at the expense of projecting herself as less educated than she really was, does Tsvetaeva deliberately conceal her knowledge of earlier renditions of the Phaedra legend. Some suggest that 'Tsvetaeva often presented herself as a somewhat uneducated poet in deliberate contrast to others such as Innokenty Annensky and Vyacheslav Ivanov'<sup>27</sup>. Others elucidate the problems faced by women in literary circles in terms of an 'anxiety of authorship, experienced by many Russian women writers, who have been made to feel that it is 'presumptuous' to take up the pen, and that intellectual women, especially women writers, are nothing but a freak, a 'crocodile in flannel or a dancing monkey' in Russian society'<sup>28</sup>. An alternative answer, however, can be proposed, namely that Tsvetaeva deliberately wanted to divorce herself from earlier renditions, so as to have more freedom to

---

ни внесли мы — / Всегда зима. / Последним солнцем розовея, / Распахнутый лежит Платон... / Бюст Аполлона — план Музея - / И все.

<sup>22</sup>Тот, кто станет изучать античность в творчестве Марины Цветаевой, не сможет пройти мимо профессиональной увлеченности ее отца.' in Kagan Ю. М. (1987.) С.152.

<sup>23</sup>R. D. B. Thomson, P.343.

<sup>24</sup>'У Эврипида все выдает кормилица, без воли Федры. Ложь. — Кормилица только передает' in M. Makin, *ibid.*, (1993), P.289.

<sup>25</sup>R. D. B. Thomson, P.343.

<sup>26</sup>R. D. B. Thomson, P.343.

<sup>27</sup>M. Stadter Fox, *The Troubling Play of Gender: The Phaedra Dramas of Tsvetaeva, Yourcenar, and H.D.*, Associated University Presses, (2001), P.21.

<sup>28</sup>R. Marsh ed., *Women and Russian Culture: Projections and Self-Perceptions*, Berghahn Books, (1998), P.7.

portray a new, nobler kind of Phaedra. This view is supported by one of Tsvetaeva's drafts, in which she states that her goal was to present Phaedra as a woman of moral integrity, who by her own virtue was free from criminal intent: "дать Федру, не Медею, вне преступления, дать безумно любящую женщину, глубоко понятную".<sup>29</sup> Thus, by concealing her knowledge of primary sources, Tsvetaeva underlined that her interest in retelling the Phaedra legend lay not in polemicalising with early renditions, but in stepping away from them and presenting a completely new character, a tragic and, as shown below, completely blameless, female.

Unlike Tsvetaeva, who chooses to conceal her knowledge of her literary predecessors, the author of her potential French source, Racine highlights his proximity to both Euripides<sup>30</sup> and the mythological tradition<sup>31</sup>. By underlining his reliance on earlier traditions so prominently, Racine implicitly draws attention to his own innovations and their overall effect on his representation of the Phaedra myth. At the beginning of his play Racine offers his readers a preface, in which he outlines his approach to the Phaedra legend: to make 'Phaedra neither completely guilty nor completely innocent' (*Phèdre n'est ni tout à fait coupable, ni tout à fait innocente.*). The playwright inverts the balance between Phaedra as the negative and Hippolytus as the positive character, 'remaking them to a human measure'.<sup>32</sup> By mollifying the stark contrast between Phaedra's corruption and Hippolytus' purity, an integral element of both Euripides' and Seneca's renditions, Racine makes the traditional binary opposition between the two characters weaker and consequently softens the negative representation of Phaedra. At an extra-textual level, Racine offers his readers an elucidation of his understanding of the Phaedra myth, directing them from the very start of the play towards a more forgiving view of Phaedra and her fateful passion.

Unlike Tsvetaeva and Racine, Euripides and Seneca, by merit of their antiquity, have left no external accounts of how their characters are to be interpreted, instead leaving their audiences to deduce their views of gender relations from the texts of the plays themselves (the two internal macrocosmic and microcosmic levels) or from external sources (the extra-textual plane). Historiographical accounts, which can be

---

<sup>29</sup>In this quotation Tsvetaeva enters into polemics with Seneca, who, as seen below, also compared Medea to Phaedra but arrived to the opposite conclusion from Tsvetaeva.

<sup>30</sup>'Here is another tragedy whose subject was treated before by Euripides', 'Voici encore une tragédie dont le sujet est pris d'Euripide'.

<sup>31</sup>'I refer to these sources because I wish to adhere very carefully to the myth', 'Je rapporte ces autorités, parce que je me suis très scrupuleusement attaché à suivre la fable.'

<sup>32</sup>I even tried to make [the character of Phaedra] a little less hateful than in the tragedies of ancient poets' ('J'ai même pris soin de la rendre un peu moins odieuse qu'elle n'est dans les tragédies des Anciens') and 'I thought I should give [the character of Hippolytus] some weakness, which would make him a little more blame-worthy in the eyes of his father' ('J'ai cru lui devoir donner quelque faiblesse qui le rendrait un peu coupable envers son père.). In R. Lauriola, K. N. Demetriou, P.461.

viewed as an extra-textual background to the Greek and Latin literary representations of gender, reveal that the societies of ancient Greece and Rome were constructed on patriarchy, excluding women from public life. Apart from the role of priestess (particularly of Bacchus in Greece and of Vesta in Rome), women were consigned to the private domains of their own quarters or gynaeceums<sup>33</sup> and were assigned the functions of minding the house and bearing children.<sup>34</sup> Both these roles, although performed from an internal domain, were of paramount importance to the preservation of the external patriarchal society: by tending her husband's house and ensuring that all its inmates (including herself), were acting in accordance to expectations, a woman was in effect preserving a man's honour. By slacking in her duties, a wife could potentially ruin her husband's reputation and consequently his public role in society.<sup>35</sup>

The importance of the second female role, that of bearing children, is obvious; in addition to ensuring the continuation of lineage it provided a new influx of soldiers,<sup>36</sup> crucially important to an ancient society with either a defensive or expansive policy. Since women were allotted the binary function of both preserving and potentially destroying a male-orientated society, the plays of Euripides and Seneca gain an additional dimension: Phaedra's behaviour is damaging not only to herself and Hippolytus but also to the entire polis in which she reigns. Thus, at least to some degree both Euripides and Seneca must unavoidably condemn Phaedra, as a woman who has failed in both her 'venerated' roles of wife and mother and put herself in the position of a potential threat to Greece's principal city, Athens.<sup>37</sup>

## Part II. Intra-textual plane

### Section 1. Macrocosmic level

The representation of gender in Tsvetaeva's *Phaedra* can be revealed with the help of the second, intra-textual plane, exploring divine motivation in exonerating or condemning the female (as opposed to male) protagonist. Tsvetaeva's play contains allusion to divine motivation but does not feature gods as physical entities, synthesizing the approach to divine intervention of her literary predecessors. The two traditions, Tsvetaeva draws upon, go back to Euripides' *Hippolytus*, which contains active divine intervention in the form of speaking gods, and Seneca's *Phaedra* and

<sup>33</sup>R. D. B. Thomson, 'Tsvetaeva's Play "Fedra": An Interpretation', *The Slavonic and East European Review*, Vol. 67, No. 3 (1989), P.344.

<sup>34</sup>N. Sorokin Rabinowitz, *Anxiety Veiled: Euripides and the Traffic in Women*, Cornell University Press (1993), P.174.

<sup>35</sup>'mak[ing] and unmak[ing] culture', N. Sorokin Rabinowitz, P.158-9.

<sup>36</sup>N. Sorokin Rabinowitz, P.174.

<sup>37</sup>M. Barker, J.M. Gheith, P.9.

Racine's *Phèdre*, which remove divine action from the stage entirely, containing only metonymical references to the gods.

In Tsvetaeva's *Phaedra* the function of the gods as motivators of human folly is to some extent indebted to Euripides' play, which opens with an introductory monologue of 57 lines by Aphrodite. In this monologue the goddess clearly states her part in the action<sup>38</sup> and outlines, in the minutest detail, the manner in which she will execute her vengeance. She looks back to the past when she made Phaedra behold Hippolytus for the first time<sup>39</sup> then to the present, when Phaedra is already sick with her passion,<sup>40</sup> and finally to the future when Hippolytus and Phaedra will be dead and their fate handed to posterity.<sup>41</sup> The three temporal spheres (past, present and future), through which Aphrodite traces her plan demonstrates the completeness of her control over Phaedra's and Hippolytus's fate. Although Aphrodite's overarching plan may seem to exonerate Phaedra fully of her crime,<sup>42</sup> one crucial detail is absent from the goddess' speech, a detail which conditions the extent of Phaedra's ultimate condemnation. Aphrodite accepts the responsibility for making Phaedra fall in love with Hippolytus and for making Theseus curse him, but leaves out the reason why Theseus is so enraged with his son. Neither does she acknowledge that Phaedra's accusation of Hippolytus is her doing. Phaedra herself must thus be held responsible for the shameful lie. The appearance on stage of a second divinity, Artemis, confirms this, since she states that Phaedra herself thought up the false accusation in order to save her reputation.<sup>43</sup> By reaffirming Aphrodite's role in the tragic events of the play and underlining that no god, let alone mortal, could stop her<sup>44</sup>, Artemis highlights

<sup>38</sup>Yet for his sins against me I shall punish Hippolytus this day,' ἄ δ' εἰς ἔμ' ἡμάρτηκε τιμωρήσομαι. Ἰππόλυτον ἐν τῆδ' ἡμέρᾳ': (21-22). All translations are taken from LOEB editions.

<sup>39</sup>'his father's high-born wife Phaedra saw him, and her heart was seized with a dreadful longing by my design.' (πατρός εὐγενῆς δάμαριδούσα Φαίδρα καρδίαν κατέσχετο,' 26-27).

<sup>40</sup>'the poor woman, groaning and struck senseless by the goad of love, means to die in silence, and none of her household knows of her malady.' (ἐνταῦθα δὴ στένουσα κάκπεπληγμένη κέντροις ἔρωτος ἡ τάλαν' ἀπόλλυται σιγῆ, ξύνοιδε δ' οὔτις οἰκετῶν νόσον,' 38-40).

<sup>41</sup>I shall reveal the matter to Theseus and it will come to light,<sup>2</sup> and the young man who wars against me shall be killed by his father with the curses of the sea-lord' δεῖξω δὲ Θησεῖ πρᾶγμα κάκφρανήσεται. καὶ τὸν μὲν ἡμῖν πολέμιον νεανιάνκτενεῖ πατῆρ ἀράσιν ἄς ὁ πόντιος ἀναξ Ποσειδῶν ὤπασεν Θησεῖ γέρας,' 42-45); 'But Phaedra, noble though she is, shall nonetheless die. I do not set such store by her misfortune as to let my enemies off from [50] such penalty as will satisfy my heart.' ('ἡ δ' εὐκλεῆς μὲν ἀλλ' ὅμως ἀπόλλυται Φαίδρα: τὸ γὰρ τῆσδ' οὐ προτιμῆσω κακόν τὸ μὴ οὐ παρασχεῖν τοὺς ἐμοὺς ἐχθροὺς ἐμοὶ δίκην τοσαύτην ὥστ' ἐμοὶ καλῶς ἔχειν,' 46-50)

<sup>42</sup>A. Powell ed., *Euripides, Women and Sexuality*, Routledge, (1990), P.47.

<sup>43</sup>'And Phaedra, fearing lest she be put to the proof, wrote her false letter and destroyed your son by deceit, and though it was a lie, she persuaded you., 'ἡ δ' εἰς ἔλεγχον μὴ πέση φοβουμένη ψευδεῖς γραφὰς ἔγραψε καὶ διώλεσε νδόλοισι σὸν παῖδ', ἀλλ' ὅμως ἐπεισέ σε.' (1310-1312).

<sup>44</sup>'Aphrodite willed that things should happen thus, sating her anger. Among the gods the custom is this: no god contrives to cross the will of another, but we all stand aside,' πληροῦσα θυμόν. θεοῖσι δ'

Phaedra's powerlessness to resist her love. Thus, as compared to the mythological and Homeric traditions which make no mention of divine participation in Phaedra's fate, Euripides executes a partial rehabilitation of Phaedra: her love and death are given clear divine motivation, but her shameful accusation against Hippolytus is not. She cannot be viewed entirely as a tool<sup>45</sup> or puppet<sup>46</sup> of the gods and yet cannot be held fully responsible for her criminal passion: divine motivation can be used to exonerate her 'body' but not her 'mind'.<sup>47</sup>

Since explicit divine motivation is absent from Tsvetaeva's other potential sources (Seneca's *Phaedra* and Racine's *Phèdre*), it can be suggested that the poet owes her allusion to [perhaps 'conception of?'] the gods' intervention to Euripides, although unlike him she does not give them active, speaking roles. From the beginning of her play a clear-cut tension is set up between the same two goddesses, Artemis and Aphrodite, which feature so prominently in Euripides. Hippolytus serves the former ('Αρτεμίδε σλужу. Α τυ?') and Phaedra the latter ('Αφροδίτε σлужу'). Although this opposition is steadily but subtly kept up throughout the play, with Hippolytus hating Aphrodite as much as he does in Euripides ('хулителъ Афродитин'<sup>48</sup>) and Phaedra still performing the function of the woman in whom the goddess has inspired seductive charms, expressions of divine motivation are rare. The Nurse gives one when she calls her mistress 'Aphrodite's slave' ('Αφροδίτιной рабы.') and Theseus the other when he proclaims hatred for 'персты / Афродитиной!', implicitly accusing Aphrodite of the present tragedy. The only real confirmation that, as in Euripides' play, Aphrodite is responsible for Phaedra's death is delayed until the very end of the play, after Hippolytus has been killed and the queen's secret has been revealed. Since in the course of the play Phaedra had uttered no false accusation against Hippolytus but was guilty only of love (now revealed to have been inspired by Aphrodite), she is fully exonerated.

Tsvetaeva keeps this divine motivation behind Phaedra's actions back till the closing scene, allowing almost the whole play to unfold with the audience believing Phaedra to be responsible for her love (Tsvetaeva idealises Hippolytus's beauty and courage making his own merits responsible for the queen's illicit passion).<sup>49</sup> This innovation signals that Tsvetaeva's interest lay less in pointing a finger of blame at

ὧδ' ἔχει νόμος: οὐδεὶς ἀπαντᾶν βούλεται προθυμίᾳ 1330 τῆ τοῦ θέλοντος, ἀλλ' ἀφιστάμεσθ' αἰεί.', (1328-1330).

<sup>45</sup>N. Sorokin Rabinowitz, P.167.

<sup>46</sup>R. Lauriola, K. N. Demetrious, P.453.

<sup>47</sup>A. M. Barker, J.M. Gheith, P.9.

<sup>48</sup>R. Lauriola, K. N. Demetrious, P.453.

<sup>49</sup>'Ипполита оленьегокого / С ртом негоупругим, / С ртом - луком неломким! /... / Дивен слух чей, чуден взгляд чей. / Легче скока никто не имывал - / Ипполита необгонимого'.

one individual or another, than in analysing the tragedy of a loving woman, rejected and forced to die of grief for an ideal youth. Because Phaedra's body is cleared of all blame only at the end of the play, she is able, with no god acting as intermediary, to convince the audience of her innocence by the purity and power of her mind. It would not be an exaggeration to state that in Tsvetaeva's play the female body is represented as being fully conquered by her mind.

In contrast to Tsvetaeva's play in which divine intervention is made explicit through references to gods by other characters, in Seneca's *Phaedra* and Racine's *Phèdre*, divine action is removed entirely, only leaving room for the gods, Venus and Diana in particular, to be addressed with abstract apostrophes or envisaged as metonymies for love and chastity. In Seneca's 1280-line play the name of Venus appears only five times, Diana's once, and only twice in the whole play is any divine being blamed for the tragic events on stage. On one occasion Phaedra tries to attribute some responsibility for her love to Venus,<sup>50</sup> but the Nurse hastily brushes it off, asserting that Venus is nothing more than a phantom.<sup>51</sup> The second time Phaedra claims that she is cursed by some divinity with a 'fatal evil' ('fatale... malum', 113), which she cannot resist.<sup>52</sup> Since no active participation of the divinity or even its name is encountered in the play, Phaedra's claim has been regarded as insincere and a mere mask behind which she tries to hide her own morally culpable passion.<sup>53</sup>

In Racine's play Venus has much the same function as in Seneca's version, since although she is regularly invoked by Phaedra, she is not explicitly shown as a direct motivator of the queen's passion. Rather than an active deity, Venus is regarded more as metonymy for love: 'Implacable Venus, am I sufficiently in your thrall?' (814) and 'Of what brave men has Venus not been conqueror' (123); or as an evil force inspiring criminal sexual passion: 'O Venus' hatred! O fatal anger! To what distraction did love not drive my mother!'. Racine's reluctance to use the deity as an active, thinking presence in his play can be accounted for two ways; firstly, the French playwright was a devout Christian and secondly, he, like Seneca, wanted to make Phaedra solely responsible for her immoral actions. By removing any divine motivation for Phaedra's actions, both Seneca and Racine make it impossible for her name to be cleared at the end of the play – for posterity she remains the sole perpetrator of her crimes. Thus, at the macrocosmic level, Seneca's and Racine's rendition of the Phaedra legend are

<sup>50</sup>Venus, detesting the offspring of the hated Sun, is avenging through us the chains that bound her to her loved Mars', 'stirpem perosa Solis inveni Venus per nos catenas vindicat Martis sui suasque, probris omne Phoebeum genusonerat nefandis:' (124-127).

<sup>51</sup>'Love-mad souls adopted these vain conceits and have feigned Venus' divinity and a god's archery., 'vana ista demens animus acscivit sibi Venerisque numen finxit atque arcus dei.' (202-203).

<sup>52</sup>A. S. Gérard, *The Phaedra Syndrome: Of Shame and Guilt in Drama*, Amsterdam – Atlanta, GA (1993), P.21.

<sup>53</sup>A. S. Gérard, P.21.



unequivocally more damning than their predecessor Euripides or literary successor Tsvetaeva.

### *Section 2. Microcosmic level*

The last of the three planes through which each individual poet reveals his understanding of the Phaedra legend can be termed the microcosmic, since it is concerned with the psychology of individual characters within the text. Although, in contrast to the epic and mythological tradition, all four dramatic renditions of the legend endow Phaedra with a complex psychology, this new dimension is given a different impetus by each poet.

In each of the plays, the characterisation of Phaedra falls into four separate stages, each serving to illuminate a different aspect of her character. The first stage occurs when the queen comes on for the first time and is presented to us via her opening speech. Here we are able to glimpse her internal psychology, the way in which she herself, as yet without the external influence of her Nurse, copes with her illicit passion, repressing or expressing it, depending on the degree of moral integrity with which she is endowed.<sup>54</sup> The second stage occurs when the queen enters into dialogue with her Nurse and when she ultimately reveals her love for Hippolytus. The relative speed with which she does so and her response to the reaction of her Nurse both serve to represent the level of her internal strength and ability to withstand pressure. The third stage occurs when the passion of Phaedra is revealed to Hippolytus, and when she, rejected, decides upon a course of action. This stage is the most crucial of the four, in that it introduces a male character with whom the female Phaedra (and indeed her Nurse) is compared and contrasted. Whereas the previous two stages of her characterisation revealed her internal 'mind', this stage allows each poet to determine the external impact she has on the male characters around her. Finally, the fourth and most problematic stage occurs at the end of the play when Phaedra is either dead (as in Euripides and Tsvetaeva) or planning her suicide (as in Seneca and Racine). At this stage the poet analyses the motivation for her suicide, and can reveal conclusively to the audience how far Phaedra is to be condemned or absolved in the play. The first stage of Phaedra's representation is shown by Tsvetaeva through the queen's initial 'sickness' and her subsequent interchange with the Nurse. Almost from the very first line of the second scene Phaedra utters a prophecy, foreshadowing her own hanging on a myrtle branch and the killing of Hippolytus by his horses. Neither the Chorus nor the Nurse can appreciate Phaedra's prophetic vision and both continue guessing at a physical malady which they believe is afflicting her. Even when Phaedra elaborates her prophecy with 'Трещи, кожа! Теки, сок!', conjuring up

---

<sup>54</sup>R. Lauriola, K. N. Demetrious, P.461.

the image of her splitting skin and the sap trickling out of a broken branch and with 'На суку тяжелый плод. / Бьется плод, гнется сук' an explicit allusion to her convulsing body dangling from a tree, the Nurse remains oblivious to her superior foresight and brusquely dismisses Phaedra's vision as nothing 'А плод каков?... / А скок каков? Да вовсе нет.' Phaedra, the heroine who gives Tsvetaeva's play its name, speaks fewer lines than even her slave the Nurse<sup>55</sup> and participates only in two extended stichomythic dialogues. There is a marked difference in register between her and the Nurse: whereas the colloquial speech of the latter is reminiscent of a 19th century Russian peasant woman<sup>56</sup> and approaches vulgarity,<sup>57</sup> the speech of the former is marked by a higher register, more like the speech of the male characters, who, according to Karlinsky, have an 'archaised diction that suggests the Bible and the Iliad'.<sup>58</sup> By introducing a differentiation in registers, Tsvetaeva raises her Phaedra out of the petty gynaeceum,<sup>59</sup> her traditional female location, and places her on the same level as the male characters: she is able to make her own decisions and act according to them.

Tsvetaeva's innovation in Phaedra's characterisation is made explicit through comparison with Euripides. In the Greek play Phaedra is carried onto the stage, weak and broken, her spiritual sickness concealed behind a physical malady: 'Raise up my body, hold my head erect! My limbs are unstrung.' ('ἀίρετέ μου δέμας, ὀρθοῦτε κάρα: λέλυμαι μελέων σύνδεσμα φίλων.', 198-199). By making Phaedra allude only to a physical weakness and remain silent about her psychological turmoil (which is only shown later, when she reveals her inner conflict to the chorus), Euripides firmly locates her in her socially acceptable female niche: her body is visible to (the entirely male) audience but her mind is repressed.<sup>60</sup>

The whole of the first stage of Phaedra's representation reveals a desire to cast off this repression and free herself from female bonds: she asks the Nurse to carry her out of the traditional, closed female locus and take off the burdensome veil concealing her hair (202).<sup>61</sup> Phaedra proceeds to express an even more disturbing desire to go out of the city and into the mountains, where Hippolytus lives. The colouring which Euripides gives to Phaedra's desire to leave her lawful location is negative: both the Nurse and the Chorus condemn it and encourage Phaedra to return to her

---

<sup>55</sup>R. D. B. Thomson, P.340.

<sup>56</sup>S. Karlinsky, *Marina Tsvetaeva: The woman, her world and her poetry*, Cambridge University Press, (1985), P. 201.

<sup>57</sup>M. Makin, P.292.

<sup>58</sup>S. Karlinsky, P. 201.

<sup>59</sup>R. D. B. Thomson, P.344.

<sup>60</sup>R. Lauriola, K. N. Demetrious, P.447.

<sup>61</sup>N. Sorokin Rabinowitz, P.162

'female' role before she violates the boundaries of propriety and says too much.<sup>62</sup> In Euripides this stage of characterisation reveals a more violent conflict between the queen's repressed mind and her visible body than the one presented by Tsvetaeva: the Russian Phaedra is not silenced and reminded of her female role. She does not yearn to break free from her physical location, and most importantly, unlike Euripides' Phaedra, she is not starving herself to death in an effort to preserve the outward, physical semblance of honour.<sup>63</sup>

In contrast to Tsvetaeva, the French playwright Racine offers a similar first stage of characterisation to that of Euripides. Although Phaedra enters the stage herself, her femininity betrays her: she is weak compared to the young, boisterous Hippolytus who has just left the scene: 'I can't support myself: my strength has left me. / My eyes are dazzled, on seeing the light of day, / My knees, trembling beneath me, have given way' (154-6). Much as in Euripides' Phaedra and unlike Tsvetaeva's queen, who foresees her own punishment, Racine makes her rave about leaving the palace of her husband and retiring to watch a chariot, an explicit reference to Hippolytus, the tamer of horses and lord of chariots.<sup>64</sup> However, as soon as Phaedra comes to her senses, she reports her conscious decision to end her life: 'Oh Sun, I come to look on you for one last time', ('Soleil, je te viens voir pour la dernière fois.', 172). Like the Greek Phaedra, she appears to have accepted what is required of her by society (physical presence, mental absence) and sees no other way out but to destroy her body by the power of her repressed mind.<sup>65</sup>

Of all the four plays studied in this article and in direct contrast to Tsvetaeva, Seneca's *Phaedra* has the most negative first stage representation, due mainly to the fact that his queen is given a more active, 'masculine' role and so robbed of the sympathy elicited by Tsvetaeva's feminine Phaedra. To begin with, Seneca's queen is not physically sick and does not adopt a sitting or reclining position as in the other plays. Instead her sickness is purely spiritual: 'a malady feeds and grows inside me,' ('alitur et crescit malum' (99-102)). Secondly, she is not surrounded by a Chorus of caring women, but is instead accompanied only by her slave (the Nurse) and dominates the stage alone. She gives no speech of weak delirium, which only hints at her passion, but, instead of concealing her love like the Greek, French and Russian Phaedras, casts off customary female propriety and begins her opening speech with a confession. Her

<sup>62</sup>Won't you stop saying such things before the crowd, hurling wild words that are mounted on madness?, 'οὐ μὴ παρ' ὄχλω τάδε γηρύσημανίας ἔποχον ῥίπτουσα λόγον;' (213-214).

<sup>63</sup>N. Sorokin Rabinowitz, P.162.

<sup>64</sup>Gods! Why am I not sitting in that dark forest? When shall I follow the chariot with my eyes, charging nobly on, through the dust that flies?' ('Dieux! que ne suis-je assise à l'ombre des forêts! / Quand pourrai-je, au travers d'une noble poussière, / Suivre de l'oeil un char fuyant dans la carrière?', (176-178)).

<sup>65</sup>N. Sorokin Rabinowitz, P.162.

character radically differs from the other three representations in that she is made the protagonist of the play: Hippolytus and Theseus seem to participate only as foils to illuminate her character rather than as independent individuals.<sup>66</sup> In addition to this, Seneca's Phaedra stands out from the Russian, Greek and French renditions by her independence of mind: in order to set the play in motion she has to persuade the Nurse to approve her passion; she decides to approach Hippolytus herself and confess her passion; she decides to accuse him herself, and is executed without resorting to written means of communication or the intercession of her slave woman.<sup>67</sup> Such a radical departure from the mythological tradition and the earlier version of Euripides, has led scholars to suggest various explanations: some claim that Seneca wanted to portray a more emancipated Phaedra, in order to provide social commentary on the moral corruption of the early Principate;<sup>68</sup> others that he is displaying nostalgia for the old Republican value of *verecundia*;<sup>69</sup> others still that he used Euripides' first version of *Hippolytus*, *Hippolytus Veiled*, a play with a more outspoken, active Phaedra as his source.<sup>70</sup> The most likely reason, however, appears to be that Seneca sought to recast Phaedra, the traditionally silent female character of Greek tragedy, in a new light and endow her with a louder voice and more developed mind than before. This new dimension to Phaedra's representation necessarily entails a more negative reading of her character, since she is made to transcend social norms and step out of her role as a silent, visible woman, a body without a mind, into a masculine role inclusive of mental and vocal abilities.

In the second stage of Phaedra's characterisation the extent of her psychological strength and ability to withstand pressure from the Nurse is analysed through her verbal interchange with her servant. Of the four dramatic renditions of the Phaedra legend, Tsvetaeva again proves the most forgiving, since she amplifies the negative characteristics of the Nurse in order to contrast Phaedra's own innocence and morality with them. The Nurse, a Mephistopheles-like character,<sup>71</sup> not only manipulates Phaedra into renouncing her husband and confessing to adulterous thoughts but also

<sup>66</sup>R. Lauriola, K. N. Demetrious, P.453.

<sup>67</sup>A. S. Gérard, P.27.

<sup>68</sup>M. Budzowka, trans. A Grzelak-Krzymianowska, *Phaedra - Ethics of Emotions in the Tragedies of Euripides, Seneca and Racine*, Peter Lang AG, (2012). P.122.

<sup>69</sup>M. Budzowka, P.122. The second version is alluded to frequently (for example by Aristophanes), and seems to have been very well known in Antiquity. The surviving Euripidean play is commonly regarded as the later one although there some controversy surrounds precise dating (cf. Gibert, J., "Euripides' Hippolytus Plays: Which Came First?", *Classical Quarterly* 47, 1997, p. 85ff; Hutchinson, G.O., 'Euripides' *Other Hippolytus*', *Zeitschrift für Papyrologie und Epigraphik* 149, 2004, pp.15-28).

<sup>70</sup>R. Lauriola, K. N. Demetrious, P.453.

<sup>71</sup>Анненский, И., 'Трагедия Ипполита и Федры', Серия "Литературные памятники": Анненский И. Ф., "Книги отражений М., "Наука", (1979).

encourages them.<sup>72</sup> She emphasises her own closeness to Phaedra by setting up an alternative form of relationship; she denounces blood ties, instead claiming that a milk bond between nurse and child is stronger: 'молочный голос – млеку покоримся – есть второе материнство'.<sup>73</sup> By reminding Phaedra that she had brought her up, the Nurse subtly arouses in Phaedra a sense of gratitude and manipulates her into believing that her counsel, like that of parents, will be most beneficial for her. The scene begins with the speech of the Nurse, who narrates the fate of Pasiphae and Ariadne, Phaedra's mother and sister respectively, and reminds her mistress that Theseus, before becoming her husband, was her sister's lover and betrayer: 'Ариадну супруг твой нынешний / Богу продал во время сна', implying that after Theseus' desertion of Ariadne she was raped by Dionysus. To further the effect of her accusations, the Nurse highlights Theseus' age and his consequent unsuitability for Phaedra: 'А с царем-то – с лица старик!... Федра, он тебе в отцы!'. Despite these provocations the queen is able to remain steadfast, replying to her Nurse's jibes with a series of aphorisms stressing her love for her husband. Instead of falling for the Nurse's manipulative speech and being drawn into a confession, Tsvetaeva's queen begs her to desist: 'Няня, я ему жена. / И оставь свои речи глухие!'. This is a noticeable departure from the earlier renditions of Euripides, Seneca and Racine, where each respective Nurse begs Phaedra to be silent, not vice versa.<sup>74</sup> The strength of Phaedra's character is further revealed by the number of lines required by the Nurse to pry out her secret (330 lines belonging to the Nurse as compared to the queen's 101 lines<sup>75</sup>) and her moral integrity is illuminated by the series of kind remarks with which, despite his absence and evident neglect of her, she describes Theseus:<sup>76</sup> 'Во-первых, – храбр... С каждым проходим запросто / Говорит... Щедр... Дальнего чтит... Седина ль Тезеева / Не мудра... На побежденных с кротостью / Зрит'. By making the Nurse a cunning temptress and Phaedra (initially) a loyal wife this scene in Tsvetaeva's play furthers the representation of the queen's strength of resolve.

In Euripides, the second stage of Phaedra's characterisation is, like the first stage, dramatically different from Tsvetaeva's, since, rather than making a conscious effort to pry Phaedra's illicit passion out of her, the Nurse stumbles upon it by mistake. The dramatic irony of the long dialogue between the Nurse and Phaedra in Euripides' tragedy consists mainly in the servant's inability to understand why the queen is hurt

---

<sup>72</sup>R. D. B. Thomson, P.349.

<sup>73</sup>M. Stadter Fox, P.55.

<sup>74</sup>Euripides' Nurse does however undertake the role of intermediary for Phaedra, which may have inspired Tsvetaeva in her negative portrayal of the character.

<sup>75</sup>R. D. B. Thomson, P.349.

<sup>76</sup>R. Lauriola, K. N. Demetrious, P.462.

by her speech.<sup>77</sup> The Nurse erroneously attributes to Phaedra the feeling of hatred for Hippolytus rather than love for him. When at last Phaedra's fatal attraction is discovered by the Nurse, it elicits dread, rather than support from the old woman.<sup>78</sup> In this scene Euripides makes Phaedra and her illicit desires too complex for her simple servant to understand and her mind too deep for the Nurse, a more conventionally portrayed female, to fathom. It is worth noting that in Tsvetaeva's play the nurse is unconventionally presented as having a sharp 'male' devious mind (cf. Annensky's comparison of the Nurse with Mephistopheles).

Like his Greek predecessor and in contrast to his Russian successor, Racine, in the second stage of Phaedra's characterisation, makes the Nurse unable to understand her mistress and similarly stumble upon the name of Hippolytus by mistake. When at last Phaedra reveals her illicit love, the Nurse's response is as piteous as her Greek counterpart's: 'Just heaven! All the blood's frozen in my veins. O despair! O crime! O you race without shame! Unfortunate voyage! O, miserable shore! Why did you come then to this place of danger?' (Juste ciel! tout mon sang dans mes veines se glace. / O désespoir! ô crime! ô déplorable race! / Voyage infortuné! Rivage malheureux, / Fallait-il approcher de tes bords dangereux?', (265-268). Like Euripides, Racine makes Phaedra too complex for the Nurse to understand, highlighting the dichotomy between the simple, open Nurse and her troubled mistress. Although her intelligence matches her Greek counterpart, the French Phaedra has a clearer sense of self-awareness; she analyses her own speech and comes to the realisation that verbally she has overstepped the boundaries of propriety.<sup>79</sup> However, when these boundaries are overstepped, Racine's Phaedra, unlike that of Euripides, shows little to no fear or remorse for what she has done: 'I've confessed it all: and I repent of nothing.' (312)<sup>80</sup> At this point Phaedra's 'conscious mind is subordinate to her passion'<sup>81</sup> and her previous self-awareness has irrevocably gone.

Finally, turning to the most damning of the three male accounts of the Phaedra legend, Seneca presents the dialogue between the Nurse and Phaedra in direct oppos-

<sup>77</sup>'a bastard with thoughts of legitimacy, you know him well, Hippolytus'; 'νόθον φρονοῦντα γνήσι', οἰσθά νιν καλῶς, Ἰππόλυτον' (309-10).

<sup>78</sup>'This is my death! Women, this is unendurable, I can not endure to live! Hatelful to me is the day, the light I see! I shall throw myself down, die and be quit of life! Farewell, I am gone! (οἴμοι, τί λέξεις, τέκνον· ὡς μ' ἀπώλεσας. γυναῖκες, οὐκ ἀνασχέτ', οὐκ ἀνέξομαι ζωσ': ἐχθρὸν ἡμαρ, ἐχθρὸν εἰσορῶ φάος ῥίψω μεθήσω σῶμ', ἀπαλλαχθήσομαι βίου θανούσα: χαίρετ', οὐκέτ' εἶμ' ἐγώ.' (353-357).

<sup>79</sup>'Maddened, where am I! What did I say? Where have I let my will and spirit go play? I have lost them' ('Insensée, où suis-je? et qu'ai-je dit? / Où laissé-je égarer mes vœux et mon esprit? / Je l'ai perdu', (179-181).

<sup>80</sup>A. S. Gérard, P.87.

<sup>81</sup>A. S. Gérard, P.87.

ition to that of Tsvetaeva; where in the Russian version the Nurse requires 330 lines to entice Phaedra into indulging her guilty thoughts, in the Latin play almost two hundred lines are needed for the Nurse to give up trying to persuade Phaedra against her passion. Seneca continues his representation of Phaedra in a 'masculine role', making her assert herself independently from the Nurse and announce both her illicit passion and its object. The only thing the Nurse can do for this new version of Phaedra is to caution her: 'quickly drive guilty thoughts from thy pure breast, put out these fires, nor show thyself obedient to this dread hope of love' ('nefanda casto pectore exturba ocus, extingue flammam ne te dirae spei praebe obsequentem;', (130-132). The effect achieved by this radical departure from both the earlier rendition of Euripides, is to transpose the responsibility for not concealing her passion from Phaedra and the Nurse, to Phaedra alone. Since in Seneca the queen's mind is more independent than its previous and subsequent counterparts, it necessarily carries more blame for her sinful conduct.

The third stage of Phaedra's characterisation has the greatest bearing on the overall representation of her 'mind', since it introduces a binary opposition between the male and the female<sup>82</sup>, which allows the poet to reveal the psychology of the two women (the Nurse and Phaedra) against a patriarchal backdrop. For two of the four plays studied in this article, this stage also completes Phaedra's characterisation and provides conclusive evidence as to whether the playwright intended to condemn or to rehabilitate the queen.

From the very beginning of the 'revelation' stage, Tsvetaeva, in her choice to have the Nurse deliver a letter from Phaedra, without orally relaying the passion of her mistress, radically departs from the plot-line supplied both by the mythological tradition and her predecessor Euripides. This innovation, which can be traced back to Ovid's *Heroides* 4,<sup>83</sup> gives scope for Phaedra to address Hippolytus herself, revealing at first hand the inner workings of her mind: her intelligence in adapting to Hippolytus' interests<sup>84</sup> and the strength of her character, required to keep her raving passion under control.<sup>85</sup> The first crucial detail in Tsvetaeva's representation of a no-

<sup>82</sup>A. Kondjak, K. Pomorska, S. Rudy eds. *Myth in Literature*, New York University Slavic Papers Volume V, (1985), P.105.

<sup>83</sup>Phaedra is given a voice in Ovid's *Heroides* 4, where she able to describe both the physical effects of her passion (7ff) and her changing pursuits, which she adapts to suit Hippolytus (37ff).

<sup>84</sup>Searching for the first spark of her love, Phaedra picks wholly natural phenomena: 'Началом / Взгляд был. . . Шаг был. Ошибаюсь: куст был началом звук был Рога, – чащ звук'. Phaedra is endowed with the sensitivity to adapt herself closely to Hippolytus' interest, and, at least in her eyes, increases her chances of winning Hippolytus over.

<sup>85</sup>At this stage in the play Phaedra fully reveals the extent of her passion: 'Дерево стояло, щедрой / Тенью путников поило. / Это я его спалила / Иступлением, тоскою' and consequently her strength with which hereto she suppressed her love.

bler, purer Phaedra, is to make Hippolytus and Phaedra meet for the first time<sup>86</sup>: the young man does not recognise Phaedra and she is forced to introduce herself to him: 'Не меня ль отец твой вдовый...'. Tsvetaeva thus distances Phaedra from Hippolytus almost to the point of displacing her from her role of step-mother, making her love-declaration less perverse. This initial mollification however is dramatically overthrown by the plethora of nouns which Hippolytus showers on Phaedra, highlighting their mother-son relationship and eliciting, as she is reminded of her maternal duty, both sympathy and contempt for the queen: 'Мачеха! Жена царева!... Тезееву супругу. . . Слово сына!'. In addition to this, Tsvetaeva remoulds other elements of the traditional Phaedra legend: the queen is closer in age to Hippolytus than to her husband Theseus, and her marriage is childless, something highly unnatural in an ancient society.<sup>87</sup> The effect of this is initially to soften the impact of the queen's confession by distancing her from her bodily roles of wife and mother, and then to emphasise her mental sin: of the 'dual domains' to which her female body is consigned, the 'venerated... maternal' is absent, whereas the 'sexual' is highlighted by her youth and childlessness.<sup>88</sup> The tragedy of Phaedra's representation is deepened by her death; the only body she can produce is her own hanged corpse.<sup>89</sup> Instead of giving life and earning merit through 'correct' female behaviour, Phaedra brings death, reversing her traditional role and earning condemnation, mixed however, as shown later, with pity. The scene concludes with Hippolytus' blunt rejection and silence from Phaedra, who as we learn from the next stage direction ('Кормилица (над телом Федры)'), will never speak again.

At this point in the play Tsvetaeva introduces the most crucial element which lifts blame from Phaedra: unlike the three other renditions of the legend, the Russian queen dies before the accusation of rape is launched and is completely innocent of Hippolytus' death. Ulterior motives such as fear for her eukleia (good name) present in the shame-culture driven Phaedra of Euripides or jealousy which motivates Racine's Phaedra, are removed from her psychology.<sup>90</sup> Tsvetaeva's Phaedra is above pragmatism or pettiness; her death is caused by unrequited love.<sup>91</sup> Whereas in earlier versions of the Phaedra legend, the queen accuses Hippolytus herself, Tsvetaeva fully clears Phaedra of any accusation concerning the rape. The blame lies fully upon the shoulders of the vengeful Nurse: 'Берегись, пес-женодав! / ... черное – белым, ... / ... белое – черным, / Явь – ложью, ложь – явью

---

<sup>86</sup> M. Stadter Fox, P.51.

<sup>87</sup> M. Stadter Fox, P.55.

<sup>88</sup> A. M. Barker, J.M. Gheith, P.7.

<sup>89</sup> M. Stadter Fox, P.56.

<sup>90</sup> A. S. Gérard, P.6.

<sup>91</sup> R. Lauriola, K. N. Demetrious, P.461.



/ Предстанет, представлю.' The effect achieved by this radical departure from tradition is to heighten the audience's sympathy for the dead queen; her sin was only illicit love, incestuous but fruitless, not murder. Consequently, the pity which earlier Phaedra loses on account of guilt, remains fully with Tsvetaeva's queen.<sup>92</sup>

Although Euripides' departure from the mythological tradition is less stark than that of Tsvetaeva, the third stage of characterisation in his play is marked by a crucial innovation: it is the Nurse who reports Phaedra's passion to Hippolytus, whereas the queen does not encounter the young man at all. This absolves the queen of breaking her proper female silence (a noble woman explicitly declaring her love on stage would have been inconceivable for the audience<sup>93</sup>) and to further her representation as a passive instrument of the well-meaning Nurse. Phaedra can only listen in on the conversation which determines her fate and comment on it for the audience: 'I am destroyed. Stand next to this door and listen what kind of din is being raised in the house.' (ἀπωλόμεσθα: ταῖσδ' ἐπιστάσαιπύλαις ἀκούσαθ' οἷος κέλαδος ἐν δόμοιςπίτνει.). The Nurse's confession brings to light a further crucial element in Euripides' gender representations, that of distorted communication.<sup>94</sup> Throughout the whole play mortal women converse with mortal women (Phaedra with the Nurse and the Chorus) and mortal men with mortal men (Theseus with Hippolytus).<sup>95</sup> The only breaches of this pattern result in death: the Nurse's speech to Hippolytus ends in Phaedra's death, Phaedra's written communication to Theseus ends in Hippolytus's death and the Nurse's disappearance from the plot. The play clearly shows that any puncture of the male bubble by a female is severely punishable and women, in order to avoid death and destruction, should keep to their own silent world.<sup>96</sup> So Hippolytus, a character traditionally regarded as the epitome of an ideal youth but, at closer investigation, irreverent and full of hubris<sup>97</sup>, outlines in his diatribe against women: 'men should put to live with [women] beasts, which bite, not talk, in which case they could not speak to any one nor be answered back by them' (ἄφθογγα δ' αὐταῖς συγκατοικίσεις δάκηθηρῶν, ἵν' εἶχον μήτε προσφωνεῖν τιναμήτ' ἐξ ἐκείνων φθέγμα δέξασθαι πάλιν' (646-650)).

Hippolytus's speech as a whole, delivered by the male protagonist of the play, reveals the prevalent attitude to women in ancient Greece: a female was regarded

<sup>92</sup>R. D. B. Thomson, P.343.

<sup>93</sup>R. Lauriola, K. N. Demetrious, P.462.

<sup>94</sup>C. Segal, *Euripides and the Poetics of Sorrow: Art, Gender and Commemoration in Alcestis, Hippolytus and Hecuba*, Duke University Press, (1993), P.92.

<sup>95</sup>Save for a few very brief exchanges between Theseus and the female chorus (e.g.790-810).

<sup>96</sup>For a more detailed study of the function of silence as opposed to character speech in Euripides' *Hippolytus* see Knox, B. M., (1952).

<sup>97</sup>R. Lauriola, K. N. Demetrious, P.453.

first and foremost as a creature who could 'threaten social structures'<sup>98</sup> and who, consequently, should be kept under lock and key.<sup>99</sup> Hippolytus shares Pericles' view that the best fate for a woman is to be both silent herself and unspoken of by men: 'It will be much for your honour not to recede from your sex and to give as little occasion of rumour amongst the men, whether of good or evil, as you can.' (Thucydides, 2.45.2).<sup>100</sup> In addition to depriving women of their voice, Hippolytus offers to strip them of their fundamental role in society, that of child-bearing,<sup>101</sup> suggesting that buying children from temples would be better: 'not from women should [men] draw their stock, but in thy temples offer gold or iron or ponderous bronze and buy a family, each man in proportion to his offering, and so in independence dwell, from women free.' (618-624). Understanding that his hope of buying children is an adynaton, Hippolytus rails against clever women, indicating his preference for bodies without minds: 'But a clever woman—that I loathe! May there never be in my house a woman with more intelligence than befits a woman! For Aphrodite engenders more mischief in the clever. The woman without ability is kept from indiscretion by the slenderness of her wit.' (640-644). For Hippolytus a woman's mind is so inexorably linked with her sexual body that it cannot function independently; it pollutes a man with its filth, resulting in his need to physically wash himself.<sup>102</sup> The destructive circuit of female body to female mind to male mind to male body, outlined by Hippolytus, is complete in Euripides' play, resulting in death and the destruction of the Athenian royal house.

Before the Nurse confesses to Hippolytus, Phaedra is able to restrain herself<sup>103</sup> but as soon as her secret is out and the female sphere is violated, Phaedra is compelled to act swiftly and provide an answering breach of her own: a letter to her husband. Whereas previously her communication was limited to the female sphere of her Nurse and the Chorus of women, now Phaedra, like her Nurse, enters the 'public realm of men'.<sup>104</sup> Although the queen is herself physically powerless to withstand the male influence of Hippolytus, her mind is as yet capable of providing a defense mechanism, playing one man (Theseus) off against another (Hippolytus). Ironically, although her living body is powerless to protect her from dishonour, her dead body, once its thinking mind has been exterminated, is able to do so. In other words, by

---

<sup>98</sup>'Women can both make and unmake culture – a man's honour can be tarnished through her...' - N. Sorokin Rabinowitz, P.157.

<sup>99</sup>'Hippolytus enumerates three systems of exchange – financial, sexual, linguistic- from which women must be excluded or in which they must be controlled. Women are seen as gifts exchanged by men to cement relations between men.' N. Sorokin Rabinowitz, P.158.

<sup>100</sup>N. Sorokin Rabinowitz, P.159.

<sup>101</sup>N. Sorokin Rabinowitz, P.158.

<sup>102</sup>N. Sorokin Rabinowitz, P.158.

<sup>103</sup>N. Sorokin Rabinowitz, P.165.

<sup>104</sup>N. Sorokin Rabinowitz, P.158.

claiming that her female body has been violated, she is able to protect her reputation and her mind.<sup>105</sup> Furthermore, when Phaedra resolves to die: 'I know but one thing, to die with all speed, the sole remedy for my present troubles', 'οὐκ οἶδα πλὴν ἓν, κατθανεῖν ὅσον τάχος, τῶν νῦν παρόντων πημάτων ἄκος μόνον.' (599-600), she does so without commonplace hysteria or fatalism, but wisely and responsibly<sup>106</sup>, realising that she has no other recourse to save not only her own reputation but also that of her children.<sup>107</sup> Although Tsvetaeva's Phaedra chooses what for a modern reader may anachronistically seem the most 'honourable' way out of her dilemma, Euripides' Phaedra, from the point of view of a shame-culture society, cannot kill herself before speaking out in defense of herself.<sup>108</sup> The only three roles Phaedra is allowed to adopt in this play are either silent, as at the start of the play or sexual, when she gives rein to her passion, or devious, as when she accuses Hippolytus.<sup>109</sup>

In Racine's play the third stage of characterisation is also marked by a considerable departure from tradition since he, unlike all the other playwrights, remolds the character of Hippolytus to suit the taste of his audience: he removes Hippolytus' hubris and chastity, which could have seemed curious if not comic elements from the point of view of a 17th French audience.<sup>110</sup> In addition to this, he also reduces the potential guilt of Phaedra through her belief that when she declares herself to Hippolytus she is a widow.<sup>111</sup> By introducing these two dramatic innovations Racine adds a 'new psychological element', female jealousy of Aricia, which in turn allows him to trace the heroine's 'moral decline' from self-control to vindictive vengefulness.<sup>112</sup> Instead of having the Nurse confess Phaedra's love for Hippolytus, Racine makes the queen speak to her beloved herself, revealing by this tactic the extent of the queen's self-criticism. She has a mind, which she uses to disclose and then analyse her guilt.

At this stage in Phaedra's characterisation it is more fruitful to examine Racine alongside, rather than separately from, Seneca, since their versions are mirror opposites. In Racine's play Phaedra's confession of love for Hippolytus is set in the con-

<sup>105</sup>N. Sorokin Rabinowitz, P.166.

<sup>106</sup>Анненский, И., (1979).

<sup>107</sup>But I, with all my thought, can but one way discover out of this calamity, that so I may secure my children's honour, and find myself some help as matters stand. For never, never will I bring shame upon my Cretan home' ('ἐγὼ εὐρημα δῆτα τῆσδε συμφορᾶς ἔχω, ὥστ' εὐκλεῆ μὲν παισὶ προσθεῖναι βίοναυτῆ τ' ὀνάσθαι πρὸς τὰ νῦν πεπτωκότα. οὐ γάρ ποτ' αἰσχυρῶ γε Κρησίουσ' δόμοσ' (715-719)).

<sup>108</sup>N. Sorokin Rabinowitz, P.166.

<sup>109</sup>N. Sorokin Rabinowitz, P.167.

<sup>110</sup>R. Lauriola, K. N. Demetrious, P.460.

<sup>111</sup>This probably had precedent in Sophocles' *Phaedra*, where Theseus undertook a journey to the underworld and went missing, presumed dead. See Barrett, W.S., 1964, ff. 29.

<sup>112</sup>A. S. Gérard, P.92.

text of her request that he protect her son after the reported death of Theseus.<sup>113</sup> In Act I the upright Oenone had already suggested this course of action.<sup>114</sup> Seneca, on the other hand, has Phaedra only feign the countenance and position of a poor widow, revealing in an aside that the queen's true intent is to arouse sympathy and love in Hippolytus: '[Aside.] O credulous hope of lovers, O deceitful Love! Have I said enough? I'll bring my prayers to bear upon him and attack,' 'O spes amantum credula, o fallax Amor! satisne dixi? precibus admotis agam. [the following words are not translated in your preceding text] Miserere, tacitae mentis exaudi preces—libet loqui pigetque.' (634-636).<sup>115</sup> Racine makes his Phaedra genuine, Seneca – false. At a fundamental level the psychology of the two women differs, eliciting compassion at the plight of one and horror at the shamelessness of the other.

In addition to this, Seneca has Hippolytus condemn Phaedra, whereas Racine makes his prince silent, too noble to censure his own step-mother.<sup>116</sup> Instead Phaedra, acknowledging her own fault, herself utters the very words which Seneca has his Hippolytus speak.<sup>117</sup> Racine's Phaedra urges Hippolytus to punish her,<sup>118</sup> whereas Seneca's prince draws his sword himself.<sup>119</sup> Racine's Phaedra labels herself as a 'dreadful monster' ('monstre affreux', (703)), whereas Seneca's prince condemns the queen as unnatural ('your monster-bearing mother, thou worse than she who bore thee' (688-689)); Racine's Phaedra understands that her blood is too filthy to be shed by Hippolytus' noble hand,<sup>120</sup> whereas it is Seneca's prince<sup>121</sup> who feels disgust at her touch.<sup>122</sup> The build-up of the French Phaedra's self-loathing makes her final wish to die appear much more genuine than that of her Latin counterpart, whose submission to Hippolytus' sword can be read as a manipulative gesture to make the prince feel sorry for her great love.<sup>123</sup> Whereas Racine's Phaedra is deeply aware of her own wickedness and is spontaneous in her confession, Seneca makes Phaedra act

---

<sup>113</sup>I come, on my son's behalf, to explain my fears. My son is fatherless... you alone can protect him from his enemies.' 'Je vous viens pour un fils expliquer mes alarmes. / Mon fils n'a plus de père... / Vous seul pouvez contre eux embrasser sa défense', (586-590).

<sup>114</sup>A. S. Gérard, P.81.

<sup>115</sup>A. S. Gérard, P.81.

<sup>116</sup>A. S. Gérard, P.84.

<sup>117</sup>A. S. Gérard, P.84.

<sup>118</sup>'Take vengeance: punish me for loathed delight,' 'Venge-toi, punis-moi d'un odieux amour.' (699).

<sup>119</sup>'Out, sword, and mete her just punishment. See, with left hand in her twisted hair have I bent back her shameless head.' (705-6).

<sup>120</sup>'Or if you think it not worthy of your blow... Or if your hand by my vile blood would be stained,' 'Ou si tu le crois indigne de tes coups... / Ou si d'un sang trop vil ta main serait trempée', (707-709).

<sup>121</sup>'And let this sword, polluted by thy touch, quit my chaste side,' 'hic contactus ensis deserat castum latus.' (714).

<sup>122</sup>A. S. Gérard, P.84.

<sup>123</sup>A. S. Gérard, P.77.

and speak deliberately: from the very start of her conversation with Hippolytus, she consciously tries to seduce him using every trick she can think of and 'embarks on her evil course of action in full knowledge that she is violating the rational-ethical principles that should govern human conduct'.<sup>124</sup>

The fourth and final stage of Phaedra's characterisation is the most problematic since in two of the four plays (those of Euripides and Tsvetaeva) Phaedra is already dead and so incapable of condemning or absolving herself of her crime. The role of ultimate judge is handed over to the male protagonists of the play, who speak after the females are robbed of their voices.<sup>125</sup> The first time we encounter Theseus in Tsvetaeva's play is almost at the very end when he enters on stage, wondering why his house is abandoned: 'Двор вымер, дом вымер.' Like Euripides' Theseus, the king of Tsvetaeva's play first asks if the silence at court is caused by the death of his son 'Сын помер, / Что ль?'. Since Phaedra has no children, Theseus can only be referring to Hippolytus, his illegitimate son from the Amazon Hippolyta. The king treasures this son more than his wife, fearing, when he attributes to his death the same weight as an enemy intrusion or plague ('Враг в доме? ... Мор грянул?'), for Hippolytus' health as much as for the safety of the whole house. This homo-centric attitude finds precedent in Euripides, who, like Tsvetaeva, has his Theseus ask first<sup>126</sup> after his father and then after his sons.<sup>127</sup> When at last, after eleven lines, Theseus understands that his wife is dead, his mourning is directed more at himself and his wretched fate than at lamenting Phaedra's untimely death.<sup>128</sup> The king's callousness becomes all the more conspicuous when juxtaposed to the grief of the Chorus, expressed entirely from the point of view of Phaedra: 'Alas, poor woman, how luckless you are!... Who was it, poor woman, that brought your life down to darkness? ('ὦ ἰὼ τάλαινα μελέων κακῶν: τίς ἄρα σάν, τάλαινα', ἀμαυροῖ ζόραν;' (8111; 816)). This negative characterisation of Theseus allows both Euripides and Tsvetaeva to explore the fate of a woman, psychologically more developed (as has been shown in stages one, two and three of Phaedra's characterisation) than her husband, contrasting her dynamic but repressed conscience with his free but one-dimensional mind.

The final revelation of Hippolytus' innocence and Phaedra's criminal passion has a slightly different manifestation in each play. In Tsvetaeva, Theseus takes full responsibility for the tragedy of Phaedra's and Hippolytus' death, claiming that the gods

<sup>124</sup>A. S. Gérard, P.27-28.

<sup>125</sup>M. Stadter Fox, P.51.

<sup>126</sup>'Has something happened to old Pittheus? ... Alas! Does death rob me a child's life...?' ('μῶν Πιτθέως τι γῆρας εἴργασται νέον', 'οἴμοι: τέκνων μοι μή τι συλᾶται βίος;' (794; 799).

<sup>127</sup>N. Sorokin Rabinowitz, P.68.

<sup>128</sup>N. Sorokin Rabinowitz, P.68.

(on the macrocosmic plane analysed above) are punishing him.<sup>129</sup> Theseus, a flat, one-dimensional character, appears at the very end of the tragedy and re-evaluates the whole play, making himself the centre of the whole plot.<sup>130</sup> His masculine desire to impose himself at the centre of all action is severely ironised by Tsvetaeva both in the name of the play itself (*Fedra*) and by the overall title of the trilogy into which *Fedra* falls: *Thesee*. Although the character of Theseus is used by the poetess to unite three women (Ariadne, Phaedra and Helen) he is 'almost an empty centre which functions only as a binding element.'<sup>131</sup> The main interest of Tsvetaeva lies not in analysing Theseus's desire for Ariadne or Phaedra (or indeed Helen, who, according to some Classical sources, was only seven at the time of Theseus' rape), but in exploring their characters and individual female tragedies.

The concluding scene of Euripides' tragedy similarly revolves around the characters of Theseus and Hippolytus and their reunion after the accusation of rape has been absolved by Artemis as *dea ex machina*. Leaving aside the immortals, the resolution of the plot takes place between two men once the temporary, female interference has been abolished.<sup>132</sup> Only after both Phaedra's 'sexuality' (body) and her 'speech' (mind) have been 'controlled' by the two males can the play end in relative harmony (father and son reconcile at last).<sup>133</sup> By concluding *Hippolytus* with the embrace of father and son, while the female, formerly princess of Crete, daughter of King Minos and Queen of Athens, now a speechless body, is removed from the prominent role she occupied through the entire tragedy, Euripides reaffirms the prince's earlier claim that women are nothing more than an object of monetary value, used in this play to explore male relationships.<sup>134</sup> In Phaedra the 'binary division between mind and body' has been displaced, leaving only her body on stage, her name has survived not in her spiritual or maternal hypostasis but in the corporeal, sexual domain where she is 'damned' in the eyes of posterity.<sup>135</sup>

The other two plays studied in this article, those of Seneca and Racine manifest the fourth stage of Phaedra's characterisation in a dramatically different way from their predecessor Euripides and successor Tsvetaeva, in that both of them allow Phaedra to survive until after the death of Hippolytus. This allows Phaedra to speak out for herself rather than surrender, like Tsvetaeva's or Euripides' heroines, to a male judgment of her sin. Both Seneca and Racine give their respective Phaedras

---

<sup>129</sup>M. Stadter Fox, P.62.

<sup>130</sup>M. Stadter Fox, P.62.

<sup>131</sup>M. Stadter Fox, P.52.

<sup>132</sup>A. S. Gérard, P.34.

<sup>133</sup>N. Sorokin Rabinowitz, P.156.

<sup>134</sup>N. Sorokin Rabinowitz, P.156.

<sup>135</sup>A. M. Barker, J.M. Gheith, P.7.

the option of escaping death and punishment by upholding their initial lie about Hippolytus' crime.<sup>136</sup> This is facilitated by the fact that in both plays the Nurse, the only witness to the false accusation, conveniently disappears from the plot.<sup>137</sup> However, both Phaedras make their conscious decision to forego this opportunity and confess to their crime, simultaneously robbing themselves of their reputation and life. In Racine Phaedra punishes herself more severely than the Roman Phaedra, since before appearing on stage to make her confession, she takes a slow-acting poison brought by Medea from Colchis.<sup>138</sup> Even in the unlikely event that Theseus forgave her crime and begged her not to commit suicide, Phaedra has no way of turning back: she has evaluated her crime and dispensed justice herself without resorting to male intervention. Much like Racine's Phaedra, Seneca's queen undergoes what has been termed 'a shockingly abrupt psychological change' in that, whereas formerly her mind and her body worked in unison, both deliberately seductive, by the end of the play her mind has broken free of her body.<sup>139</sup> Unlike Euripides, Seneca has his Phaedra realise her crime and ask forgiveness.<sup>140</sup> Her characterisation, initially the most negative of the four versions of the Phaedra legend, is redeemed at the end of the play. As in the case of Euripides' Phaedra there is a 'binary division between [her] mind and body',<sup>141</sup> but in an altogether different form: her mind has overcome her body and led her to the noblest course of action available to a fallen queen, death.

### Conclusion

In this article the representation of gender in Tsvetaeva's play has been analysed through the reception and transmission of the Phaedra myth in preceding literary sources from three different cultures, Greek, Roman and French. The three plane approach (extra-textual, macrocosmic and microcosmic) offered in the article has facilitated a precise analysis of the extent to which the female body and mind have been condemned or rehabilitated by each poet. It can be concluded that the earliest of the four plays, Euripides' *Hippolytus*, sets up a relatively sympathetic interpretation of the Phaedra legend: although on the extra-textual plane his society, constructed around a patriarchy, condemned Phaedra's bodily crime as wife and mother, at the macrocosmic and microcosmic levels Euripides provides sufficient mollification to adopt a balanced view of Phaedra's body and mind respectively: she is represented as

---

<sup>136</sup> A. S. Gérard, P.34.

<sup>137</sup> A. S. Gérard, P.34.

<sup>138</sup> A. S. Gérard, P.103.

<sup>139</sup> A. S. Gérard, P.35.

<sup>140</sup> A. S. Gérard, P.45.

<sup>141</sup> A. M. Barker, J.M. Gheith, P.9.

both a victim and a wrongdoer. Seneca's play, chronologically the next after Euripides, presents a completely different representation of gender in the Phaedra myth: both the queen's body and mind are condemned at the macrocosmic level and extra-textual plane, with only the fourth stage of the microcosmic level providing a mitigating feature, the voluntary confession of her bodily crime and the redemption of her mind. Racine's rendition of the Phaedra legend reverts back to Seneca in that it attempts to absolve Phaedra's mind of her sin on two planes, the external and microcosmic, only the macrocosmic plane which reveals an absence of divine motivation serves to condemn the queen. Thus, at all the three levels, Tsvetaeva's play, the last of the four in terms of chronology, can conclusively be named the most sympathetic rendition of the Phaedra legend, since the poet, inheriting the schematic plot of the myth, has recast it through her own female viewpoint. Tsvetaeva has rescued Phaedra 'from her portrait as [a] destructive... being – the familiar... interpretation of masculine tradition'<sup>142</sup> and reshaped her into an independently thinking, innocent woman trapped in a homo-centric society. Tsvetaeva's heroine has overcome the 'binary division between mind and body – spirit and matter'<sup>143</sup> liberating herself of the bonds of a patriarchal society. In Tsvetaeva's play Phaedra is endowed not only with a mind which is independent of her body (she is able to suppress her passion and fight back against the tempting Nurse) but also noble enough to lead her to the finest course of action, death without accusation. ❁

### **Acknowledgements**

This work was supported by the Arts and Humanities Research Council (AHRC).

---

<sup>142</sup>B. Heldt, *Terrible Perfection: Women and Russian Literature*, Indiana University Press (1987), P.131.

<sup>143</sup>A. M. Barker and J.M. Gheith ed., *Ibid.*, P.7.



### References

- Barker, A.M., Gheith, J.M., ed., *A History of Women's writing in Russia*, Cambridge University Press, 2002.
- Barrett, W. S., *Euripides: Hippolytos*, Clarendon Press: Oxford University Press, 1964.
- Budzowka, M., *Phaedra - Ethics of Emotions in the Tragedies of Euripides, Seneca and Racine*, trans. A. Grzelak-Krzymianowska, Peter Lang AG, 2012.
- Coudenys, W., «Те последние вы можете исправлять с точки зрения стилистики» : четыре забытых французских перевода Марины Цветаевой, *Revue des études slaves*, V. 66, 1994.
- Gérard, A.S., *The Phaedra Syndrome: Of Shame and Guilt in Drama*, Amsterdam – Atlanta, GA, 1993.
- Heldt, B., *Terrible Perfection: Women and Russian Literature*, Indiana University Press, 1987.
- Hutchinson, G.O., 'Euripides' Other Hippolytus', *Zeitschrift für Papyrologie und Epigraphik* 149, 2004.
- Gibert, J., "Euripides' Hippolytus Plays: Which Came First?", *Classical Quarterly* 47, 1997.
- Lauriola, R., Demetrious, K.N., ed. *Brill's Companion to the Reception of Euripides*, v.3, Brill, 2015.
- Karlinsky, S., *Marina Cvetaeva: Her Life and Her Art*, University of California Press, 1966.
- Karlinsky, S., *Marina Tsvetaeva: The woman, her world and her poetry*, Cambridge University Press, 1985.
- Kelly, C., *A History of Russian Women's Writing 1820-1992*, Oxford: Clarendon, 1998.
- Knox, B. M., *The Hippolytus of Euripides*. Yale University Press, 1952.

Kondjak, A., Pomorska, K., Rudy, S., eds. *Myth in Literature*, New York University Slavic Papers Volume V, 1985.

Makin, M., *Marina Tsvetaeva: Poetics of Appropriation*, Clarendon Press Oxford, 1993.

Marsh, R., ed., *Women and Russian Culture: Projections and Self-Perceptions*, Berghahn Books, 1998.

Powell, A., ed., Euripides, *Women and Sexuality*, Routledge, 1990.

Sorokin Rabinowitz, N., *Anxiety Veiled: Euripides and the Traffic in Women*, Cornell University Press, 1993.

Stadter Fox, M., *The Troubling Play of Gender: The Phaedra Dramas of Tsvetaeva, Yourcenar, and H.D.*, Associated University Presses, 2001.

Thomson, R.D.B., 'Tsvetaeva's Play "Fedra": An Interpretation,' *The Slavonic and East European Review*, Vol. 67, No. 3, 1989.

Torlone, Z., "Voicing Passions: Phaedra of Euripides and Marina Tsvetaeva," ASEES, Washington, DC, 2008.

Torlone, Z., *Soul and Passion: Marina Tsvetaeva's Classical Plays. Introduction, Translation and Commentary*, 2011.

Torlone, Z., "Muted Voices: Marina Tsvetaeva's Classical Heroines" In *Brill's Companion to Classical Reception and Modern World Poetry* (ed. Polina Tambakaki), 2019.

Анненский, И.Ф., 'Трагедия Ипполита и Федры', Серия "Литературные памятники", 1979.

Анненский, И.Ф., "Книги отражений", М., "Наука", 1979.

Войтехович, Р.С., Античные мотивы в творчестве Марины Цветаевой. Тарту, 2007.

Войтехович, Р.С., Цветаева и античность. Москва - Тарту, 2008.

Войтехович, Р.С., Психея в творчестве М. Цветаевой: Эволюция образа и сюжета. Тарту, 2005.

Войтехович, Р.С., Ариадна Марины Цветаевой: Поэтика античности // Северный сборник: Proceedings of the №rFA Network in Russian Literature 1995–2000. Stockholm Studies in Russian Literature. 34. Stockholm, 2000.

Каган, Ю. М., И.В Цветаева. Жизнь. Деятельность. Личность. Москва, 1987.

# Critical Theory's Objectives: Unfulfilled in Praxis, Fulfilled in Critique

*Makiko Miyazaki\**

*Department of Politics and International Relations*

Critical theory has been one of the prominent theoretical schools in the discipline of International Relations (IR) since the so-called 'critical turn in IR' in the 1980s. However, given the sheer diversity and complexity of what constitutes critical theory, what exactly critical theory purports to do and whether it has, or even can, achieve its objectives remain a contested topic.<sup>1</sup> This article attempts to contribute to this debate by introducing a heuristic tool of praxis/critique to delineate different ways of assessing critical theory.

The article proceeds as follows. First, I provide a brief overview of what critical theory is and identify four objectives that critical theory purports to pursue: 1) denaturalise the world to expose its underlying social relations and power dynamics; 2) examine how the world became what one observes through historical inquiry; 3) make judgments on how the world could have been, could be, and/or ought to be; and 4) explore the possibilities for transforming the world.<sup>2</sup> I then argue that these four objectives can either be interpreted as oriented primarily towards praxis or critique/reflexivity (henceforth, 'critique'), and the choice of interpretation colours whether one believes critical theory to have achieved its objectives. Considered in terms of praxis, critical theory has not and cannot achieve its objectives, because of the difficulty in formulating universal and non-hegemonic definitions, the never-

---

\*makiko.miyazaki@politics.ox.ac.uk

<sup>1</sup>For instance, see Craig Murphy, "The Promise of Critical IR, Partly Kept," *Review of International Studies* 33, no. 1 (2007): 117–33; Davide Schmid, "The Poverty of Critical Theory in International Relations: Habermas, Linklater and the Failings of Cosmopolitan Critique," *European Journal of International Relations* 24, no. 1 (2018): 198–220; Kimberly Hutchings, "Happy Anniversary! Time and Critique in International Relations Theory," *Review of International Studies* 33 (2007): 71–89.

<sup>2</sup>Richard Shapcott, "Critical Theory," in *The Oxford Handbook of International Relations*, ed. Christian Reus-Smit and Duncan Snidal (Oxford: Oxford University Press, 2008), 337.

ending nature of the search for universal well-being, and the lack of access to the truly marginalised. Considered in terms of critique, however, critical theory has achieved its objectives by denaturalising the field of IR, raising awareness of the marginalised voices, and highlighting the political implications of theories. After I highlight critical theory's room for improvement, I suggest that it is more useful to judge critical theory's accomplishments from the perspective of critique and to hold praxis as an important guiding force for future improvement.

Before proceeding, I emphasize that while critical theory has its origins in the wider philosophy and social science traditions that predate the 1980s, the focus of this article is critical theory as adopted by and used in IR. Therefore, my use of the term 'critical theory' henceforth is a shorthand for critical theory in IR. I also note that the term 'critical theory' originally referred narrowly to the tradition called the Frankfurt School, which coined the term in 1937.<sup>3</sup> However, scholarly convention has broadened the term to encompass various strands of theory and methods – as discussed below – while designating the Frankfurt School as a large "C" Critical Theory<sup>4</sup>, and I adopt this distinction for the purposes of this article. The narrow focus on critical theory in IR leads to my second clarification regarding 'mainstream theory:' this term will also only refer to IR theories from which critical theory distinguishes itself. The distinction between critical theory and mainstream theory will be delineated in the next section. Finally, 'praxis' can be understood as an act of actualising change in the world, most famously inspired by Marx, while 'critique'/'reflexivity' refers to one's awareness of and reflection on the backdrop of one's theorising, whether it be one's role as a theorist or the context of time, social/political space, ideology, perspective, and/or interest.<sup>5</sup> Such awareness and reflection is premised on the notion, made famous by Cox, that "[t]heory is always for someone and for some purpose."<sup>6</sup> Praxis, critique, and reflexivity are discussed further in the 'Praxis vs. Critique' section.

<sup>3</sup>Max Horkheimer, "Traditional and Critical Theory," in *Critical Theory: Selected Essays* (New York: Continuum, 1972) (1937 article published in 1972 collection of essays); Steven C. Roach, "Critical Theory of International Relations," Oxford Bibliographies, May 29, 2019, <https://www.oxfordbibliographies.com>.

<sup>4</sup>Shapcott, 329.

<sup>5</sup>g. Jurgen Haacke, "Theory and Praxis in International Relations: Habermas, Self-Reflection, Rational Argumentation," *Millennium* 25, no. 2 (1996): 257; Diana Boros, "Review: Critical Theory and the Challenge of Praxis," *Contemporary Political Theory* 18, no. 1 (2017): 5–7; Fabian Freyenhagen, "Critical Theory: Self-Reflexive Theorizing and Struggles for Emancipation," Oxford Research Encyclopedia of Politics, 2018; Robert W. Cox, "Social Forces, States and World Orders: Beyond International Relations Theory," *Millennium* 10, no. 2 (1981): 128.

<sup>6</sup>Cox, 128. Emphasis original.

## Overview

### *What is critical theory?*

Critical theory emerged in IR during the 1980s as a school of thought that challenged the 'mainstream' or dominant theoretical paradigms in the discipline at the time, particularly realism.<sup>7</sup> The challenge extends to at least three realms: ontological positions, the role of values, and the role of critique. First, critical theory rejects the mainstream theories' positivist basis, which posits the existence of an objective, independently existing world. Instead, critical theory contends that the world is constructed by interpretations, and relations of power privilege certain interpretations over others.<sup>8</sup> Second, critical theory criticizes mainstream theories for their "value neutrality" or objectivity in analysing the world as it is, with little regard to how it is undergirded by domination (whether by class, race, gender, or knowledge production), and characterizes itself as value-bound.<sup>9</sup> Third, critical theory emphasises reflexivity, raising questions about the assumptions, perspectives, and biases that shape theories and by extension the world with which theories intersect.<sup>10</sup> Although works of mainstream theories do engage in reflection to a certain extent by critiquing previous scholarly works, critical theorists place reflexive critique at the centre of their endeavour.

Granted, once we define critical theory against mainstream theories, important differences within critical theory become visible. After all, critical theory is often considered an umbrella term over at least three distinct strands: 1) the Gramscian tradition and the Frankfurt School, both a reformulation of Marxist thought; 2) post-structuralism, focused on genealogy, discourse, and power/knowledge; and 3) gender studies and post-colonialism, focusing not just on class but also gender and race as structures of domination.<sup>11</sup> These strands differ in significant ways. For instance, the Frankfurt School holds as its goal cosmopolitan human 'emancipation' – allowing everyone to achieve freedom both "from unnecessary suffering" and "to partake in dialogue, consent, and deliberation" in matters of shared concern<sup>12</sup> – yet post-

<sup>7</sup>Faruk Yalvaç, "Critical Theory: International Relations' Engagement With the Frankfurt School and Marxism," Oxford Research Encyclopedia of International Studies, December 1, 2015, <https://oxfordre.com>.

<sup>8</sup>Cox, 129-30.

<sup>9</sup>Shapcott, 328.

<sup>10</sup>Cox, 128.

<sup>11</sup>Shapcott; Edward Keene, "Methods and Approaches in Critical Theory" (Research Design and Method in IR Hilary Term, University of Oxford, Oxford, February 4, 2021); J. Ann Tickner and Laura Sjoberg, *Feminism and International Relations: Conversations about the Past, Present and Future* (Abingdon: Routledge, 2011); Sanjay Seth, "Postcolonial Theory and the Critique of International Relations," *Millennium: Journal of International Studies* 40, no. 1 (2011): 167–83.

<sup>12</sup>Shapcott, 328.

structuralists, gender studies scholars, and post-colonialists are sceptical of emancipation as a universalising notion, as discussed later. Whereas post-structuralists often focus on revealing power/knowledge by examining how a particular discourse is made dominant, propagated, implemented, and reinforced in the social world, Gramscian/Frankfurt School theorists do not, and gender studies scholars and post-colonialists often criticise post-structuralists for not promoting (more substantial) change.<sup>13</sup> While this is far from a comprehensive list of differences, these differences alone demonstrate that scholars are right to refrain from assigning any strand-specific objective, such as emancipation, to critical theory. Nevertheless, concluding that these strands should not be considered together is to dismiss the significant commonalities among them. An analogy to realism is pertinent here. There are important distinctions within realist thought, such as the respective roles of human nature versus anarchical structure in explaining war and whether the ultimate state interests are survival or power maximisation/hegemony.<sup>14</sup> Yet they share fundamental assumptions, such as egoism and the importance of material power, that distinguish the realist thought from others, and scholars have found it useful to recognise realism as an “intellectual tradition” whose definitions “vary considerably in their details but reveal a striking family resemblance.”<sup>15</sup> I contend that adopting the same approach is useful for critical theory.

#### *What are critical theory’s objectives?*

That critical theory in all its diversity is characterized by its post-positivist, value-bound, and reflexive nature could be translated into the four objectives outlined in the introduction. Critical theory seeks to 1) denaturalise the world, or in other words demonstrate that what appears natural in the world are socially constructed and enforced through power dynamics<sup>16</sup>; 2) reveal, through historical inquiry, how the present world came to be; 3) make judgments on how the world could have been, could be, and/or ought to be; and 4) explore the possibilities for transforming the world. As one may suspect from the above discussion of critical theory’s diversity, the strands differ in the extents to which they attempt to meet these objectives. The

<sup>13</sup>“Methods and Approaches in Critical Theory.”

<sup>14</sup>E.g. Hans J. Morgenthau, *Politics among Nations: The Struggle for Power and Peace* (New York: Knopf, 1954); Kenneth N. Waltz, *Theory of International Politics* (Reading: Addison-Wesley, 1979); John J. Mearsheimer, *The Tragedy of Great Power Politics* (New York: W.W. Norton Company, Inc., 2001).

<sup>15</sup>William C. Wohlforth, “Realism,” in *The Oxford Handbook of International Relations*, ed. Christian Reus-Smit and Duncan Snidal (Oxford: Oxford University Press, 2008), 134, 132.

<sup>16</sup>For instance, the capitalist relationship between labour and wages used to buy labour are underpinned by the social relationship between the worker and the manager, both human beings with different levels of power. “Methods and Approaches in Critical Theory.”

Gramscian tradition, Frankfurt School, post-structuralism, gender studies, and post-colonialism arguably engage in the first two objectives to similar extents, differing in their focus (class, discourse, gender, or race). The process of denaturalisation and study of the world's trajectory to the present invite judgments on how the world could have been and could be. As for articulating what ought to be, the Frankfurt School may be the most normative, but even post-structuralism, which as discussed earlier is often criticized for not envisioning enough change, could potentially be said to engage in normative critique when it sheds light on "subjugated knowledges" marginalized by the dominating discourses.<sup>17</sup> By engaging in the first three objectives, critical theory's strands explore the possibilities for transforming the world, although that again is most clearly demonstrated by the Frankfurt School as articulated below.

### Praxis vs. Critique

Considering the objectives as leaning towards praxis reflects critical theory's inspiration from Marx: "philosophers have only interpreted the world, in various ways; the point is to change it."<sup>18</sup> This attitude is most visible in the Frankfurt School's focus on emancipation and criticism that scholars have not have enough "political impact" in the real world, although the attitude is present in other strands as well.<sup>19</sup> Meanwhile, the interpretation of the objectives as leaning towards critique mirrors Cox's understanding of critical theory: one that is "more reflective upon the process of theorising itself: to become clearly aware of the perspective which gives rise to theorising, and its relation to other perspectives."<sup>20</sup> While critical theory in this sense makes "a normative choice" and envisions changes to the status quo, it does not immediately necessitate tangible action: "[a] principal objective of critical theory is to *clarify* [the] range of possible alternatives," or "feasible transformations," of the current world.<sup>21</sup> Praxis and critique are both important suggestions for what critical theory should do, but choosing which one is more pertinent colours how we understand critical theory.

Considered in terms of praxis, critical theory has not and cannot achieve its objectives because it is difficult to create universal and non-hegemonic definitions, achieving universal well-being is never-ending, and scholars lack access to the truly marginalised. Emancipation is a case in point. Emancipation, inspired by figures including Kant, Marx, and Hegel, focuses on the "use [of] reason to reflect upon...

<sup>17</sup>Tom Keenan, "The 'Paradox' of Knowledge and Power: Reading Foucault on a Bias," *Political Theory* 15, no. 1 (1987): 7.

<sup>18</sup>Karl Marx, "Theses on Feuerbach," in *Marx/Engels Selected Works*, trans. W. Lough, vol. 1 (Moscow: Progress Publishers, 1969), 13–15. Article X.

<sup>19</sup>Murphy, 118.

<sup>20</sup>Cox, 128.

<sup>21</sup>Cox, 130. Emphasis added.



‘mankind’s self-incurred immaturity,’” and to use critical scrutiny so that “humankind could ‘actively determine its own way of life.’”<sup>22</sup> That this notion, reflecting Western Enlightenment ideals, should be applied globally is often criticised for emphasising universality and hegemony at the expense of cultural diversity, such that some have called to abandon the search for a notion of emancipation acceptable to everyone.<sup>23</sup> While Linklater has addressed this by incorporating the principle of ‘no harm’ to the understanding of emancipation,<sup>24</sup> it remains questionable whether ‘doing no harm’ is an effective alternative: if our commitment to others is negative and interventions to correct injustices are discouraged, to what extent is this emancipation? Is it not closer to sympathetic inaction instead? This is further complicated by critical theory’s aversion to considering emancipation in relativist terms. Critical theory’s rejection of relativism is based on the association of relativism with “withdrawal,” “passivity,” and “defeatism” with regards to the world as it is.<sup>25</sup> As long as critical theory faces problems in both the universalist and relativist approaches, there would be disagreement on what emancipation entails, and little action could be launched without such guidance. Furthermore, imposing a definition is a form of power itself: by critiquing it, formulating alternatives, and critiquing the alternatives again, critical theory becomes constrained to this never-ending cycle. If critical theory holds that empirical research and praxis “ought not be begun until [the normative and metatheoretical groundwork] has been laid,”<sup>26</sup> it is questionable when critical theory could actually move to praxis.

Also never-ending is the task to achieve universal well-being. Self-other distinctions and the relative lack of commitment to improve the lives of those far removed from us are too innate to be fully dismantled.<sup>27</sup> The process of achieving emancipation could also elicit new sources of discontent, inequality, and marginalisation. We have to draw an arbitrary line as to where emancipation is achieved, choosing incomplete emancipation at best. Finally, “the social worlds of critical IR scholars and those [they] wish to serve are so disconnected” by virtue of the former being part of the privileged group and not the latter.<sup>28</sup> Regardless of good intentions, there is a

<sup>22</sup>Horkheimer, quoted in Bernstein 1976, 181, quoted in Shapcott, 331.

<sup>23</sup>Diez and Steans, 135-7; Cox, 137. For discussions of universality and hegemony, see Stephen Hopgood, *The Endtimes of Human Rights* (Ithaca: Cornell University Press, 2013), 2.

<sup>24</sup>Andrew Linklater, “The Harm Principle and Global Ethics,” *Global Society* 20, no. 3 (n.d.): 329–43.

<sup>25</sup>Espen Hammer, “Critical Theory and the Challenge of Relativism,” in *The Routledge Handbook of Philosophy of Relativism*, ed. Martin Kusch (London: Routledge, 2019), 247–8. Horkheimer and Habermas were strong proponents for this view.

<sup>26</sup>Shapcott, 334.

<sup>27</sup>Diez and Steans, 138; David Miller, “Cosmopolitanism: A Critique,” *Critical Review of International Social and Political Thought* 5, no. 3 (2002): 80–85.

<sup>28</sup>Murphy, 131.

limit to how much scholars and the truly marginalised could understand each other, let alone work together to achieve emancipation in the terms that the scholars have set.

A similar struggle is also present outside of the Frankfurt School. For instance, feminists and post-colonialists, who seek to reform the IR discipline as a necessary step to changing the real world, are generally removed from the truly marginalised population to whom they claim to give voice, in terms of high educational attainment, employment, race, and education in Western academic institutions. This is exacerbated by the dilemma that the privileged speaking for the truly marginalised is criticised for reinforcing 'discursive hierarchies,' but the often-preferred solution of speaking with them is complicated by the differences in language, dispensable time, worldviews, and so on, which are difficult to surmount.<sup>29</sup> Furthermore, while important progress has been made, scholarship still arguably cannot apply to the experiences of the truly marginalised unless more analyses combine the insights on class, race, and gender. In this sense, praxis seems more aspirational than attainable, at least in the foreseeable future.

Yet, as critique, critical theory has made significant contributions to IR. First, it has provided fierce and prolonged challenges to the assumptions held in mainstream IR, such as the masculine notions of anarchy and self-help, discursive constructions of anarchy and nations, and the use of Western history for universal theories.<sup>30</sup> The final point in particular has contributed to IR's efforts to "provincialise Europe" and to establish a 'Global IR,' as well as to increase "[a]wareness worldwide of the discipline's West-centrism."<sup>31</sup> These in turn have raised awareness of the extent to which marginalised voices are missing from the discipline.<sup>32</sup> Finally, critical theory has shed light on the political implications of theorising by narrowing the gap between what Levine calls the positive, knowledge-building aspect and the critical-reflexive aspect of IR. Levine notes that while IR used to emphasise both aspects as key to theorising around the time of its founding, ultimately "the positive, knowledge-building side of IR flourished, [while] its critical-reflexive side grew only in irregular fits and starts."<sup>33</sup> Given the mainstream IR's focus on the positive, knowledge-

<sup>29</sup>Linda Alcoff, "The Problem of Speaking For Others," *Cultural Critique* 20 (1992 1991): 3, 16.

<sup>30</sup>E.g. Tickner and Sjoberg; Schmid; Seth; David Campbell, *Writing Security: United States Foreign Policy and the Politics of Identity*. (Minneapolis: University of Minnesota Press, 1998).

<sup>31</sup>Seth, 168; Amitav Acharya and Barry Buzan, "Why Is There No Non-Western International Relations Theory? An Introduction," *International Relations of the Asia-Pacific* 7, no. 3 (2007): 287–312; Acharya and Buzan, "Why Is There No Non-Western International Relations Theory? Ten Years On," *International Relations of the Asia-Pacific* 17 (2017): 341–70.

<sup>32</sup>E.g. Murphy.

<sup>33</sup>Daniel Levine, *Recovering International Relations: The Promise of Sustainable Critique* (Oxford: Oxford University Press, 2012), 10.

building side of IR – or problem-solving, in Cox’s words – critical theory has been crucial to supplement this critical-reflexive side.<sup>34</sup> In these three respects, therefore, critical theory has fulfilled its objectives.

Despite its fulfilment, however, there remains an acute need for improvement. For instance, “the persistence of American and Western dominance” in mainstream IR theory, as well as the insufficient incorporation of feminist critique by mainstream IR compared to its efforts for a Global IR, must be better addressed.<sup>35</sup> As for critical theory itself, particularly the Frankfurt School, it must take seriously Schmid’s concern that there is a “progressive uncoupling of normative critique from substantive social and political-economic analysis,” with the former “framed in increasingly rarefied and ideal terms,” and the latter delegated to “functionalist, a-critical theory.”<sup>36</sup> Critical theory urgently needs to recalibrate and reflect on itself, and Cox’s focus on critique and restraint of utopian idealism (by focusing on “feasible transformations of the existing world”<sup>37</sup>) seems useful in that endeavour.<sup>38</sup>

### **Conclusion**

This article has argued that critical theory purports to denaturalise the world, inquire how the world evolved, make normative judgments, and explore possibilities for transforming the world. It has contended that critical theory from the perspective of praxis has not and cannot achieve its objectives, while that in terms of critique has achieved its objectives, albeit with room for improvement. While outside the scope of this article, exploring whether critical theory holds particular salience for the 21st century – in terms of increasing awareness of global inequality, widespread access to discursive politics through social media, and so on – seems a fruitful endeavour for future scholars. 🌸

---

<sup>34</sup>Cox, 128.

<sup>35</sup>E.g. Acharya and Buzan, “Why Is There No Non-Western International Relations Theory? Ten Years On.”

<sup>36</sup>Schmid, 200, 216.

<sup>37</sup>Cox, 130. Emphasis added.

<sup>38</sup>E.g. see Levine for the need for critical theory for more self-reflexivity. Levine, 81.

## References

Acharya, Amitav, and Barry Buzan. "Why Is There No Non-Western International Relations Theory? An Introduction." *International Relations of the Asia-Pacific* 7, no. 3 (2007): 287–312.

"Why Is There No Non-Western International Relations Theory? Ten Years On." *International Relations of the Asia-Pacific* 17 (2017): 341–70.

Alcoff, Linda. "The Problem of Speaking For Others." *Cultural Critique* 20 (1992 1991): 5–32.

Boros, Diana. "Review: Critical Theory and the Challenge of Praxis." *Contemporary Political Theory* 18, no. 1 (2017): 5–7.

Campbell, David. *Writing Security: United States Foreign Policy and the Politics of Identity*. Minneapolis: University of Minnesota Press, 1998.

Cox, Robert W. "Social Forces, States and World Orders: Beyond International Relations Theory." *Millennium* 10, no. 2 (1981): 126–55.

Diez, Thomas, and Jill Steans. "A Useful Dialogue? Habermas and International Relations." *Review of International Studies* 31, no. 1 (2005): 127–40

Freyenhagen, Fabian. "Critical Theory: Self-Reflexive Theorizing and Struggles for Emancipation." Oxford Research Encyclopedia of Politics, 2018.

Haacke, Jurgen. "Theory and Praxis in International Relations: Habermas, Self-Reflection, Rational Argumentation." *Millennium* 25, no. 2 (1996): 255–89.

Horkheimer, Max. "Traditional and Critical Theory." In *Critical Theory: Selected Essays*. New York: Continuum, 1972.

Hammer, Espen. "Critical Theory and the Challenge of Relativism." In *The Routledge Handbook of Philosophy of Relativism*, edited by Martin Kusch, 247–55. London: Routledge, 2019.

Hopgood, Stephen. *The Endtimes of Human Rights*. Ithaca: Cornell University Press, 2013.

Hutchings, Kimberly. "Happy Anniversary! Time and Critique in International Relations Theory." *Review of International Studies* 33 (2007): 71–89.

Keene, Edward. "Methods and Approaches in Critical Theory." Presented at the Research Design and Method in IR Hilary Term, 2021, Oxford, February 4, 2021.

Keenan, Tom. "The 'Paradox' of Knowledge and Power: Reading Foucault on a Bias." *Political Theory* 15, no. 1 (1987): 5–37.

Levine, Daniel. *Recovering International Relations: The Promise of Sustainable Critique*. Oxford: Oxford University Press, 2012.

Linklater, Andrew. "The Harm Principle and Global Ethics." *Global Society* 20, no. 3 (n.d.): 329–43.

Marx, Karl. "Theses on Feuerbach." In *Marx/Engels Selected Works*, translated by W. Lough, 1:13–15. Moscow: Progress Publishers, 1969.

Mearsheimer, John J. *The Tragedy of Great Power Politics*. New York: W.W. Norton Company, Inc., 2001.

Miller, David. "Cosmopolitanism: A Critique." *Critical Review of International Social and Political Thought* 5, no. 3 (2002): 80–85.

Morgenthau, Hans J. *Politics among Nations: The Struggle for Power and Peace*. New York: Knopf, 1954.

Murphy, Craig. "The Promise of Critical IR, Partly Kept." *Review of International Studies* 33, no. 1 (2007): 117–33.

Roach, Steven C. "Critical Theory of International Relations." Oxford Bibliographies, May 29, 2019. <https://www.oxfordbibliographies.com>.

Schmid, Davide. "The Poverty of Critical Theory in International Relations: Habermas, Linklater and the Failings of Cosmopolitan Critique." *European Journal of Inter-*

*national Relations* 24, no. 1 (2018): 198–220.

Seth, Sanjay. "Postcolonial Theory and the Critique of International Relations." *Millennium: Journal of International Studies* 40, no. 1 (2011): 167–83.

Shapcott, Richard. "Critical Theory." In *The Oxford Handbook of International Relations*, edited by Christian Reus-Smit and Duncan Snidal, 327–45. Oxford: Oxford University Press, 2008.

Tickner, J. Ann, and Laura Sjoberg. *Feminism and International Relations: Conversations about the Past, Present and Future*. Abingdon: Routledge, 2011.

Vincent, R.J., and Moorhead Wright. "Special Issue on the Balance of Power." *Review of International Studies* 15, no. 75 (1989): 1.

Waltz, Kenneth N. *Theory of International Politics*. Reading: Addison-Wesley, 1979.

Wohlforth, William C. "Realism." In *The Oxford Handbook of International Relations*, edited by Christian Reus-Smit and Duncan Snidal, 131–49. Oxford: Oxford University Press, 2008.

Yalvaç, Faruk. "Critical Theory: International Relations' Engagement With the Frankfurt School and Marxism." *Oxford Research Encyclopedia of International Studies*, December 1, 2015. <https://oxfordre.com>.

# Is the Concept of Truth Part of Every Proposition?

Mariona E. Miyata-Sturm\*

Faculty of Philosophy, University of Oxford

Philosophers have tried and failed to give an illuminating analysis of the concept of truth for thousands of years. Since we can count uncontested conceptual analyses on one hand this is perhaps not evidence for anything except that very few concepts can be non-circularly analysed. Some, however, think that there is something special about the concept of truth: *truth*,<sup>1</sup> on this view, is not just another hard-to-analyse concept but one of the most fundamental concepts we have, perhaps even underlying all our other concepts (Davidson 1996:265, Asay 2013a:82–3). This view – primitivism about the concept<sup>2</sup> of truth – has gained in popularity over the past two decades and now has a good number of defenders and sympathisers.<sup>3</sup> One of the philosophers driving this new interest in primitivism is Jamin Asay, who, with one book and multiple papers defending the theory, is one of its strongest proponents.<sup>4</sup>

One of Asay’s central arguments for primitivism is the so-called ‘omnipresence argument’, which is (an attempt at) a *reductio* of the claim that *truth* can be non-circularly analysed. Why is it worth our time finding out whether the argument is sound or not? Because if it is, that has serious ramifications both for theories of *truth* and of concepts in general. If sound, it would show that *no* non-circular account of *truth* can be correct, which is to say that most attempts at accounting for *truth* are

---

\* mariona.miyata-sturm@philosophy.ox.ac.uk

<sup>1</sup>I use italics to mention concepts and propositions.

<sup>2</sup>It is quite possible to be a primitivist about the property of truth instead or in addition to accepting primitivism about the concept of truth (as in Merricks 2007), but here I will only be concerned with *conceptual* primitivism.

<sup>3</sup>E.g., Davidson (1990), (1996), Patterson (2010), McGinn (2000), Merricks (2007), Sosa (2001), Asay (2013a), (2013b), (2013c), (2018), (2021). Boghossian (2010), Lowe (2009), and Armstrong (2009) have also expressed their sympathies with primitivism, but, to the best of my knowledge, have neither articulated a full primitivist account nor defended it. See also Moore (1899), Russell (1937), and – in particular – Frege’s “Logic” (1997:227–50) and “The Thought” (1997:325–45) for early versions of alethic primitivism.

<sup>4</sup>See references in the previous note.

doomed to fail. It would also show that some concepts are so fundamental that they are part of every thought we think and statement we make. In this paper, I argue that one of the central premises of Asay's argument – the identity thesis – is insufficiently supported. I start in section 2 by briefly presenting the argument and its significant premises before turning to Asay's arguments for the identity thesis in section 3. There, I argue that his arguments at best lead to a stand-off of judgements about when two statements express the same proposition. I tentatively conclude that since Asay is the one bringing a radical thesis to the table the burden of proof is on him. If that is right, then the stand-offs amount to a defeat of the Identity thesis; if not, it still shows that his defence of the thesis is insufficient to convince those not already convinced of its truth.

## 2. The Omnipresence Argument

The omnipresence argument contains two significant premises: that the proposition *that p* is identical to the proposition *that it's true that p*, and that the concept of truth is part of the conceptual content that composes the proposition *it's true that p* (2013a:166, 2013b:507–8).<sup>5</sup> From this, it concludes that *truth* cannot be non-circularly analysed because the concept is contained in any proposition whatsoever; in particular, in any proposed analysis of *truth*.

In a bit more detail, if we assume that propositions are the content of declarative sentences, the first premise says that the content of a declarative sentence is identical to the content of the same sentence with the truth operator, 'it is true that', added. Let's call this the identity thesis ('Identity' for short). Identity says that, e.g., "Juno orbits Jupiter" and "it is true that Juno orbits Jupiter" express the same proposition even though their surface structures differ. Identity on its own is of course not enough to show that *truth* is part of every proposition, as it could just as well mean that any sentence whose surface structure is 'it is true that *p*' only expresses the proposition *that p* without containing *truth* at all. Asay goes for the first option, i.e., that *truth* is part of every proposition – this is the 'the omnipresence thesis' which gives the argument its name.<sup>6</sup> For Omnipresence to follow from Identity, Asay needs to show that the

<sup>5</sup>Asay uses 'thought' where I use 'proposition'. 'Thought' is dangerously ambiguous between mental states and the content of such states, and as it is clear that it's the latter that Asay is interested in (e.g., 2013a:153) I use 'proposition' to avoid this ambiguity. In his (2021) paper, Asay has presented a different argument that confusingly enough is given the same name where he takes thoughts to be mental states. In this paper, I will focus exclusively on the argument as it is given in his (2013a) and (2013b).

<sup>6</sup>Omnipresence is one of the most central theses of Asay's theory of truth. He has presented different versions of this thesis at different points in time (e.g., 2013a, 2021), but although the content has changed its central place in his theory has not. Here I will only be concerned with his views as they are presented in connection with the 2013 version of the omnipresence argument.



concept of truth is part of *it's true that p*, which is why he also needs the second significant premise. This premise says that *truth* is part of the content of any declarative sentence prefixed with 'it is true that'. In this paper I'm only concerned with showing that Asay has not presented us with sufficient reasons to accept the claim that *p* is identical to *it's true that p*. If that is so, then we ought to reject the omnipresence argument regardless of the truth of the omnipresence thesis.

### 3. The Identity Thesis

What reasons could there be for believing that *p* is identical to *it is true that p*? In this section, I will discuss a handful of Asay's arguments for Identity. We start on the least promising end, in section 3.1, with what I'll call the 'addition' and 'separation' arguments and the substitutivity claim that is left implicit in these and other arguments. I argue that, so far, we are at best left in a stand-off of intuitions about sameness of content. In section 3.2, I turn to a thought experiment of Asay's meant to undermine one of the objections discussed in section 3.1 and argue that it fails to do so. The last argument to be discussed, in section 3.3, brings together ideas from his arguments and thought experiment into a challenge to the opponent of Identity. I argue that this challenge will seem easy to meet for those not already committed to Identity and so we are again left in a stand-off where both sides can accuse the other of begging the question without gaining any headway. I tentatively conclude that since Identity is such a radical thesis the burden of proof is on Asay, which means that it's on him to present us with compelling reasons to accept Identity.

#### 3.1 Addition, Separation and Substitutivity

Asay uses what we may call the omnipresence thesis of existence that both Hume and Kant held to introduce this argument (2013a:154, 2013b:508–9). The omnipresence of existence has it that thinking of something as existing and simply thinking of that thing is the very same act. "Similarly, if I ask you to consider whether kangaroos live in Australia, and then to consider whether it's true that kangaroos live in Australia, I have not asked you to do two separate things. Hence, just as adding existence to an idea adds nothing, so too does adding truth to a thought add nothing" (2013a:154–5). This is clearly meant to generalise to all acts involving propositions, such as expressing, grasping, hoping, denying, asserting, believing, presenting, judging, etc., *that p*.

That was the addition argument. The separation argument is something of a mirror image of the addition argument: if *p* and *it's true that p* are *not* identical, then it should be possible to token (express, entertain, grasp, think of, etc.) one without tokening the other. They should be separable. It should be possible to take conflicting

attitudes towards them: you should be able to believe that it's raining without believing that it is true that it's raining, and hope that it is true that there is life on other planets without hoping that there is life on other planets. But (or so the argument goes) it doesn't seem possible to separate them in this way, so  $p$  and *it's true that*  $p$  are identical (Asay 2013b:509; see also Frege 1997:328).

These arguments come dangerously close to simply *stating* Identity rather than arguing for it: in the addition argument, for instance, we are just told that our two kangaroo thoughts amount to the same thing. But is this claim true? Well, it doesn't make a difference to the truth-value of what we're saying whether we use, say, "Juno orbits Jupiter" or "it is true that Juno orbits Jupiter": either both are true or both are false.<sup>7</sup> That in itself is not enough to show that the sentences express the same proposition, but if we add to this the claim that propositions that are substitutable in this way are identical then we have an argument for Identity. Asay implicitly relies on such a substitution claim: for if the possibility of substitution did *not* indicate sameness of content, then it is irrelevant to the content of, e.g., the kangaroo sentences that it doesn't make a difference which one we use (that is, that they are substitutable). Why believe that substitutable expressions are identical in content, and not, say, necessarily equivalent but distinct? Here is one reason. If substitution is always permissible, *including in belief ascriptions and other intensional contexts*, then there is nothing to distinguish the contents of the two expressions. If there's nothing to distinguish them then we have no reason to postulate that the expressions have distinct content, and since we ought to avoid postulating superfluous entities, we should assume that they express the same content.<sup>8</sup>

---

<sup>7</sup>At least not in cases like this: but what if we have a truth-value gap? Let  $p$  be a gappy proposition – say, *that it's raining* when it's drizzling and not clear whether it amounts to rain or not. Then it seems like *it's true that*  $p$  will be false, since  $p$  is not true, and thus that we have a counterexample to Identity. Whether there are any truth-value gaps at all is controversial, though, and here I don't have the space to go into this, but it seems that if one accepts gappiness then one should reject Identity.

<sup>8</sup>In general, I think such an argument from simplicity makes much sense. However, in this case it's clear that it is more of a motivating consideration than a conclusive argument. Consider what would follow if it were indeed the case that substitutivity always holds. From this fact alone, are we entitled to infer that the relevant propositions are identical? No, because propositions are theoretical entities whose individuation depends on much more than speakers' judgements about the content of statements. Propositional individuation depends, for instance, on the theoretical utility of the larger framework within which it is embedded; that's why possible world-propositions are still a live option even though they are much too coarse-grained to conform to speaker' intuitions about content. If this is correct, then it's possible that we can always substitute one expression for another without any change in truth-value, but that the two expressions nonetheless have distinct content. A full assessment of this objection requires going through different theories of propositions and the various individuation criteria which, implicitly or explicitly, underlie these accounts. That would take us too far afield, so I'll leave this objection as a challenge to proponents of the substitution claim: perhaps it can be worked around rather easily, perhaps not, but to show either requires more work.

Let's get back to the addition and separation arguments. What happens if we embed the sentences in belief ascriptions or other intensional constructions, as in 'Asay believes that...'? Asay is not alone in believing that this doesn't make a difference: Crispin Wright, for instance, has expressed a similar view:

[it] would seem to be an absolutely basic and constitutive characteristic of the notion of truth, that P and "It is true that P" are, as it were, attitudinally equivalent: that any attitude to the proposition that P – belief, hope, doubt, desire, fear, etc. – is tantamount to the same attitude to its truth. (2001:782, fn. 1)

Many will probably agree with Asay and Wright that *it's true that p* and *p* are 'attitudinally equivalent'.<sup>9</sup> If most competent speakers do, that is good support for this claim. But the radicalness of the identity thesis should not be underestimated: it is the claim that every single proposition contains *truth* as an (often) unarticulated constituent. A particularly implausible upshot of this thesis is that even very small children and others with very limited cognitive capacities possess the concept of truth and that they token it whenever they token a proposition. Here is William Alston:

Isn't it clear that a toddler could carry on a conversation ... without being able to claim, as implicitly as you like provided it is a genuine claim, that the propositional content of the original utterance is true, or even that any proposition is true, lacking at that point the conceptual development necessary for making such a claim?  
(2007:13)<sup>10</sup>

The claim about claims is not necessary: it may be that the child has just not learnt the word 'true' or its synonym in her language.<sup>11</sup> We can still raise the question: isn't it clear that a toddler could (say) think that she wants porridge for breakfast without thinking that it's true that she wants porridge for breakfast, or believe that Lisa knows that the ball is in the blue basket without believing that this is true? This seems more plausible to me than assuming that she unwittingly tokens *truth* from the

<sup>9</sup>However, note that this 'attitudinal equivalence' is weaker than Identity: it might be that we cannot help but take the same attitude towards *p* and *it's true that p* without them being identical (compare: we might always take the same attitude towards simple mathematical equations and logical equivalencies – say,  $2+2=4$  and  $3+4=7$ , or *p* and *not not-p* – without thereby claiming that they are identical). To get Identity one needs an additional argument: for instance, an argument from substitution and simplicity, as outlined in the previous paragraph.

<sup>10</sup>Künne presents a similar critique of Identity (Künne 2003:51) (I discuss Asay's response to it below). See Künne (2003:42–52) for a relevant discussion and references to others who reject Identity on a similar basis.

<sup>11</sup>This assumes that concept possession can come apart from word mastery. If that's not possible then Asay's argument clearly fails as there are people who don't know words like 'true' in any language.

very first time she tokens a proposition and continues to do so whenever she entertains any proposition at all.

Also, a good case can be made for the claim that grasping *truth* requires a more sophisticated conceptual machinery than what most children have, as the concept is on the face of it similar to abstract concepts such as *justice* and *time*; concepts we don't expect toddlers to grasp.<sup>12</sup> If possessing *truth* requires more conceptual sophistication than what is required for tokening simple propositions then we cannot always substitute *p* for *it's true that p*, contrary to what Asay claims. One might object to this that some (fundamental) concepts underlie many of our other concepts such that grasping the less fundamental concepts require grasping the more fundamental ones, and that truth is one such fundamental concept. This is indeed a common view in the primitivist literature.<sup>13</sup> The thought is then that even toddlers possess *truth*, as evidenced by their possession of other, less fundamental concepts. In response to this we can point out that a concept might be fundamental without being part of *every* proposition, so this does not in itself support either Identity or substitutivity. Even if possession of *truth* is required to entertain more propositions than is indicated by the use of words like 'true', it does not follow that each and every proposition contains the concept.

A more serious worry one might have about my objections is that they trade on a confusion between different accounts of what propositions are and, in particular, how fine-grained they are.<sup>14</sup> For is it not the case that in one legitimate sense of 'proposition', *p* and *it's true that p* are trivially identical, while in another, equally legitimate sense they might come apart? On a coarse-grained account of propositions, 'two' propositions with the same extension count as one. So, for instance, on a possible world account, the proposition *that p* is just the set of worlds where *p* is true, so the proposition expressed by (say) "triangles have three sides" is the same as the proposition expressed by "triangles have three angles". On such an account, *p* and *it's true that p* are identical, as the proposition is just the set of worlds where *p* is true. On a fine-grained account, however, two extensionally equivalent sentences can express two different propositions depending on the 'mode of presentation' of what it is about (or, the cognitive significance of its constituent terms). For example, on

---

<sup>12</sup>Though it is quite possible that this aura of being a difficult concept is an artefact of being a concept long discussed by philosophers, rather than being a concept discussed by philosophers because it is hard to grasp.

<sup>13</sup>See, e.g., Davidson (1996), McGinn (2000), Asay (2013); see also Strawson (1992) and Chalmers (2011) for two ways of arguing for fundamental concepts, and Miyata-Sturm (2019: ch. 3) for an overview and discussion of this way of arguing for primitivism about truth. Thanks to an anonymous referee of this journal for raising this worry.

<sup>14</sup>Thanks to an anonymous referee of this journal for pushing me to address this objection.

a (neo-)Fregean account the sentences “Jamin writes about truth” and “Asay writes about truth” express two different propositions as someone who does not know that ‘Jamin’ and ‘Asay’ refer to the same person might understand or accept the one but not the other. Here  $p$  and *it’s true that p* can indeed come apart, as someone who doesn’t possess *truth* might take conflicting attitudes towards the propositions. The objection, then, is that Asay might simply opt for a coarse-grained account of propositions, for against such an account my objections have no force.

But this line of reply doesn’t seem to be open to Asay. First, he claims that he can be neutral on the nature of propositions (2013a:27), which shows that he intends Identity to hold regardless of what account of propositions one favours. However, and this leads us to the second point, the way he presents the omnipresence argument that Identity is part of is incompatible with (at least) the most prominent accounts of coarse-grained propositions, namely possible worlds and (neo-)Russellian accounts. In the omnipresence argument, Asay claims that *truth* “is a part of the conceptual content that composes the thought that it is true that  $p$ ” (2013a:166), but if propositions are sets of possible worlds, then *truth* isn’t a part of the proposition expressed by, say, “it’s true that Hanna misses Casper” – this is just the set of all worlds where she misses Casper. Also, on a (neo-)Russellian account propositions are composed of objects, properties, and relations and not of concepts; in particular, they don’t contain *truth*.<sup>15</sup> (Asay is a deflationist about the *property* of truth, so he would not be willing to change his account to fit the Russellian mould.)

So what, you might think, if his account is incompatible with accounts taking propositions to be coarse-grained and not composed of concepts: what about taking them to be coarse-grained *and* composed of concepts? Well, one of the reasons for talking about concepts in the first place is to capture how we think and talk, as opposed to facts about the objects and properties we think and talk about, and that requires staying at roughly the same level of grain as we do when we navigate the world. For instance, Frege introduced the level of sense, or mode of presentation, to make sense of the fact that identity statements are in fact often informative (although they shouldn’t be if they expressed coarse-grained propositions). Whatever concepts are taken to be – and here there is not much agreement, with claims about their nature ranging from definitions via abilities to prototypes<sup>16</sup> – they must be fine-grained enough to make sense of how people perceive the world and of our lim-

<sup>15</sup> See Miyata-Sturm (2019: ch. 2) for a fuller discussion.

<sup>16</sup> See Margolis and Laurence (1999, 2014) and Machery (2009) for good overviews of theories of concepts. The extreme variety of theories of concept is in itself a strong reason to be sceptical of concept-talk: it is just not clear what we are talking about when we talk about concepts. I think this is an important point but have unfortunately had to set it aside here due to space limits (with the regrettable effect that I’m contributing to this woolly concept-talk).

ited knowledge. Any account of concepts coarse-grained enough to guarantee the truth of Identity would also make it the case that, e.g., “Louis Lane believes that Clark Kent is Superman” is true even though she would be surprised to learn that the man she thinks of as Clark Kent is the same as the man she thinks of as Superman.<sup>17</sup> This tension can be captured by saying that she has one concept of Clark Kent and a distinct one of Superman, though that is not available to someone who takes content to be coarsely individuated. Thus, adopting a coarse-grained account to make Identity come out (trivially) true would undermine the purported benefits of talking about concepts in the first place.

So, Asay’s account is on the face of it not compatible with coarse-grained accounts of propositions or concepts and modifying his account to allow for them is unlikely to help.<sup>18</sup> We are back in our stand-off where some, like Asay, claim that *p* and *it’s true that p* are identical and others, like Alston, claim that they are not identical. A potential way out of this situation is offered by Asay in a thought experiment designed to show that objections like the one by Alston fails, which is what we now turn to.

### 3.2. The Thought Experiment

In response to an objection by Künne which is similar to Alston’s, Asay presents a thought experiment meant to show that everyone who can believe, assert, deceive, etc., must possess truth (Künne 2003:51, Asay 2013b:511–3). If so, and even if Identity is false, *truth* is somehow involved every time we believe, assert, etc.<sup>19</sup> It follows

<sup>17</sup>Unless there are some principled reasons for why truth is a special, coarse-grained concept among fine-grained ones.

<sup>18</sup>What about relativizing propositional individuation to the perspective of particular persons? Instead of saying that, say, “Alex believes that Asay writes about truth” express a different proposition from “Alex believes that Jamin writes about truth” because someone (most notably Alex) could take different attitudes towards the propositions, we might say that the sentences express different propositions depending on whether the person entertaining it/them actually takes different attitudes towards them. So, if Alex knows that ‘Asay’ and ‘Jamin’ refer to the same person then the two sentences express the same proposition *from his perspective* but they will simultaneously express two distinct propositions from the perspective of Abdul who believes that Alex believes that ‘Asay’ and ‘Jamin’ refer to two people and that at least one of them doesn’t write on truth. The idea is that from the perspective of someone who possesses the concept of truth, *p* and *it’s true that p* are identical, though they are not identical *tout court*. But this line of reply is clearly not helpful to Asay, as it wouldn’t show that all propositions, even those entertained by people apparently without the concept of truth, contains that concept. To put it differently, it would not show that *truth* is part of all propositions considered from all perspectives, which is what perspective-relative Identity would amount to, but at most that from the perspective of those who already possess the concept of truth, *p* and *it’s true that p* are identical. Thanks to an anonymous referee for asking me to address this.

<sup>19</sup>The list of acts and attitudes provided by Asay is open-ended, but I think we can safely assume that he intends to include every propositional attitude and act whose content is a proposition.

that even the toddler possesses the concept of truth. I do not think that the thought experiment succeeds in establishing that; here is why.

Consider two persons, A and B, who are stuck on a remote island “with only a coin to toss” (2013a:160). Their language contains no semantic vocabulary, but it contains the sentences (H) “the coin landed heads” and (T) “the coin landed tails”. A tosses the coin, hiding the result from B. A “typically utters (H) when the coin lands heads, and (T) when it lands tails, and always shows the result to [B] afterwards” (2013a:160–1). But sometimes A utters (T) when the coin lands heads, and vice versa, and when she shows it to B they have a good laugh. According to Asay, B “understands what it is for [A] to tell the truth, and [B] fully understands the distinction between truth and falsity as well ... The moral of this story is that the concept of truth is a part of [A]’s and [B]’s cognitive lives, in spite of ‘true’ not being a part of their linguistic lives” (2013a:161). Furthermore,

[t]he concept of truth enters their mind not when ‘true’ does, but when they form any thought at all ... The case ... shows that it is appropriate to attribute to them a concept of truth (as revealed by their ability to engage in belief, assertion, deception, etc.) in spite of having no truth predicate in their language ... [T]heir behavior involves the kinds of activities that are indicative of possessing TRUTH, activities that they could not perform without having the concept. (*ibid.*)

I have three problems with this story. First, the claim that they react to amusing deception the same way we (i.e., possessors of *truth*) do is question-begging. Asay assumes that A and B would react the way we would and that these reactions would reveal possession of truth, but that is exactly what’s in question here. The conclusion Asay draws from the case – that the acts of people without any alethic vocabulary reveals that they possess *truth* – is contained in the description of it, so we cannot conclude that A and B possess it.

Second, it could be the case that B laughs when A says (H) even though the coin landed tails because she finds it funny that A says “the coin landed heads” even though the coin landed tails, and similarly for when A says “the coin landed tails” even though it landed heads (and not because she finds it funny that A tells a lie). There is nothing here necessarily involving more than a deflationary account of *truth*: *truth* may enter the picture only as a generalization device for when explaining what all these cases that B finds funny have in common. Thus, we cannot conclude that laughing at *p* when one can see that *not-p* reveals possession of a substantive concept of truth.

Third, the thought experiment doesn’t generalise the way Asay needs it to. Even if it showed that one must possess *truth* to lie, it doesn’t show that “an ability to believe, judge, assert, deceive, deny, etc.” (2013a:162) requires it. It may be that deception is

an especially sophisticated ability (which is very plausible, as it requires understanding both the distinction between something and a (false) representation of it and that another's point of view is distinct from one's own). In any case, lying is closely connected to the distinction between truth and falsity, so it should not be surprising if it requires possession of *truth* even if other acts and attitudes do not. So, at the very least Asay would need to say more about why potentially less sophisticated attitudes like believing or wanting and acts like stating and questioning involve *truth*, and how the thought experiment supports this further claim.

So far, we have seen examples which on the face of it show that it is not always possible to substitute *p* from *it's true that p* and seen that the thought experiment doesn't give the intended support to Identity. Nevertheless, Asay can deny – and has denied – that examples such as Alston's toddler are coherent descriptions of the world (we come back to this below), regardless of the status of his thought experiment. So even if what I have argued so far is correct *and* the arguments presented in the next section require substitution, they might still be available to Asay and so are worth discussing.

### 3.3 Explanatory Difference and Belief Ascriptions

A few of Asay's arguments for the identity thesis are based on the alleged 'empirical indistinguishability' between what is said by *it's true that p* and *p*, and relatedly, the 'explanatory indifference' of saying either that someone believes *that p* or *that it's true that p* (2013a:156–8, 2013b:509–10). If Identity were false, or so the argument goes, then believing that *p* would be a different mental state from believing that *it's true that p*. If so, we should find an empirical difference between these states, and which state we invoke should make a difference to our explanations. The upshot, as Asay sees it, is that

[t]o deny omnipresence, one must articulate some possible phenomenon that could only be accounted for by appeal to the thought that *p*, but not the thought that it is true that *p* (or vice versa). There could never *be* any situation where only one of those was true, so the phenomenon would have to involve someone's taking a cognitive attitude toward one of them but not the other. ... Could there be someone who believes that *p* but not that it is true that *p* [at the same time]? I do not see how. ... [S]o long as I am inclined to describe a person as believing that *p*, I am equally inclined to describe that person as believing that it is true that *p*. (2013a:157–8; italics original)

We can separate this into two challenges to the opponent of Identity: i) she must show that there can be some explanatory difference between saying that *p* and that *it*



is true that  $p$ , and ii) she must show that it is coherent to have or ascribe conflicting attitudes towards  $p$  and *it is true that*  $p$ . If these challenges cannot be met, then, arguably, we should accept Identity or else we unnecessarily inflate propositional content by claiming that there are two propositions where really there is only one.

The first challenge is to show that there can be some explanatory difference between saying that  $p$  and that *it's true that*  $p$ . Suppose that Alfred is against the death penalty. To explain this, I say that Alfred believes that murder is wrong, whereas you say that Alfred believes that it's true that murder is wrong. Can there be any explanatory difference between your explanation and mine? Well, everyone who believes that there is more to *it's true that*  $p$  than simply  $p$  thinks that an explanation by means of the first is different from one made by means of the latter.<sup>20</sup> Perhaps the opponent of Identity believes that saying *it's true that*  $p$  is to say that  $p$  has a truth-maker, whereas saying  $p$  is not – that is simply saying whatever  $p$  says. On this view, you attribute to Alfred the belief that the proposition *murder is wrong* has a truth-maker, whereas I attribute to him simply the belief that murder is wrong. Or perhaps the opponent of Identity believes that saying that *it's true that*  $p$  is to say that  $p$  corresponds to a fact. If so, you attribute to Albert the belief that *murder is wrong* corresponds to a fact, whereas my attribution doesn't make any such reference to any relation Albert may believe to hold between a proposition and what it represents. And so on, for all the explanations that one can give in terms of a substantive (i.e., non-deflationary) concept of truth *given that one doesn't already accept Identity*. In these cases, there is an alleged explanatory difference between saying that  $p$  and that *it's true that*  $p$ . Most, if not all, who deny Identity but believe that the concept of truth is explanatorily potent – that is, almost everyone except for deflationists and primitivists – will take *it's true that*  $p$  or any other explicit truth-ascription to have more explanatory potential than  $p$  alone. To deny this is to deny that truth-making and correspondence, or anything else one may invoke to explicate what's going on in truth-ascriptions, play any explanatory role.

But what Asay is asking is for his opponent to show in what way  $p$  and *it's true that*  $p$  are explanatorily different: what can *only* be explained by invoking one and not the other? The examples given in the previous paragraph suggest some potential explanatory differences, e.g., between simply expressing a proposition and making the further claim that the proposition has a truth-maker or that it corresponds to a fact. The question is one of who has the burden of proof: is it on Asay to show that explanations in terms of *it's true that*  $p$  say no more than explanations in terms of  $p$ , or is it on the opponent of Identity to show that the explanations are distinct? It's

---

<sup>20</sup>The degree of the presumed difference between the explanations is arguably equal to how substantive one takes *truth* to be.

unlikely that there is a definite answer to who has the burden of proof, but a good rule of thumb is that the person presenting the most controversial view has the burden on her side.<sup>21</sup> If so, the burden of proof is on Asay, all the while he claims that *truth* is necessarily possessed by everyone who can token a proposition, that expressions with very different surface structure nevertheless have the same content, and that *truth* is part of each and every proposition. This is of course not a knockdown argument, but I do think Asay has to do more to show that an explanation of the form *it's true that p* and an explanation of the form *p* cannot differ in their explanatory potential than simply questioning whether that is so.<sup>22</sup>

The second challenge Asay poses to the opponent of Identity is to show that it can be coherent to have or ascribe different attitudes towards *p* and *it's true that p*. Say that I believe that David Lewis was creative: is it then possible for me to disbelieve – at the same time – that it's true that he was creative? Could the claim that I believe that *p* and disbelieve that *it's true that p* (at the same time) ever be a coherent description of my mental states? Alston's toddler example shows, I believe, that it can indeed be coherent to ascribe conflicting attitudes towards *p* and *it's true that p*: perhaps two-year-old Ayumi believes that Miuna took her Lego without believing that it's true that Miuna did so. Here is another example: Kevin Scharp (2013) argued that *truth* is inconsistent and ought to be replaced by two other *truth*-like concepts, each of which would conform to half of the rules we expected *truth* to obey but that the alethic paradoxes allegedly show cannot be satisfied by one concept in combination with classical logic and a self-referential language. These two concepts are *ascending true*, which is constituted by the rule 'if *p* then <*p*>', and *descending true*, constituted by 'if <*p*> then *p*'. If Scharp or anyone else follows his advice then we can coherently describe them as thinking *that p* without thinking *that it's true that p* (but instead, e.g., that it's *ascending-true that p*). I see no good reason to deny that it is possible to replace *truth* as there is no obvious inconsistency in revising, regimenting, or precisifying concepts (compare how we may decide that the concept of red only covers *those* nuances or decide that *woman* also applies to trans women).<sup>23</sup>

<sup>21</sup>See, e.g., Williamson (2011), Pigliucci and Boudry (2014), and Rescorla (2009) for interesting discussions of burden of proof-shifting.

<sup>22</sup>It can, of course, be countered that this begs the question against Asay. This amounts to the same thing as saying that the burden of proof is on those who claim that *it's true that p* and *p* can be explanatorily different, and to this I can only reiterate what I just said: it's reasonable to put the burden of proof on the person with the most radical thesis, and in this case that person is Asay. Both claims can be challenged – perhaps we should put the burden on the least controversial thesis as we already have a *status quo*-bias, and perhaps it's less radical to think that each and every proposition contains truth than that *it's true that p* and *p* can come apart – but that does not seem promising to me.

<sup>23</sup>There is now a large and growing literature on 'conceptual engineering', or the evaluation and replacement of concepts, and there are many open questions here that has a bearing on whether one can change

If that is right, then the second challenge Asay poses to those who deny Identity is met. Asay may, of course, argue that even though Scharp believes that he can think that  $p$  but not that *it's true that*  $p$ , he is deluding himself: there *just is no difference* between the thoughts, so anyone who believes she can entertain one without the other is wrong. Also consider the claim that children can token a proposition without possessing *truth*. Against this Asay may argue that even though Alston thinks that children can, say, think that they are hungry without thinking that it's true that they are hungry, he is simply wrong – there just is no difference between the thoughts.

At best, such replies lead directly to the now-familiar stand-off between the proponent and opponent of Identity. Sharp and Alston think that they can separate  $p$  and *it's true that*  $p$ ; Asay thinks that they cannot. It is doubtful that we can resolve such a stand-off without introducing principled criteria for propositional individuation (which, unfortunately, Asay doesn't discuss). Thus, as it stands, this argument for the identity thesis is at the very best inconclusive. At worst, Asay has simply failed to provide strong enough arguments in favour of it. Regardless, I think we can safely conclude that one of the central premises of the omnipresence argument is insufficiently supported.

#### 4. Conclusion

I have argued that Asay's identity thesis is insufficiently supported. We have seen that his addition and separation arguments and the substitution claim that they rely on can be challenged, that Asay's thought experiment doesn't support the conclusion he draws from it, and that the challenges he poses to the opponent of Identity can be met. Crucially, however, Asay can and has denied the counterarguments and so long as the basis on which to decide who is right is judgements about content unrestrained by principles of propositional individuation, we are left in a stand-off where both sides can accuse the other of begging the question. I briefly argued that the burden of proof is on Asay as he brings the most radical thesis to the table, and if that is correct then we should reject Identity and thus the omnipresence argument of which it is a central premise. If that is not correct, I believe that I have still showed that Asay's defence of Identity is unconvincing and that we might not be able to get further without more principled ways of individuating propositions.<sup>24</sup> 🌟

---

a concept at all – for instance, what a concept *is*, if there are any concepts at all, and if so, how one can intentionally replace one (see, e.g., the papers in Burgess, Cappelen, and Plunkett (2020) and Marques and Wikforss (2020)). Here I will not take a stand on these issues beyond assuming that concepts exist and claiming that it is not given that concepts cannot be changed, which is all we need for present purposes.

<sup>24</sup>Thanks to Conrad Bakka, Herman Cappelen, Max Johannes Kippersund, Ainar Miyata-Sturm, Miriam Natvig, and two anonymous referees of this journal for discussions of the material presented here and comments on earlier drafts.

## References

- Alston, W. P. (2007) "Illocutionary Acts and Truth", pp. 9–30, in D. Greimann and G. Siegart (eds.) *Truth and Speech Acts: Studies in the Philosophy of Language*. New York: Routledge.
- Armstrong, D. M. (2009) "The General Theory of Truth-Making", pp. 115–33, in E. J. Lowe and A. Rami (eds.) *Truth and Truth-Making*. King's Lynn: Acumen.
- Asay, J. (2013a) *The Primitivist Theory of Truth*. Cambridge: Cambridge University Press.
- Asay, J. (2013b) "Primitive Truth", *Dialectica*, 67, 4, pp. 503–19.
- Asay, J. (2013c) "Tarski and Primitivism about Truth", *Philosophers Imprint*, 13, pp. 1–18.
- Asay, J. (2018) "Putting Pluralism in its Place", *Philosophy and Phenomenological Research*, 96, 1, pp. 175–91.
- Asay, J. (2021) "TRUTH: A Concept Unlike Any Other", *Synthese*, 198 (Supplement 2), pp. S605–S630.
- Boghossian, P. A. (2010) "Our Grasp of the Concept of Truth: Reflections on Künne", *Dialectica*, 64, 4, pp. 553–63.
- Burgess, A., Cappelen, H., and Plunkett, D. (eds.) (2020). *Conceptual Ethics and Conceptual Engineering*. Oxford: Oxford University Press.
- Cappelen, H. W. (2018) *Fixing Language: An Essay on Conceptual Engineering*. Oxford: Oxford University Press.
- Davidson, D. (1990) "The Structure and Content of Truth", *The Journal of Philosophy*, 87, 6, pp. 279– 328.
- Davidson, D (1996) "The Folly of Trying to Define Truth", *The Journal of Philosophy*, 93, 6, pp. 263–78. Frege, G. and Beaney, M. (ed.) (1997) *The Frege Reader*. Oxford: Blackwell.

Künne, W. (2003) *Conceptions of Truth*. Oxford: Clarendon Press.

Lowe, E. J. (2009) "An Essentialist Approach to Truth-Making", pp. 201–15, in E. J. Lowe and A. Rami (eds.) *Truth and Truth-Making*. King's Lynn: Acumen.

Machery, E. (2009) *Doing Without Concepts*. Oxford: Oxford University Press.

Margolis, E. and Laurence, S. (eds.) (1999) *Concepts: Core Readings*. Cambridge and London: MIT Press.

(2014) "Concepts" in E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, URL= <<https://plato.stanford.edu/archives/spr2014/entries/concepts/>>.

Marques, T. and Wikforss, Å. (eds.) (2020). *Shifting Concepts: The Philosophy and Psychology of Conceptual Variability*. Oxford: Oxford University Press.

McGinn, C. (2000) *Logical Properties: Identity, Existence, Predication, Necessity, Truth*. Oxford: Clarendon Press.

Merricks, T. (2007) *Truth and Ontology*. Oxford: Clarendon Press.

Miyata-Sturm, M. E. (2019) *End of Analysis, Beginning of Explanation: Making the Case for Alethic Primitivism*. MA thesis, the University of Oslo. [www.duo.uio.no/handle/10852/70481](http://www.duo.uio.no/handle/10852/70481)

Moore, G. E. (1899) "The Nature of Judgment", *Mind*, 8, pp. 176–93.

Pigliucci, M. and Boudry, M. (2014) "Prove it! The Burden of Proof Game in Science vs. Pseudoscience Disputes", *Philosophia*, 42, pp. 487–502.

Rescorla, M. (2009) "Shifting the Burden of Proof?", *The Philosophical Quarterly*, 59, 234, pp. 86– 109.

Russell, B. (1937) *The Principles of Mathematics*. 2nd ed. New York: Norton Company, Inc.

Schärp, K. (2013) *Replacing Truth*. Oxford: Oxford University Press.

Sosa, E. (2001) "Epistemology and Primitive Truth", pp. 641–62, in M. Lynch (ed.) *The Nature of Truth: Classic and Contemporary Perspectives*. Cambridge and London:

MIT Press.

Williamson, T. (2011) "Philosophical Expertise and the Burden of Proof", *Metaphilosophy*, 42, 3, pp. 215–29.

Wright, C. (2001) "Minimalism, Deflationism, Pragmatism, Pluralism", pp. 751–88, in M. Lynch (ed.) *The Nature of Truth: Classic and Contemporary Perspectives*. Cambridge and London: MIT Press.

# Reducing Networks Sizes

*Paul Dubois\**

*Mathematical Institute, University of Oxford*

## **Abstract**

With the recent explosion of the volume of data we produce and analyze, there is a need for techniques to reduce the size of the objects we study with minimal loss of information. The networks that we aim to understand are no exception. Networks are mathematical objects composed a set of nodes, and links between those nodes. Nowadays we have graphs with thousands, sometimes even millions of vertices.

It is therefore crucial to be able to reduce the size of networks, without, or with minimal loss of information. We present how role and block modeling can be used for this purpose.

We will be interested in the study of dynamics on graphs, in particular, the consensus dynamics. We will give a mathematical intuition and demonstrate with experiments that (under some assumptions) the dynamics on a large network are similar to the one from an induced weighted graph.

## **Introduction**

The study of networks became very popular in the recent years, as they can be used to represent phenomena from many scientific disciplines: physics and biology (e.g. protein interactions), computer science (e.g. web routers connections), sociology (e.g. friendship on social networks), economics (e.g. trade relationships between firms, producers-distributors-consumers sales transactions).

The networks that arise from the real world data are, nowadays, typically very large, and their study complicated. In such a context, it is crucial to reduce the calculations on the full network, as they are computationally demanding. We therefore try to model the networks accurately, so that we can extract some information from smaller objects. We will see how this can be done using an "image graph".

---

\*pauldubois98@gmail.com

We will analyze the structural properties of (static) networks. We show that block modeling with role models can be used to reduce our problem to a smaller graph: we can drastically decrease the network size while keeping some properties (in particular, consensus dynamics) of the network unchanged.

## 1. Notations

The vertices (also called "nodes") and edges (or "links") of a network (also called "graph") will be denoted by  $\mathcal{V}$  and  $\mathcal{E}$  respectively (with  $|\mathcal{V}| = N$ ,  $|\mathcal{E}| = M$ ). We will use subscripts for vector and matrix indices. In the case of an unweighted network (i.e. a network where an edge either exists or do not), we use the standard convention  $w_{ij} = 1$  if there is an edge between nodes  $i$  and  $j$ , and  $w_{ij} = 0$  otherwise. In the case of a weighted network (i.e. each edge is assigned an associated value, called "weight"), we will write  $w_{ij} = \omega$  if there is an edge between nodes  $i$  and  $j$  of weight  $\omega$ , and  $w_{ij} = 0$  otherwise. We denote the  $N \times N$  adjacency matrix by  $W = (w_{ij})$ .

Two nodes  $i$  and  $j$  are neighbors if there is an edge  $ij$  between them. The neighborhood  $N_i$  of a node  $i$  is the set  $\{j \in \mathcal{V} \mid ij \in \mathcal{E} \text{ i.e. } w_{ij} \neq 0\}$  of nodes  $j$  sharing a link with  $i$ .

We call classes groups of nodes. Networks will typically be divided in classes, we write  $i \in a$  if the node  $i$  is in class  $a$ , furthermore, we write  $c$  for the vector of classes names<sup>1</sup> (i.e.  $c_i = a$  if node  $i$  is in class  $a$ ). We write  $W_{[a][b]}$  for the entries of the sub-matrix of  $W$  with rows  $i \in a$  and columns  $j \in b$ , that is, the sub-matrix of  $W$  describing the relation of nodes from class  $a$  to class  $b$ .

## 2. Role Model

The concept of role model was introduced by Reichardt and White in [1]. Here, we focus on the inter-dependence of groups of nodes in a network. This is a different approach from the classic community decomposition. In a community analysis, we try to split the network into groups of nodes with a high density of edges inside the groups, and a low density across them. However, a role model decomposes a graph's nodes into classes of nodes that are "equivalent". A class of equivalent nodes may have a low edge density (i.e. there are very few connections among nodes in the class). In some contexts, one may want to group nodes that have no or few links between them, but behave similarly in the network. The typical example (see fig. 1) is the one of producers-distributors-consumers: producers interact mainly with distributors, and very few times with other producers or consumers; similarly, consumers interact mainly with distributors, and very few times with other consumers or producers; and distributors do not often interact with other distributors.

---

<sup>1</sup>Typically, we label classes with numbers.



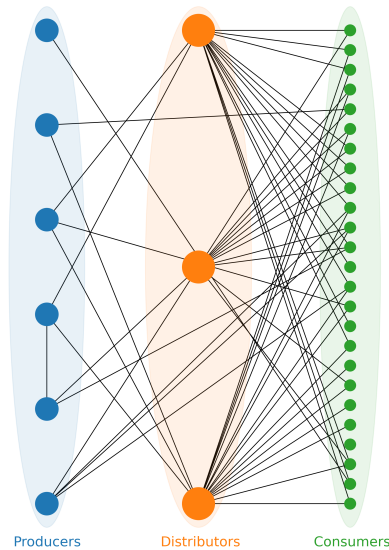


Figure 1: Standard network example of a Producers-Distributors-Consumers architecture.

As defined by Lorrain and White in [2], two nodes are called structurally equivalent if they have the exact same neighbors (i.e.  $i$  and  $j$  are structurally equivalent if  $N_i = N_j$ ). This definition is rather strong, and is almost never achieved in practice. A related concept applicable to a graph endowed with a partition of the set of vertices into classes (introduced by White and Reitz in [3]) is "regular equivalence": two nodes are regularly equivalent if they are connected to the same classes (see fig. 2, inspired from Reichardt and White [1]). That is, the set of classes reached by two equivalent nodes are the same. Formally,  $i$  and  $j$  are regularly equivalent if  $\{c_k \mid k \in N_i\} = \{c_k \mid k \in N_j\}$ . Note that this definition depends on a class attribution.

Eventually, we wish to group every node of regularly equivalent classes as a single node in an "image graph" or "induced graph". This image graph is intended to be (much) smaller than the original one, while still capturing all phenomena happening in the original network. Unlike communities, these nodes classes may have a very low (possibly even zero) internal edge density.

Despite being weaker, regular equivalence is still an exact definition: a single edge placed unpleasantly can completely change the equivalent classes. To thwart this problem, we create a quality function to tell how "good" a given couple of image graph and nodes classes is. This is somehow similar to the modularity approach for

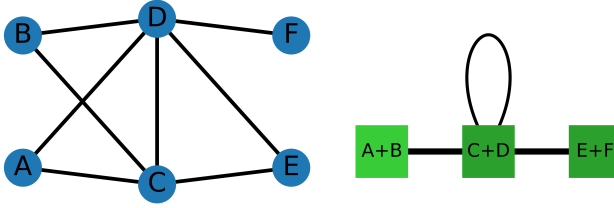


Figure 2: Example network: A and B are the only structurally equivalent nodes, while when paired as on the green model, we have three equivalent classes: AB, CD and EF. the community finding problem. The quality function  $Q$  is proposed by Reichardt and White in [1] It is defined as follows:

The quality of an image graph with roles  $\mathcal{R} = \{1, \dots, q\}$  and  $q \times q$  adjacency matrix  $B$  coupled with role classes  $c : \mathcal{V} \rightarrow \mathcal{R}$  is:

$$Q(B, c) = \frac{1}{M} \left( \sum_{i \neq j} a_{ij} w_{ij} B_{c_i, c_j} + b_{ij} (1 - w_{ij}) (1 - B_{c_i, c_j}) \right).$$

Here,  $a_{ij}$  is the gain (or reward, i.e. the amount by which the quality function  $Q$  increases) for the edge  $(i, j)$  to appear in the image graph as  $(c_i, c_j)$ , and  $b_{ij}$  is the opposite (i.e. the reward for non-edge  $(i, j)$  to not appear as  $(c_i, c_j)$  in the image graph). We then maximize the quality function to get the "best" image graph. Optimization of this quality function is further discussed in [1].

### 3. Stochastic Block Model

Stochastic Block Model (SBM) are used to model matrices that split into submatrices (called "blocks"). If we split a network vertices into classes (groups of nodes), then it makes sense to split the adjacency matrix into blocks accordingly. Far classes  $a$  and  $b$ ,  $W_{[a][b]}$  (defined in section 1) will form one of the blocks of  $W$  in our model. Eventually, the classes we will use will be the group of nodes with the same role, in the sense we discussed in the previous section.

#### 3.a Theory of SBM Modelization.

Following [4], we introduce the stochastic block model (SBM): [SBM] Let  $W$  be an  $n \times n$  matrix of random variables, and consider a partition  $\mathbf{c}$  of the set of indexes. We say that  $W$  is the matrix representing a SBM with respect to the partition  $\mathbf{c}$  if:

- The entries  $w_{ij}$  are independent.

- If  $i, i' \in a$  and  $j, j' \in b$ , then  $w_{ij}$  and  $w_{i'j'}$  are identically distributed.

Now, in our case, we will use the random matrix  $W$  to generate the adjacency matrix of our network. Since we do not study networks with multi-edges,  $w_{ij} \in \{0, 1\}$ , so the  $w_{ij}$  must follow a Bernoulli distribution. We therefore just need to specify a probability to parameterize the Bernoulli process for each pair of distinct classes.

Thus, the SBM for a partition  $\mathbf{c}$  with  $k$  classes is parameterized by a  $k \times k$  matrix  $\Theta = [\theta_{ab}]$ , with  $\theta_{ab} \in [0, 1]$  the probability that an edge occurs from a node of class  $a$  to a node of class  $b$ . So we have  $w_{ij} \sim \mathcal{B}(\theta_{c_i, c_j})$ <sup>2</sup>. Therefore, the adjacency matrix  $W$  is divided into  $k^2$  blocks, namely  $W_{[a][b]}$ ,  $a, b \in \{1, \dots, k\}$ , each being parameterized by  $\theta_{ab}$ ;  $a, b \in \{1, \dots, k\}$ .

We will write  $W \sim \mathcal{SBM}(\Theta, \mathbf{c})$  to denote this setup.

We let  $m_{ab}$  count the number of observed edges in the block  $(a, b)$ :

$$m_{ab} = \sum_{i \in a, j \in b} w_{ij}$$

and  $n_{ab}$  count the number of potential edges in the block  $(a, b)$ :

$$n_{ab} = \begin{cases} |a||b| & \text{if } a \neq b \\ (|a| - 1)|a| & \text{if } a = b \end{cases}$$

So the density of  $y_{ab}$  of edges in the block  $(a, b)$  is  $y_{ab} = \frac{m_{ab}}{n_{ab}}$ .

As we are summing Bernoulli random variables, we have:

**Property 1 (Distribution of  $m_{ab}$  in the SBM model)** *In the previous setup, we have:*  
 $m_{ab} \sim \mathcal{B}(n_{ab}, \theta_{ab})$ .<sup>3</sup>

Similarly, by properties of Binomial random variables, we have:

**Property 2 (Expectations of  $m_{ab}$  and  $n_{ab}$  in the SBM model)** *In the previous setup, we have:*

- $\mathbb{E}[m_{ab}] = n_{ab}\theta_{ab}$
- $\mathbb{E}[y_{ab}] = \theta_{ab}$

<sup>2</sup> $x \sim \mathcal{B}(p)$  means the random variable  $x$  follows a Bernoulli process with parameter  $p$ .

<sup>3</sup> $x \sim \mathcal{B}(n, p)$  means the random variable  $x$  follows a Binomial process with repetition  $n$  and probability  $p$ .

### 3.b Combining with role model.

If we have a network and a role model (i.e. a partition of the set of vertices) we can associate an SBM model to the network, with respect to the partition. Eventually, we will use this SBM model (which is a random graph) to generate a network with a computer (i.e. we take different realizations of the random matrix  $W$  to get concrete adjacency matrices). It is expected to be similar to the original one. Conversely, one can just start with an SBM model and use it to generate a role model and a random networks associated.

In the context of studying the structural properties of a network through role models, we want the groups of nodes from the SBM model to correspond to the different roles that we are able to detect on the network.

When modeling a real life network, we would like the probability parameter matrix to reflect the connections between groups of nodes with the same role. This can be done by setting the probability parameters to the density of edges between role classes (note that density and probability both between 0 and 1, so there is no scaling needed). Such a choice makes the expected number of edges between classes to be the same as the one observed in practice.

Formally, for classes  $a$  and  $b$ , we let  $e_{ab} = |\{(i, j) \mid i \in a, j \in b, (i, j) \in \mathcal{E}\}|$  (i.e.  $e_{ab}$  is the number of edges between nodes from class  $a$  and nodes from class  $b$ ). Then, we set  $\theta_{ab} = e_{ab}/n_{ab}$ . Note that this implies  $\mathbb{E}[y_{ab}] = e_{ab}/n_{ab}$ , i.e. that the expected density of edges is the observed one.

### 3.c Induced Graph.

The goal of this modeling is to reduce the size of the network, to make computations faster, and interpretations simpler.

We will do so by creating an "induced graph": we treat each class as a node. For simplicity, we label the nodes on the induced graph using the class label. For the induced graph to be accurate, we will add a weight to each node and edge. The weight of a node  $a$  in the induced graph will be set to the number of nodes  $i$  in the initial network such that  $i \in a$ . We set the weight of an edge  $(a, b)$  to the number of edges between classes  $a$  and  $b$  in the initial network.

## 4. Consensus Dynamics

We turn our interest to consensus dynamics on our network. This is one of the (perhaps *the*) most well known dynamics on graphs. As one of the most popular multi-agent dynamical system, consensus dynamics can be used in several context. E.g. for individuals' opinion in society (socio-economics context), or for efficient

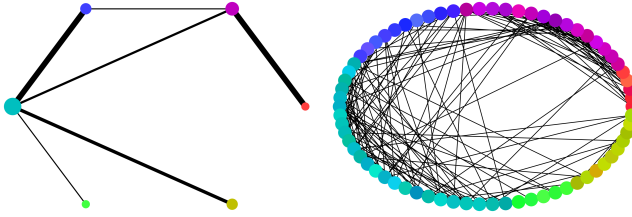


Figure 3: Example network (right hand side) and its induced graph (left hand side). distributive computation of global functions (engineering context). Our goal will be to understand the behaviour of the network, and to model a simplified version.

#### 4.a Definition.

The consensus dynamics attaches a value to each node. These values are evolving in time, according to a linear ordinary differential equation: each node influences and is influenced by its neighbors. The strength of the mutual influence between two nodes is proportional to the difference of their values.

Formally, consensus dynamics is defined as follows: the values associated to each node are stored in a vector  $\mathbf{x}$  which is indexed by nodes of the graph. The dynamics (the evolution of values) is governed by the equation

$$\dot{\mathbf{x}} = -\mathbf{L}\mathbf{x}$$

where  $\mathbf{L}$  is the Laplacian of the network<sup>4</sup>. Component-wise, this becomes

$$\dot{\mathbf{x}}_i = \sum_{j \in \mathcal{V}} W_{ij}(\mathbf{x}_j - \mathbf{x}_i)$$

so the values on each node changes in a way that reduces the difference with its neighbors (on average) [5]. For the states to be well-defined, we also need to specify initial values. In general, this is done using a normal or uniform random variable for each node.

#### 4.b Properties.

Two properties are widely known for this dynamics.

**Property 3 (Global consensus)** *Assuming the network is connected, it will tend to a global consensus (i.e. all nodes having the same value). If there is more than one connected component, a consensus will be reached for each component.*

<sup>4</sup>The Laplacian matrix  $L$  is a matrix representation of a graph. It is defined as  $L = D - W$  where  $D$  is the diagonal matrix of degrees, and  $W$  is the adjacency matrix.

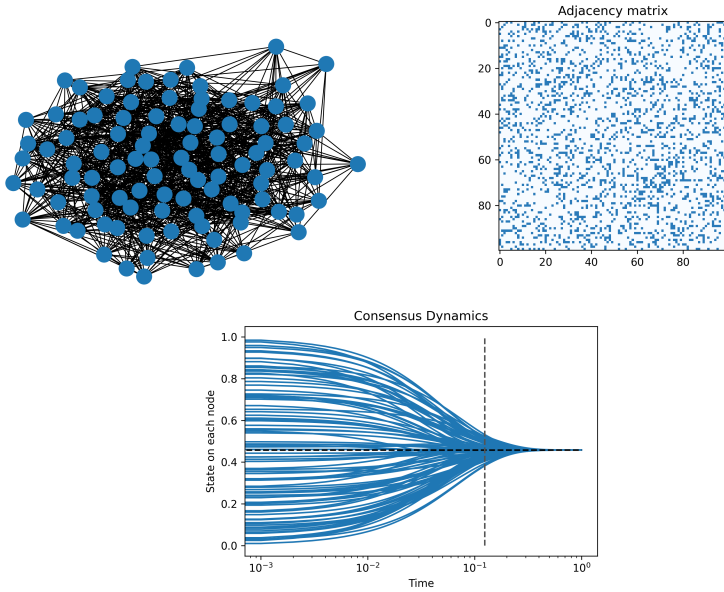


Figure 4: Standard network example: global consensus.

Considering the graph component by component, each node adjusts its value to reduce the difference to its neighbors. Eventually, this difference tends to zero. The effects are symmetric, so the sum of values on all nodes is constant over time. Thus, all nodes tend to the final state  $x^* = \frac{1}{|\mathcal{V}|} \sum_{i \in \mathcal{V}} \mathbf{x}_0$ . This is derived in more detail in [5].

This is well illustrated with the following network simulation (fig. 4). Here, we simulate the consensus dynamics on a graph taken at random following the Erdős-Rényi model with 100 nodes and probability 0.2. Each node was initialized with a uniform random variable on  $[0, 1]$ . The black dotted line is the value of the global consensus, while the gray dotted line is the time where we consider consensus is reached.

**Property 4 (Local consensus)** *If the network studied has some communities structures (i.e. some subsets of nodes have a high edge-density compare to the rest of the graph), then a consensus local to each community will be reached before the global one<sup>5</sup>.*

<sup>5</sup>The time-scales of the process are dictated by the eigenvalues of the Laplacian matrix, local consensus in  $k$  communities is reached near time  $t = 1/\lambda_{k+1}$ , while global consensus will be reached near  $t = 1/\lambda_1$  (with  $\lambda_i$  eigenvalues of the Laplacian in increasing order).

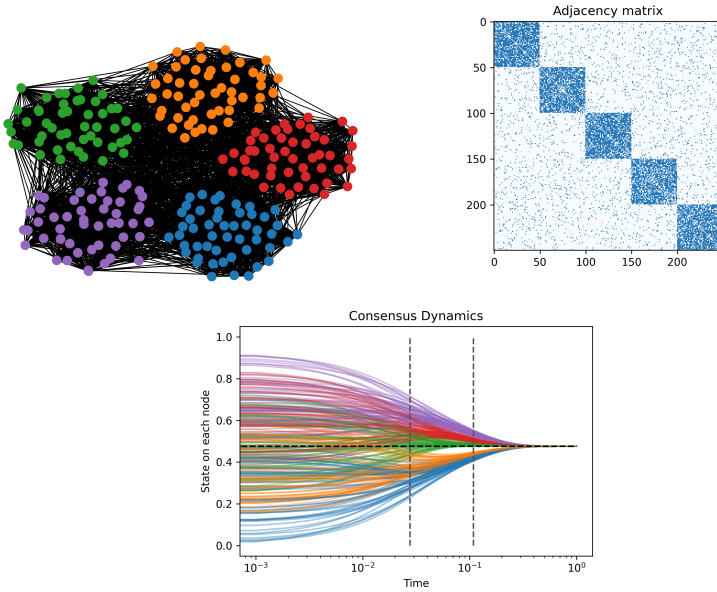


Figure 5: Communities network example: local consensus.

This is derived in [5], using time-scale separation in networks with consensus. Once again, we illustrate this with a network simulation (fig. 5). Here, we simulate the consensus dynamics on a graph taken at random following a SBM with 5 groups of 50 nodes, having external edge density 0.05 and internal edge density 0.75 (i.e.  $\theta_{ab} = 0.05$  for  $a \neq b$  and  $\theta_{aa} = 0.75$ ). Each node was initialized with a uniform random variable on  $[0, \frac{1}{2}]$ , shifted by  $0.125k$  ( $k \in 0, 4$  labels the groups). This small change is used to shift each local consensus so that we can distinguish them (otherwise, all local consensus converge to the same value). The black dotted line is the value of the global consensus, while the gray dotted line is the time where local consensus is reached. Note that eventually, a global consensus (property 3) is still reached.

#### 4.c Relation with Previous Theories

Here, we consider a network  $G$  with classes vector  $\mathbf{c}$ . We are interested in the dynamics in the case where the nodes within the same class share the same, or a very similar initial value. This is a rather strong assumption about the initial state. However, if the classes are derived from roles, then it is justified to assume that nodes with the same role start with similar values. In this particular setup, two patterns appear. We will

give some mathematical intuitions, and demonstrate them through some numerical experiments.

#### 4.d About edges density inside role classes

In the context of a role model, it is clear that if the role classes have high intra-edges density, and that we initialize the nodes with a local community consensus, then this local consensus should remain, as it appears naturally even from a random initialization (as seen in property 4).

Now, if the role classes have a low intra-edges density, we could predict that edges with other classes may give rise to different community structures. In that case, these different new communities will reach a local consensus, breaking the structured initialization that we enforced.

However, this is not what happens in practice (as we will see in section 5, the dynamics is only marginally changed for dense versus sparse classes). The mathematical intuition is the following: suppose all the nodes within a class had the exact same value. We then calculate the dynamics on one node, say  $i \in a$ :

$$\begin{aligned}
 \dot{\mathbf{x}}_i &= \sum_{j \in \mathcal{V}} w_{ij}(\mathbf{x}_j - \mathbf{x}_i) \\
 &= \sum_{j \in N_i} w_{ij}(\mathbf{x}_j - \mathbf{x}_i) \\
 &= \sum_{j \in N_i^E} w_{ij}(\mathbf{x}_j - \mathbf{x}_i) + \sum_{j \in N_i^I} w_{ij} \underbrace{(\mathbf{x}_j - \mathbf{x}_i)}_{\bar{\mathbf{x}}_a - \bar{\mathbf{x}}_a = 0} \\
 &= \sum_{j \in N_i^E} w_{ij}(\mathbf{x}_j - \mathbf{x}_i)
 \end{aligned}$$

Remember  $N_i$  is the set of neighbors of  $i$ , and  $d_i$  is the degree of  $i$  (i.e.  $d_i = |N_i|$ ). Moreover, we will split  $N_i$  into  $N_i^E$  and  $N_i^I$ :

- $N_i^E$  is the neighborhood of  $i$  outside of its class ("exterior neighborhood").
- $N_i^I$  is the neighborhood of  $i$  inside of its class ("interior neighborhood").

Accordingly,  $d_i^E = |N_i^E|$  is the exterior degree, and  $d_i^I = |N_i^I|$  is the interior degree.

In the case of an unweighted network,  $W_{ij} = 1$  if  $j \in N_i$ , so

$$\begin{aligned}
 \dot{\mathbf{x}}_i &= \sum_{j \in N_i^E} \mathbf{x}_j - \mathbf{x}_i \\
 &= -d_i^E \mathbf{x}_i + \sum_{j \in N_i^E} \mathbf{x}_j
 \end{aligned}$$



This means that for each node  $i$ , only the relations with the exterior of its class contribute to the dynamics. Now, nodes within the same class  $a$  have nearly equivalent neighbors. This means that  $N_j^E$  is different for each  $j \in a$  (which makes nodes states within the same class to eventually differ, despite the equal initialization). However, they are very similar, so states of nodes within the same class will not differ a lot.

Note that the calculations above are in expectation<sup>6</sup> exact if nodes states have the same expected values class-wise. That is, if at time  $t$ ,  $\mathbb{E}(\mathbf{x}_i^t) = f(\mathbf{c}_i)$ <sup>7</sup> (for example with  $\mathbf{x}_i^0 \sim \mathcal{N}(f(\mathbf{c}_i), \sigma_i)$ <sup>8</sup> at  $t = 0$ ), then  $\mathbb{E}(\dot{\mathbf{x}}_i^t)$  can be computed using only exterior neighbors as above. Thus, for  $\tilde{t} \geq t$ , the above calculations will remain in expectation.

#### 4.e About the dynamics on the induced graph

We now consider a large network  $H$  with classes given by  $\mathbf{c}$ , and induced graph  $G$  (with respect to  $\mathbf{c}$ ). Let  $\mathbf{x}_i$  denote consensus dynamics on  $H$  ( $i \in \mathcal{V}(H)$ ), and let  $\tilde{\mathbf{x}}_a$  denote the average of the dynamics in class  $a$  (i.e.  $\tilde{\mathbf{x}}_a = \frac{1}{|a|} \sum_{i \in a} \mathbf{x}_i$ ).

Then, we have:

$$\begin{aligned} \dot{\tilde{\mathbf{x}}}_a &= \frac{1}{|a|} \sum_{i \in a} \dot{\mathbf{x}}_i \\ &= \frac{1}{|a|} \sum_{i \in a} \left( -d_i \mathbf{x}_i + \sum_{j \in N_i} \mathbf{x}_j \right) \\ &= \frac{1}{|a|} \sum_{i \in a} \sum_{j \in N_i} (\mathbf{x}_j - \mathbf{x}_i) \end{aligned}$$

Now, we suppose that the nodes states reach classes values at some time  $t$  (that is,  $\mathbf{x}_i^t = \tilde{\mathbf{x}}_a^t$  for all  $i \in a$ ). This leads to

$$\dot{\tilde{\mathbf{x}}}_a^t = \frac{1}{|a|} \sum_{i \in a} \sum_{j \in N_i^E} (\tilde{\mathbf{x}}_{c_j}^t - \tilde{\mathbf{x}}_{c_i}^t)$$

finally, grouping edges,

$$\dot{\tilde{\mathbf{x}}}_a^t = \frac{1}{|a|} \sum_{b \neq a} |\text{edges}(a, b)| (\tilde{\mathbf{x}}_b^t - \tilde{\mathbf{x}}_a^t) \quad (1)$$

where  $|\text{edges}(a, b)|$  is the number of edges between classes  $a$  and  $b$ .

<sup>6</sup>An error analysis would make this more precise, however, we leave this to further work.

<sup>7</sup>Here, where  $f(a)$  are class shifts.

<sup>8</sup> $x \sim \mathcal{N}(\mu, \sigma)$  means the random variable  $x$  follows a Normal law with mean  $\mu$  and variance  $\sigma^2$ .

But now this is just a modified version of consensus dynamics on the induced graph  $G$ , where  $\text{edges}(a, b)$  is the weight of the edge  $(a, b)$ , and  $\frac{1}{|a|}$  is an added "drag" linked to the size (or weight)  $|a|$  of the node  $a$  in  $G$ .

Again, note that the calculations remain correct in expectation if nodes states have synchronous values at some time  $t$ , i.e.  $\mathbb{E}(\mathbf{x}_i^t) = f(\mathbf{c}_i)$ . Now, if  $\mathbf{x}_i^t$  is close to  $\tilde{\mathbf{x}}_i^t$ , then  $\dot{\mathbf{x}}_i^t$  is close to  $\dot{\tilde{\mathbf{x}}}_i^t$ . Thus, we can legitimately speculate that if we initialize  $\mathbf{x}_i^0$  close to  $\tilde{\mathbf{x}}_i^0$ , then this will remain true for all  $t \geq 0$ . In this case,  $\tilde{\mathbf{x}}_i$  will reflect the general behavior of the network  $H$ . This means that if we are only interested in the class averages  $\tilde{\mathbf{x}}_i$ , we can in fact use the dynamics on the induced graph from equation 1.

## 5. Experiments

Our aim here is to illustrate the behaviors discussed above, as well as show that we in fact observe them in practice. We will be looking at numerical simulations of consensus dynamics on a large network and on the image graph associated to it.

### 5.a Numerical Setup

It is numerically heavy, in general, to find the image graph associated to a large network. Thus, for our experiences, we will construct a small graph  $G$ , and then build a large network  $H$  with image graph  $G$ .

We first generate a double weighted (weighted on both nodes and edges) graph  $G$ . For each node  $i$  in  $G$ , we create  $w_i$  nodes in  $H$  (where  $w_i$  is the weight of  $i$ ). Then, we use a SBM to generate edges on  $H$ .

The probability parameters matrix  $\Theta$  will be set in agreement with the adjacency matrix of  $G$ . We choose  $\theta_{ab} = \frac{B_{ab}}{|a||b|}$  ( $B$  being the adjacency matrix of  $G$ ) for the off-diagonal ( $a \neq b$ ) entries. We will set  $\theta_{aa}$  (the diagonal entries) to a different value depending on how dense we want our classes to be. Dense classes mean we set  $\theta_{aa}$  close to 1, while discrete classes (i.e. classes with empty set of edges) mean  $\theta_{aa} = 0$ . We will then simulate consensus dynamics on both  $G$  and  $H$ . We denote the node values on  $G$  with the vector  $\mathbf{x}^t$ , and values on  $H$  with  $\mathbf{y}^t$ ,  $t$  being time, starting at  $t = 0$ . The two consensus dynamics are defined by the following equations:

$$\dot{\mathbf{x}} = -\mathbf{d} * \mathbf{L}(G)\mathbf{x} \quad \text{with } \mathbf{L}(G) \text{ the Laplacian of } G$$

$$\dot{\mathbf{y}} = -\mathbf{L}(H)\mathbf{y} \quad \text{with } \mathbf{L}(H) \text{ the Laplacian of } H$$

where  $\mathbf{d}$  is the drag on  $G$ , i.e.  $\mathbf{d}_a = \frac{1}{|a|}$ , and  $*$  is term by term vector multiplication.

For initialization, we proceed in a more advanced way than just a random value on each node:

**Initialization on  $G$**  For each node  $a \in \mathcal{V}(G)$ , we initialize the node's state with a normal distribution:  $\mathbf{x}_a^0 \sim \mathcal{N}(0, 1)$ .

**Initialization on  $H$**  For each node  $i \in \mathcal{V}(H)$  from class  $a$  (i.e.  $i \in a$ ), we initialize the node state according to its class  $a$ :  $\mathbf{y}_i^0 \sim \mathcal{N}(\mathbf{x}_a, \sigma)$ .

This means that the states of nodes belonging to the same class will be initialized with similar values, as we assumed it in the derivations from section 4. The quality of this assumption was also discussed in section 4.

## 5.b Simulation Results

We proceed with two simulations: the first one with dense blocks, the second one with discrete blocks.

We already know that when a network has a communities structure, the consensus dynamic is affected by the communities (consensus intra-communities is reached before the general consensus). Thus, we expect the blocks states to keep a similar value.

In the case of discrete blocks, expectations are unclear: would the states follow the roles inferred, or would the edges across blocks create communities completely different from the block classes leading to local consensus?

### 5.b1 Classes densities.

We begin by testing if the intra-density affects the consensus dynamics. To keep our first model simple, we take  $G$  to be unweighted on both nodes and edges. This makes the consensus dynamics on  $G$  to be the same as the usual one (no drag factor  $d$ , nor weighted edges in the Lagrangian  $\mathbf{L}(G)$ ).

For each vertex in  $G$ , we create 30 nodes in  $H$ . We will proceed with two simulations: one with dense blocks, setting  $\theta_{aa} = 0.8$ , and one with empty blocks, setting  $\theta_{aa} = 0$ . We set the edges weight so that the edge density is  $\theta_{ab} = 0.3$  between classes  $a$  and  $b$  ( $a \neq b$ ). Finally, we take  $\sigma = 0.05$ .

The graph  $G$  considered and networks  $H$  induced are shown next (on figure 6). We plot the adjacency matrix (see figure 6) of  $G$  with the corresponding identity matrix as a parallel to dense block in the adjacency matrix of  $H$  in the dense case. This reveals clearly the "block" structure (induced from the SBM model).

The dense blocks behave very similarly to communities (in fact, the plotting tool<sup>9</sup> groups the nodes together for that reason). It is therefore not that surprising to

<sup>9</sup>We used the default one from NetworkX (a Python library for networks).

see a very similar behavior appearing from the consensus dynamic shown on figure 6.

However, the empty blocks show a very different structure (and this is why the plotting tool did not group the nodes together). Now, surprisingly, despite the discreteness of blocks classes, experiments show that the dynamics are also similar between  $G$  and  $H$ , see fig. 6.

Finally, we remark that the dynamics on  $H$  is the same (or negligibly different) whether the classes are dense or empty. Moreover, in the special case of classes all being the same size, and edges all having the same weights, the dynamics on the induced graph summarize nicely the behavior on the two networks  $H$  with dense and empty classes.

*We have plot each node from  $G$  in a different color, and accordingly nodes from the same block classes in  $H$ .*<sup>10</sup>

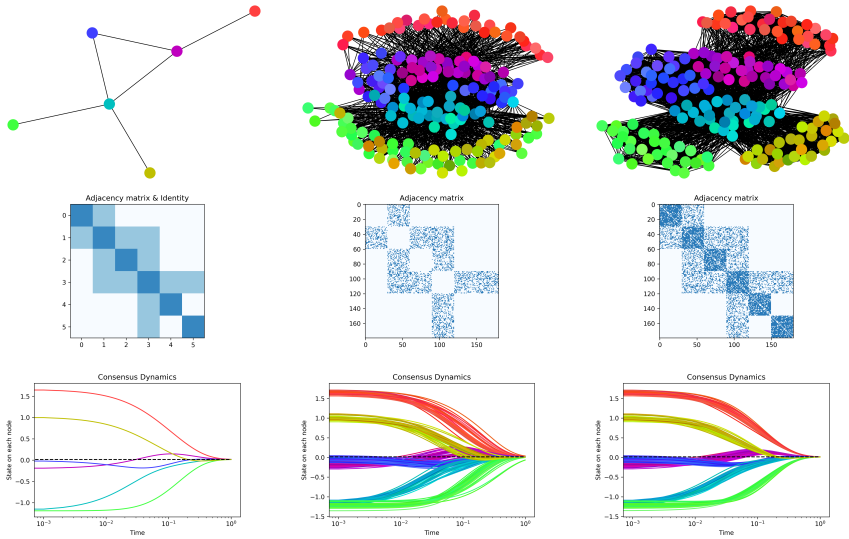


Figure 6: Left:  $G$  ; Center:  $H$  with empty classes ; Right:  $H$  with dense classes

### 5.b2 Image Graph.

We now consider a general graph  $G$  with arbitrary weights on nodes and edges. We will check that the dynamics is the same on  $G$  and on a network  $H$  with image graph  $G$ .

<sup>10</sup>This was done using a three-phase signal for the RGB color values, more details in the supplementary information.

We create a graph  $G$  with sum of nodal weights equal to 210. We then generate a network  $H$  via SBM with parameters  $\theta_{ab} = \frac{|\text{edges}(a,b)|}{|a||b|}$  for  $a \neq b$  and  $\theta_{aa} = 0.25$  (this is arbitrary, but we have already demonstrated that classes density did not impact much the dynamics). We will set  $\sigma$  to 0.05.

In the next figure (fig. 7), we plot  $G, H$ , and their associated dynamics. As expected from theory, the dynamics on  $H$  follows rigorously the one from  $G$ , with a little "blur" (some nodes slightly over, some slightly under, with a nearly perfect average).

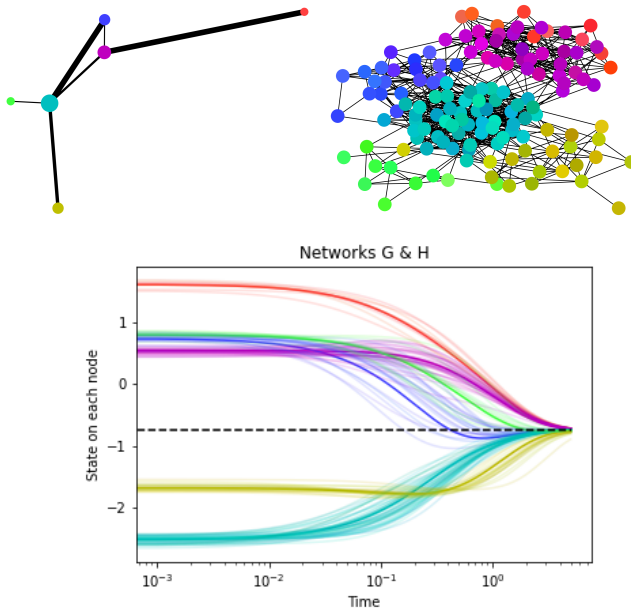


Figure 7: Left:  $G$  ; Center:  $H$  ; Right: dynamics on  $H$  (light lines) and on  $G$  (heavy lines)

Finally, we look at the effect of  $\sigma$ . From the plots (fig. 8), it seems that this phenomenon (of nodes classes keeping similar values) appears when  $\sigma \leq 0.1$  (when nodes on  $G$  are initialized following a  $\mathcal{N}(0, 1)$ ).

## 6. Conclusions

### 6.1 Model Choices

Since the dynamics are the same in  $H$  and  $G$ , we may study the small graph  $G$ , and extrapolate our analysis and conclusions to  $H$ . In our examples,  $|\mathcal{V}(H)| \sim 20 *$

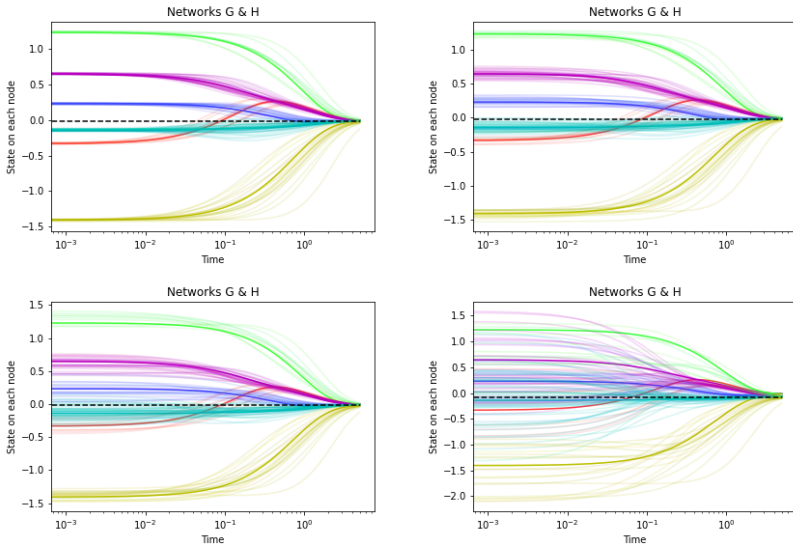


Figure 8: Top left:  $\sigma = 0.01$  ; Top right:  $\sigma = 0.05$  Bottom left:  $\sigma = 0.1$  ; Bottom right:  $\sigma = 0.5$

$|\mathcal{V}(G)|$ , so we reduced by a factor of 20 at least the size of the network to be studied. This is substantial, but the blocks could be bigger, and calculations could be cut by an even larger factor. We have also shown (experimentally) that the density inside the blocks do not impact the model's performance, surprisingly.

In a real life model, we could imagine a threshold to ignore links between two classes of nodes, if they are not numerous enough. This would avoid having some edges with extreme weights variations, and therefore simplify the model without losing too much information.

## 6.2 Comparison of communities versus roles classes

One could use communities classes instead of roles classes in the model presented above. However, this is not optimal: Communities are thought as groups of nodes with few inter-communities edges; with most of the dynamics happening within communities. On the other hand, the model using the induced graph is thought to capture how groups of nodes interact, ignoring the internal behavior. The two approaches are diametrically opposed. The best model will depend on the phenomenon we try to study.

### 6.3 Philosophical interpretation

Our role (or position) in society depends on our relationships with people with a different role (as we demonstrated the intra-class density did not impact the dynamics). 🌀

#### Acknowledgements

I would like to thank Prof. Renaud Lambiotte for his amazing teaching of Networks, despite being online. I would also like to thank The New Collection's administration for all the support they provided.

#### References

Reichardt J, White DR (2007) Role models for complex networks. *The European Physical Journal B* 60(2):217-224.

Lorrain F, White HC (1971) Structural equivalence of individuals in social networks. *The Journal of Mathematical Sociology* 1(1):49-80.

White DR, Reitz KP (1983) Graph and semigroup homomorphisms on networks of relations. *Social Networks* 5(2):193-234.

Holland PW, Laskey KB, Leinhardt S (1983) Stochastic blockmodels: First steps. *Social Networks* 5(2):109-137.

Masuda N, Lambiotte R (2016) *A Guide to Temporal Networks, Complexity Science*. (World 442 Scientific publishing).

# Brain-Machine Interfaces and Neuroplasticity: Towards Accurate and Robust Two-Learner Systems

**Peter Holderrieth**<sup>1\*</sup>, Matthew Lennon<sup>1,2</sup>, Grant Rigney<sup>3</sup>

<sup>1</sup>*Department of Physiology, Anatomy, and Genetics, University of Oxford, UK*

<sup>2</sup>*Faculty of Medicine, University of New South Wales, Australia*

<sup>3</sup>*Department of Psychiatry, Warneford Hospital, Oxford, UK*<sup>3</sup>

## 1. Introduction

Patients with severe motor deficits such as paralysis, spinal cord injury, or stroke suffer from loss of their independence to perform everyday tasks such as feeding, grooming, or controlling a computer. Brain-machine-interfaces (BMIs) (Green Kalaska, 2011) promise to improve the situation of such patients. Measuring signals from the brain and connecting these signals with a computer or a robotic arm, BMIs provide an additional pathway for patients to control and communicate with the external world. A BMI can be considered as an additional limb (e.g. robotic arm) plus nerves connecting this limb with the brain (e.g. cables or wireless communication). In addition to challenges in engineering such devices, this poses the question: is the brain able to develop the neural networks necessary to control such an artificial limb? If yes, how can we explain and describe that mechanism neuroscientifically? Are there limits? And how should that inform the design of BMIs? These are the questions this essay is concerned with.

The basic function of BMIs for movement restoration relies on the ability of patients to reproduce the electro-physiological activity in the brain associated with a movement, even in the absence of the movement itself (Buch et al., 2008; Truccolo et al., 2008). The evoked neural signals can be measured, e.g. via electrocorticography (ECoG) arrays or intracortical microelectrodes implanted (on the surface) of the corresponding cortical area. These signals are then processed and converted via

---

\*peter.holderrieth@new.ox.ac.uk



a *decoder* into control of an *effector*, i.e. a remote external device (see fig. 1). Impressive progress has been made in the design and application of such BMIs (Chapin et al., 1999; Serruya et al., 2002; Taylor et al., 2002; Musallam et al., 2010; Ethier et al., 2012; Hochberg et al., 2012; Collinger et al., 2013) allowing patients to control robotic arms (Collinger et al., 2013), tablet computers (Nuyujukian et al., 2018), or their own muscles via electrical stimulation (Moritz et al., 2008).

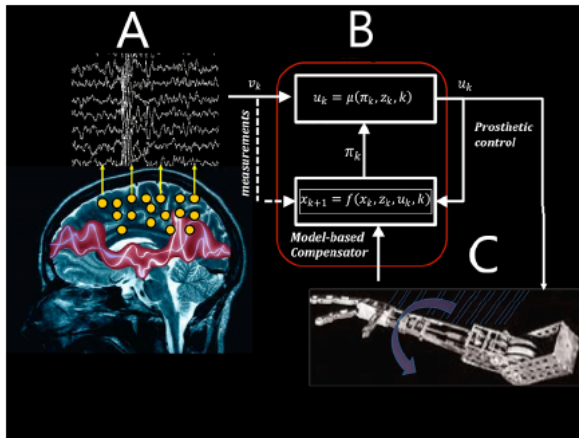


Figure 1: Figure adapted from link. Basic components of a brain-machine interface. A: neuronal activity is recorded via electrodes. B: a decoder, i.e a computer-based control program with specific decoding algorithms, transforms signals into action commands. C: action commands are converted into physical movement of an effector, here a robotic arm.

However, performance of BMIs cannot as yet replicate the subjective ease of moving a limb deliberately and accurately (Carmena, 2013). In addition to achieving high accuracy, a particular challenge is the translation from proof-of-concept studies to the everyday use of a BMI which would ideally last for years (Green Kalaska, 2011). Difficulties stem from the differences of BMIs to natural movements: to activate movement, patients have to use cortical activity - as opposed to spinal motor neuron activity. In addition, subjects control the movement by looking at the device (visual feedback) as opposed to sensing it (proprioceptive feedback).<sup>1</sup> Due to these difficulties, subjects need to *learn* how to use the non-natural BMI signalling pathway.

<sup>1</sup>BMIs detecting sensation are under development (Venkatraman Carmena, 2011; O'Doherty et al., 2011)

Naturally, understanding this learning process is crucial for the design and clinical application of BMIs.

In neuroscience, *plasticity* refers to the ability of neural networks to functionally reorganize themselves. For BMIs, an increasing body of work shows an impressive ability of the brain to adapt to neuroprosthetic devices (Ganguly Carmena, 2009; Moritz et al., 2008; Jarosiewicz et al., 2008). However, the plasticity of the brain also implies that the brain is constantly reorganizing itself. How can we ensure stable performance if the brain is always changing? A popular approach is to use *closed-loop decoder adaptation* (CLDA). With CLDA, the decoder, i.e. the part that converts electrical signals from the brain to movement initiation, adapts to changes in neuronal input to the BMI (Dangi et al., 2013). While CLDA has shown remarkable success (Li et al., 2011; Gilja et al., 2012; Orsborn et al., 2012; Jarosiewicz et al., 2013), this creates a system where two adaptation processes happen simultaneously: in the brain and in the BMI. Therefore, the ultimate role of neuroplasticity and CLDA in the applications of BMIs remains unclear.

In this essay, we will review recent work on BMIs related to neuroplasticity and CLDA and discuss how these inform the design of BMI systems. We will argue for BMIs that are true *two-learner* systems: a strategy that gives both decoders and cortical networks the space to adapt to each other maximizes long-term performance, robustness and transferability of tasks. We restrict ourselves to BMIs for restoring lost movement which use invasive recording techniques such as ECoG or intracortical microelectrodes and to BMIs which decode cortical, as opposed to subcortical, signals.

## 2. Neuroplasticity and the Acquisition of Neuroprosthetic Skill

The ability to perform tasks with a BMI appears to be a learned skill, similar to learning an instrument or a sport. Task performance usually increases over practice (Carmena et al., 2003; Wolpaw McFarland, 2004; Jarosiewicz et al., 2013; Hochberg et al., 2006), and human BMI users report a transition from conscious control to nearly automatic execution. This suggests a learning process is occurring in the brain whose neural correlates we discuss in this section.

### 2.1 Cortical maps form and can be recalled

It is known that specific properties of neurons change during learning a BMI. *Neuronal tuning* refers to the property of a neuron to encode specific sensory, motor or cognitive information with the rate or timing of action potentials. For example, certain motor neurons are tuned to a specific direction of movement, i.e. their firing

rate increases when movement is executed in the preferred direction. During learning a BMI, the directional tuning of neurons change allowing to control a BMI in space (Taylor et al., 2002; Carmena et al., 2003; Truccolo et al., 2008).

Studying not only single neurons but whole neural networks, Ganguly Carmena (2009) were among the first to demonstrate a long-term stable reorganization of motor cortex in response to the use of BMIs. They performed recordings across 19 days from primary motor cortex in nonhuman primates using a stable ensemble of neurons with a constant decoder, i.e. a decoder that is not changed during the whole training and test process. By this, they could track changes in neuronal properties over time. As shown in fig. 2, their findings suggest that motor cortex is able to maintain a stable neural representation. Observing that performance within a daily session stays relatively constant, they also find that acquired neuroprosthetic skills could be readily recalled from previous days. In addition, they observed that small disturbances to the decoder and the size of the neural ensemble could impair neuroprosthetic skill showing the specificity and robustness of this cortical map.

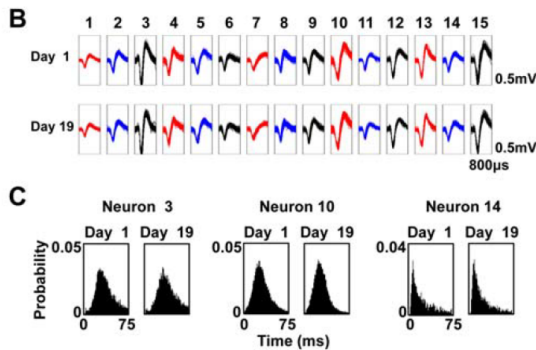


Figure 2: Figure adapted from Ganguly Carmena (2009). B: waveforms for single neurons on day 1 and 19 (height is 0.5mV and width 800 ms resp.). C: Interspike interval (ISI) distributions for days 1 and 19 for three representative units. One can observe stability of waveforms and ISI distributions across the whole experiment.

## 2.2 Neuroprosthetic skill acquisition depends on distributed networks

Next, we discuss the extent to which neuronal networks beyond those directly measured by the BMI are involved in the acquisition of neuroprosthetic skill. Wander et al. (2013) implanted ECoG electrodes recording high-gamma band signals in 7 subjects who were required to control the position of a cursor on a computer monitor. They studied changes in activation of prefrontal cortex, premotor cortex, and

posterior parietal cortex, i.e. areas that have been previously been associated with motor sequence learning and task learning (Van Mier et al., 1998). As one can see in fig. 3, they find that initially observed task-specific activation patterns diminish over the course of learning the BMI task. Therefore, this work strongly suggests that learning to control a BMI resembles natural motor skill learning and relies not only on cortical areas in M1 directly involved in the signalling path of a BMI but also on distributed neuronal networks across the central nervous system.

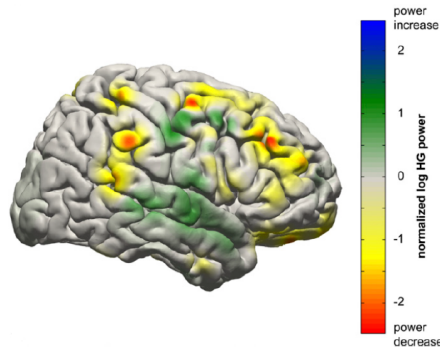


Figure 3: Figure adapted from Wander et al. (2013). Frontal areas and posterior parietal areas show a decrease in task-related activation over the course of BCI use.

### 2.3 Limitations of neuroplasticity

BMIs are non-physiological, or "non-natural", pathways which patients have to use to signal their intentions. It is therefore reasonable to ask to what extent learning to control a BMI can rely on natural motor circuits or whether it has to form "novel" BMI-specific neural networks. In particular, some earlier works suggest that any subset of neurons can be activated by a BMI, and the directional tuning of neurons can change arbitrarily (Moritz et al., 2008; Jarosiewicz et al., 2008). These learning mechanism is called *invidual-neuron* learning mechanisms by Hwang et al. (2013) since each neurons adapts independently to maximize the "reward" based on the success of BMI control.

However, Hwang et al. (2013) argue that the learning process more resembles something which they call *intrinsic-variable* process, i.e. global network structures of neurons make them dependent on each other, and hence there are limits to neuroplasticity. They study the behaviour of neurons whose activity is measured by BMIs but not used by the decoder ("untrained neurons"). They consider neurons in the

posterior parietal cortex of macaque monkeys whose spike rate increases when the subject reaches right as opposed to left. For the decoder, they turn this rule around, i.e. an increase in spike rates causes the BMI to reach the left. Interestingly, they find that during the learning process not only the trained but also the untrained neurons switch their preferred stimulus locations. Under individual-neuron learning, the untrained neurons would not behave in such a consistent way due to their independence. Hence, their work strongly suggests that BMIs leverage the repertoire of neural activation patterns associated with natural movement. This repertoire is limited and should be leveraged in the design of BMIs.

### 3. Closed-Loop Decoder Adaptation

Having seen how the brain adapts to BMIs, we discuss the opposite direction in this section: adaptation of the decoder. Before subjects start to learn how to control a BMI, the decoder is usually initialized while subjects either (1) perform real movements (Serruya et al., 2002; Taylor et al., 2002; Santhanam et al., 2006) or (2) imagine given movements (Velliste et al., 2008; Hochberg et al., 2006). In this initial starting phase, the decoder is fitted to predict the (1) real or (2) imagined movements from recorded neural activity. From this point on, subjects can start learning to control the BMI.

#### 3.1 The need for online decoder adaptation and feedback

Early works using computational decoding methods have kept the parameters of the decoder constant after initial decoder fitting, i.e. all decoder adaptations happen offline in this case (Wessberg et al., 2000; Hochberg et al., 2006; Wu et al., 2006; Ganguly Carmena, 2009). In particular, during the learning process these methods require subjects to perform all adaptations to the device. However, there are several drawbacks to this method.

Firstly, the ability of the brain to adapt might be limited as discussed in section 2.3. Several works have also shown that the neuronal tuning of arm movement can be unstable over time and change quickly - even within time spans of 60min (Li et al., 2001; Padoa-Schioppa et al., 2004; Lebedev et al., 2005; Kim et al., 2003; Rokni et al., 2007; Chestek et al., 2007). Similar observations about changes in neuronal tuning have been made for BMI control (Taylor et al., 2002; Truccolo et al., 2008). While we have seen in section 2.1 that for *constant* recordings conditions stable cortical maps can evolve, practical considerations such as recording instability may alter the recorded spike waveforms over time. Together, this can lead to a significant decline of BMI performance over time if the decoder is kept constant (e.g. see non-adaptation experiments of Li et al. (2011)).

Secondly, there are systematic differences between neural activity patterns for the offline compared to the online phase. Taylor et al. (2002) already demonstrated that having visual as opposed to no feedback in a BMI system significantly changes neural activity patterns which implies that one might fit the decoder at the start to different neural dynamics. This was also later confirmed by research showing that the prediction performance of decoders purely based on recorded signals in the offline-phase (open-loop) does not necessarily correlate with performance when a subject controls a BMI and receives feedback (closed-loop) (Ganguly Carmena, 2010; Koyama et al., 2010).

Therefore, it became common practice to use *closed-loop decoder adaptation* (CLDA) where one updates the decoder "online", i.e. not only at initial fitting but also during the learning process (Jarosiewicz et al., 2013).

### 3.2 Closed-loop decoder adaptation

We briefly discuss the most popular CLDA approach which builds on the Kalman Filter (KF). In this framework, the (intended) state of a BMI is given by its position, represented by 3d coordinates  $\mathbf{p} = (p_1; p_2; p_3)$ , and its movement velocity, represented by a 3d velocity vector  $\mathbf{v} = (v_1; v_2; v_3)$ . For every measured time point  $t = 0, 1, 2, 3, \dots$ , e.g. every 10 milli seconds, the state of the system is described by the combination of its position and its velocity  $\mathbf{x}_t = (\mathbf{p}_t, \mathbf{v}_t)$  (kinematics). The neural signals by the BMI are given by a vector  $\mathbf{y}_t$

How can we model the connection between  $\mathbf{x}_t$  and  $\mathbf{y}_t$ ? The KF models it as a so-called *linear dynamical system with Gaussian noise*. More specifically, the dynamics are given by

$$\mathbf{x}_t = A\mathbf{x}_{t-1} + \boldsymbol{\omega}_t, \quad \mathbf{y}_t = C\mathbf{x}_t + q_t, \quad \boldsymbol{\omega}_t \sim \mathcal{N}(0, W), \quad q_t \sim \mathcal{N}(0, Q) \quad (2)$$

In words, this means that the kinematics  $\mathbf{x}_{t-1}$  at the previous time point are transformed to the kinematics  $\mathbf{x}_t$  of the current point by the matrix  $A$  plus some random noise  $\boldsymbol{\omega}_t$  with Gaussian distribution that varies from time to time. The brain signals  $\mathbf{y}_t$  are a function of the kinematics  $\mathbf{x}_t$  given by the matrix  $C$  plus some random noise  $q_t$  with Gaussian distribution. Crucially, we only observe the brain activity  $\mathbf{y}_t$  and do not know the intended kinematics  $\mathbf{x}_t$  of the patient. The KF (Haykin, 2008) allows inference of  $\mathbf{y}_t$  from  $\mathbf{x}_t$  in such a model for given  $A, C, W, Q$ . Therefore, to begin, one has to find optimal parameters for the matrices  $C, A, Q, W$ , i.e. one has to fit the decoder.

In CLDA, this is not only done in the initial training phase but throughout usage of the BMI. Various versions of CLDA algorithms differ in terms of how parameters are updated, when this happens, and how updates are done. For example, Batch

Maximum Likelihood Estimation (BMLE) collects a batch of data for a certain period of time and then updates the parameters using the whole batch (Gilja et al., 2012), while the Adaptive Kalman Filter updates the parameters at every time step  $t$  (every  $\sim 100\text{ms}$ ) Dangi et al. (2011). For a systematic empirical and theoretical study of CLDA algorithms, see Dangi et al. (2013).

### 3.3 CLDA enables long-term BMI use

Using CLDA with BMLE, Gilja et al. (2012) showed that one can achieve high performance of BMIs for a long timespan - even over years. They present a new method which they call recalibrated feedback intention-trained Kalman filter (ReFIT-KF). During training, this method infers the true intentions  $\mathbf{x}_t$  for model fitting purposes using two simple principles: first, subjects intend to move the cursor straight to the target and velocities are rotated toward the target to generate estimated velocities. Second, if the cursor is on the target, one assumes zero velocity. Moreover, they propose that the state transition model parameters ( $A$  and  $W$ ) should be constrained to be consistent with physical kinematics, i.e. integrating the velocity leads to position. The authors tested their methods in rhesus monkeys which had been implanted with ECoG arrays and observed that ReFIT-KF significantly outperforms previous algorithms (see fig. 4a). They observe prolonged and continuous use over hours and generalizability to more complex tasks without retraining. As shown in fig. 4b, their method shows repeatable high performance for years. However, one can also observe high variability of performance across days.

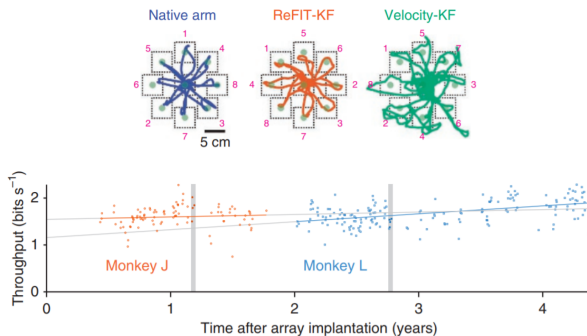


Figure 4: Figure adapted from Gilja et al. (2012). Top: qualitative comparison of ReFIT-KF with Velocity-KF. Bottom: linear regression of performance (Fitts's law metric) of two monkeys across years.

## 4. Two-learner systems

While we have seen that a stable cortical map can form under well-isolated, stable neural activity, maintaining such a BMI recording configuration over multiple years is infeasible in a clinical setting. If we use CLDA to alleviate this, it is not clear a priori what we can say about cortical map formation in such an unstable setting. In fact, it has been observed that regular re-training of decoders can lead to an absence of cortical map formation (Ganguly Carmena, 2009). In studies with daily decoder adaptation over weeks, unstable performance was observed as well (Taylor et al., 2002; Collinger et al., 2013; Gilja et al., 2012). In this section, we discuss the question of how one can create stable performance across a prolonged period of time and how neuroplasticity and CLDA interact.

#### 4.1 Maintaining accuracy

First of all, we discuss to which extent cortical maps and already successfully fitted decoders can be stabilized over time via CLDA. Li et al. (2011) perform periodic (every 2min) updates of the neuronal tuning model via two approaches of Bayesian linear regression: first, a fast and exact method ("joint formulation"), second, an approximate method using Variational Inference methods allowing the omission of neurons from updates ("factorized formulation"). They performed experiments on rhesus monkeys with implanted microwire electrodes and show that CLDA stabilizes high performance over a prolonged period of time if learning starts with already high-performing decoders.

#### 4.2 Random initialization and brain-decoder co-adaptation

In clinical situations, many patients cannot perform the natural movements often used to initially train the decoder due to severe motor deficits. Maintaining high performance is therefore not sufficient since the initial fitting process is either not possible or leads to unsatisfying performance. It is therefore crucial to know the extent to which cortical maps can emerge for "unfavourable" decoder initialization. Orsborn et al. (2012) propose the SmoothBatch algorithm which updates parameters via sliding averages and test it by training non-human primates to perform a center-out reaching task. While they find some correlation between initial performance and final accuracy, they demonstrate that their method can achieve high performance even for random decoder initializations. This was confirmed both theoretically via mathematical convergence analysis and by experimental results with non-human primates by Dangi et al. (2013). In addition, Orsborn et al. (2012) find decreasing "sizes" (Frobenius norm) of updates indicating that decoder parameters converge (see fig. 5A). This goes along with a decrease of neuroplasticity: directional tunings of BMI units converge and differentiate with decreasing speed into complex patterns (see



fig. 5B and C) and even after decoder adaptation was stopped, subjects were able to maintain accurate reaches for over 30min. Together, this work indicates that there is a co-adaptation or co-convergence process between the brain and the decoder which can succeed even for random initialization by using the right CLDA methods.

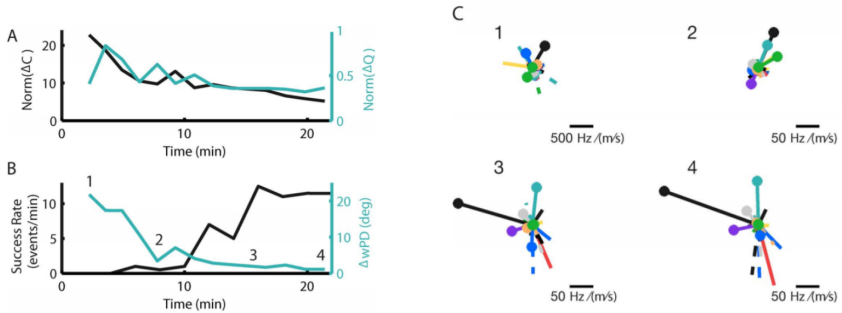


Figure 5: Figure adapted from Orsborn et al. (2012). A: Frobenius norm of changes in matrices  $C$  and  $Q$  (see eq. (1)) converges to stable changes. B: Average difference in velocity preferred directions (PD) weighted by relative tuning strength. Convergence is consistent with behavioural improvement. C: Differentiation of velocity tuning over the 4 time points seen in B (vector length equal to MD and orientation equals PD).

### 4.3 Formation of cortical maps by combining CLDA and neuroplasticity

In every day use, subjects would need to execute a big variety of functions with a BMI and they would be required to do so in combination with other cognitive tasks. To study the interaction between BMIs and neuroplasticity in such contexts, Orsborn et al. (2014) require non-human primates to control a two-dimensional cursor under two scenarios: with changing population of neurons contributing to the decoder and with non-stationary recorded neural activity. They present a novel training paradigm to leverage both neuroplasticity and CLDA building on the idea that one should only apply CLDA at targeted times: on the first day, CLDA is used to improve initial performance. Subsequently, the decoder is fixed. If performance drops or the recorded neural activity shifts (e.g. changes in the contributing units to the decoder), CLDA is activated again to adjust the decoder. Moreover, in contrast to Orsborn et al. (2012), CLDA is stopped, if activated, *before* converging to maximal performance.

Using this method, Orsborn et al. (2012) find in their experiments that performance improved clearly over days with a refinement of cursor movements and a re-

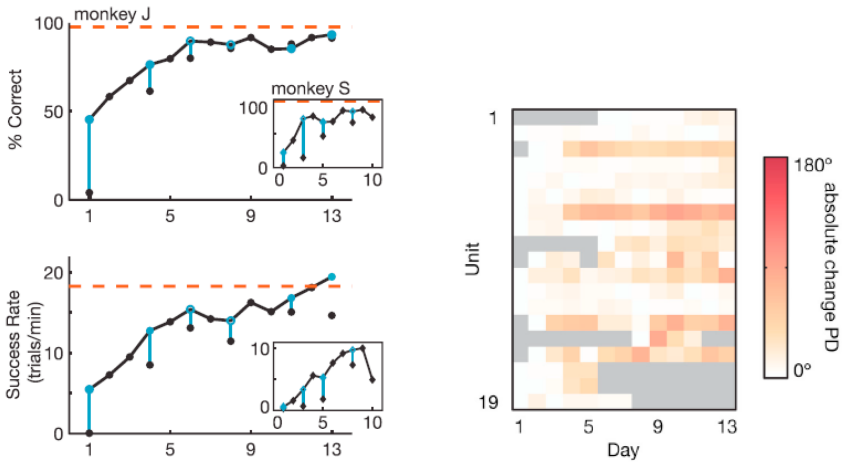


Figure 6: Figure adapted from Orsborn et al. (2014). Left: increase in performance accuracy and success rate over days. Blue indicates application of CLDA and red dashed lines show typical manual performance. Right: changes in PD relative to first day for all BMI units (grey squares were out of the ensemble).

duction in the average movement error. If performance dropped from one day to the next, the application of CLDA raised performance again to a higher level than the day before leading to a continuous increase in performance across days (see fig. 6a). This shows that using CLDA at targeted times with limited amount leads to better long-term results. Since these gradual improvements could not be observed when CLDA was applied until maximized performance given varying neural ensembles, one can also conclude that these results do not just reflect increased practice with the BMI but rather consistent improvement across days. In addition, Orsborn et al. (2012) did similar experiments to Ganguly Carmena (2009) and showed that decoder-specific cortical maps form over the course of learning. By comparing tuning curves across days, they also find that cortical maps of subsequent days became more strongly correlated at the end of the learning process, which indicates a stabilization process of cortical maps is occurring (see fig. 6b).

## 5. Conclusion

From this essay, we infer three main research directions and approaches towards the design of clinically viable BMIs.

First, a better understanding of neuroplasticity informs the design of BMIs. Previous works studying neuroplasticity in response to BMI control demonstrate the

emergence of specialized BMI control networks, that plasticity induced by BMIs happens across the brain, and that BMIs have to leverage the intrinsic repertoire of neural structures associated with natural movement. Hence, a better understanding of neuroplasticity, in particular computational models, and motor neuron circuits will be crucial for the design of better decoder models.

Secondly, CLDA addresses limitations of neuroplasticity in practice such as non-stationarity of neuronal activity and difference in ensembles of recorded in neurons across time. It can maintain accuracy of high-performing decoders and achieve co-convergence of neuroplasticity and decoder parameters even from random decoder initialization. However, while some studies comparing different CLDA methods exist (Dangi et al., 2013), a systematic understanding how to choose a particular algorithm and design for CLDA for particular settings remains a subject of future investigations.

Thirdly, the sole application of CLDA does not lead to optimal performance. Training procedures which allow subjects to adapt to the decoder and to transfer neuroprosthetic skills across days lead to consistent improvements and robust results and allow the formation of cortical maps even in the presence of non-stationary neuronal ensembles. Therefore, the path towards effective and clinically viable BMIs lies in two-learner systems where both the decoder and the brain dynamically adapt to each other in a natural manner.

BMIs promise to help millions of patients to lead a more independent life. For example, they might enable patients with paraplegia to access digital information on the web or to enable children who lost their limbs by meningitis to play board games with their friends. As this essay illustrates, there remain significant challenges to make BMIs clinically applicable but the astonishing progress made in the last decades allows us to look optimistic into the future. 🌟

## References

- Buch, E., Weber, C., Cohen, L. G., Braun, C., Dimyan, M. A., Ard, T., Mellinger, J., Caria, A., Soekadar, S., Fourkas, A., et al. Think to move: a neuromagnetic brain-computer interface (bci) system for chronic stroke. *Stroke*, 39(3):910–917, 2008.
- Carmena, J. M. Advances in neuroprosthetic learning and control. *PLoS Biol*, 11(5):e1001561, 2013.
- Carmena, J. M., Lebedev, M. A., Crist, R. E., O’Doherty, J. E., Santucci, D. M., Dimitrov, D. F., Patil, P. G., Henriquez, C. S., and Nicolelis, M. A. L. Learning to control a brain-machine interface for reaching and grasping by primates. *PLOS Biology*, 1(2), 10 2003. doi: 10.1371/ journal.pbio.0000042.
- Chapin, J. K., Moxon, K. A., Markowitz, R. S., and Nicolelis, M. A. Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nature neuroscience*, 2(7):664–670, 1999.
- Chestek, C. A., Batista, A. P., Santhanam, G., Byron, M. Y., Afshar, A., Cunningham, J. P., Gilja, V., Ryu, S. I., Churchland, M. M., and Shenoy, K. V. Single-neuron stability during repeated reaching in macaque premotor cortex. *Journal of Neuroscience*, 27(40):10742–10750, 2007.
- Collinger, J. L., Wodlinger, B., Downey, J. E., Wang, W., Tyler-Kabara, E. C., Weber, D. J., McMorland, A. J., Velliste, M., Boninger, M. L., and Schwartz, A. B. High-performance neuroprosthetic control by an individual with tetraplegia. *The Lancet*, 381(9866):557–564, 2013.
- Dangi, S., Gowda, S., Héliot, R., and Carmena, J. M. Adaptive kalman filtering for closedloop brain-machine interface systems. In *2011 5th International IEEE/EMBS Conference on Neural Engineering*, pp. 609–612. IEEE, 2011.
- Dangi, S., Orsborn, A. L., Moorman, H. G., and Carmena, J. M. Design and analysis of closedloop decoder adaptation algorithms for brain-machine interfaces. *Neural computation*, 25(7): 1693–1731, 2013.
- Ethier, C., Oby, E. R., Bauman, M. J., and Miller, L. E. Restoration of grasp following paralysis through brain-controlled stimulation of muscles. *Nature*, 485(7398):368–371,

2012.

Ganguly, K. and Carmena, J. M. Emergence of a stable cortical map for neuroprosthetic control. *PLOS Biology*, 7(7):1–13, 07 2009. doi: 10.1371/journal.pbio.1000153.

Ganguly, K. and Carmena, J. M. Neural correlates of skill acquisition with a cortical brain– machine interface. *Journal of Motor Behavior*, 42(6):355–360, 2010.

Gilja, V., Nuyujukian, P., Chestek, C. A., Cunningham, J. P., Byron, M. Y., Fan, J. M., Churchland, M. M., Kaufman, M. T., Kao, J. C., Ryu, S. I., et al. A high-performance neural prosthesis enabled by control algorithm design. *Nature neuroscience*, 15(12): 1752–1757, 2012.

Green, A. M. and Kalaska, J. F. Learning to move machines with the mind. *Trends in neurosciences*, 34(2):61–75, 2011.

Haykin, S. S. *Adaptive filter theory*. Pearson Education India, 2008.

Hochberg, L. R., Serruya, M. D., Friehs, G. M., Mukand, J. A., Saleh, M., Caplan, A. H., Branner, A., Chen, D., Penn, R. D., and Donoghue, J. P. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, 442(7099):164–171, 2006.

Hochberg, L. R., Bacher, D., Jarosiewicz, B., Masse, N. Y., Simeral, J. D., Vogel, J., Haddadin, S., Liu, J., Cash, S. S., Van Der Smagt, P., et al. Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature*, 485(7398):372–375, 2012.

Hwang, E., Bailey, P., and Andersen, R. Volitional control of neural activity relies on the natural motor repertoire. *Current Biology*, 23(5):353–361, 2013. ISSN 0960-9822. doi: [https:// doi.org/10.1016/j.cub.2013.01.027](https://doi.org/10.1016/j.cub.2013.01.027).

Jarosiewicz, B., Chase, S. M., Fraser, G. W., Velliste, M., Kass, R. E., and Schwartz, A. B. Functional network reorganization during learning in a brain–computer interface paradigm. *Proceedings of the National Academy of Sciences*, 105(49):19486–19491, 2008.

Jarosiewicz, B., Masse, N. Y., Bacher, D., Cash, S. S., Eskandar, E., Friehs, G., Donoghue, J. P., and Hochberg, L. R. Advantages of closed-loop calibration in intracortical brain–computer interfaces for people with tetraplegia. *Journal of neural engineering*, 10(4):046

012, 2013.

Kim, S.-P., Sanchez, J. C., Erdogmus, D., Rao, Y. N., Principe, J. C., and Nicolelis, M. Modeling the relation from motor cortical neuronal firing to hand movements using competitive linear filters and a mlp. In *Proceedings of the International Joint Conference on Neural Networks, 2003.*, volume 1, pp. 66–70. IEEE, 2003.

Koyama, S., Chase, S. M., Whitford, A. S., Velliste, M., Schwartz, A. B., and Kass, R. E. Comparison of brain–computer interface decoding algorithms in open-loop and closed-loop control. *Journal of computational neuroscience*, 29(1-2):73–87, 2010.

Lebedev, M. A., Carmena, J. M., O’Doherty, J. E., Zacksenhouse, M., Henriquez, C. S., Principe, J. C., and Nicolelis, M. A. Cortical ensemble adaptation to represent velocity of an artificial actuator controlled by a brain-machine interface. *Journal of Neuroscience*, 25(19):4681–4693, 2005.

Li, C.-S. R., Padoa-Schioppa, C., and Bizzi, E. Neuronal correlates of motor performance and motor learning in the primary motor cortex of monkeys adapting to an external force field. *Neuron*, 30(2):593–607, 2001.

Li, Z., O’Doherty, J. E., Lebedev, M. A., and Nicolelis, M. A. Adaptive decoding for brain-machine interfaces through bayesian parameter updates. *Neural computation*, 23(12):3162–3204, 2011.

Moritz, C. T., Perlmutter, S. I., and Fetz, E. E. Direct control of paralysed muscles by cortical neurons. *Nature*, 456(7222):639–642, 2008.

Musallam, S., Andersen, R. A., Corneil, B. D., Greger, B., and Scherberger, H. Cognitive control signals for neural prosthetics, November 2 2010. US Patent 7,826,894.

Nuyujukian, P., Albites Sanabria, J., Saab, J., Pandarinath, C., Jarosiewicz, B., Blabe, C. H., Franco, B., Mernoff, S. T., Eskandar, E. N., Simeral, J. D., et al. Cortical control of a tablet computer by people with paralysis. *PLoS one*, 13(11):e0204566, 2018.

O’Doherty, J. E., Lebedev, M. A., Ifft, P. J., Zhuang, K. Z., Shokur, S., Bleuler, H., and Nicolelis, M. A. Active tactile exploration using a brain–machine–brain interface. *Nature*, 479(7372): 228–231, 2011.

Orsborn, A. L., Dangi, S., Moorman, H. G., and Carmena, J. M. Closed-loop decoder adaptation on intermediate time-scales facilitates rapid bmi performance improvements independent of decoder initialization conditions. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 20(4):468–477, 2012.

Orsborn, A. L., Moorman, H. G., Overduin, S. A., Shanechi, M. M., Dimitrov, D. F., and Carmena, J. M. Closed-loop decoder adaptation shapes neural plasticity for skillful neuroprosthetic control. *Neuron*, 82(6):1380–1393, 2014.

Padoa-Schioppa, C., Li, C.-S. R., and Bizzi, E. Neuronal activity in the supplementary motor area of monkeys adapting to a new dynamic environment. *Journal of neurophysiology*, 91(1): 449–473, 2004.

Rokni, U., Richardson, A. G., Bizzi, E., and Seung, H. S. Motor learning with unstable neural representations. *Neuron*, 54(4):653–666, 2007.

Santhanam, G., Ryu, S. I., Byron, M. Y., Afshar, A., and Shenoy, K. V. A high-performance brain–computer interface. *Nature*, 442(7099):195–198, 2006.

Serruya, M. D., Hatsopoulos, N. G., Paninski, L., Fellows, M. R., and Donoghue, J. P. Instant neural control of a movement signal. *Nature*, 416(6877):141–142, 2002.

Taylor, D. M., Tillery, S. I. H., and Schwartz, A. B. Direct cortical control of 3d neuroprosthetic devices. *Science*, 296(5574):1829–1832, 2002.

Truccolo, W., Friehs, G. M., Donoghue, J. P., and Hochberg, L. R. Primary motor cortex tuning to intended movement kinematics in humans with tetraplegia. *Journal of Neuroscience*, 28 (5):1163–1178, 2008.

Van Mier, H., Tempel, L., Perlmutter, J., Raichle, M., and Petersen, S. Changes in brain activity during motor learning measured with pet: effects of hand of performance and practice. *Journal of Neurophysiology*, 80(4):2177–2199, 1998.

Velliste, M., Perel, S., Spalding, M. C., Whitford, A. S., and Schwartz, A. B. Cortical control of a prosthetic arm for self-feeding. *Nature*, 453(7198):1098–1101, 2008.

Venkatraman, S. and Carmena, J. M. Active sensing of target location encoded by cortical microstimulation. *IEEE Transactions on Neural Systems and Rehabilitation En-*

*gineering*, 19 (3):317–324, 2011.

Wander, J. D., Blakely, T., Miller, K. J., Weaver, K. E., Johnson, L. A., Olson, J. D., Fetz, E. E., Rao, R. P., and Ojemann, J. G. Distributed cortical adaptation during learning of a brain–computer interface task. *Proceedings of the National Academy of Sciences*, 110(26): 10818–10823, 2013.

Wessberg, J., Stambaugh, C. R., Kralik, J. D., Beck, P. D., Laubach, M., Chapin, J. K., Kim, J., Biggs, S. J., Srinivasan, M. A., and Nicolelis, M. A. Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature*, 408(6810):361–365, 2000.

Wolpaw, J. R. and McFarland, D. J. Control of a two-dimensional movement signal by a noninvasive brain–computer interface in humans. *Proceedings of the national academy of sciences*, 101(51):17849–17854, 2004.

Wu, W., Gao, Y., Bienenstock, E., Donoghue, J. P., and Black, M. J. Bayesian population decoding of motor cortical activity using a kalman filter. *Neural computation*, 18(1):80–118, 2006.





## Part II: *New Perspectives*



## *COVID, what might the future hold? A conversation with Alain Townsend FRS, Professor of Molecular Immunology.*

### **JOE: What makes SARS CoV 2 especially problematic?**

ALAIN: It depends how you define problematic. COVID came a little bit from left field. It wasn't completely unexpected in the sense that SARS 1, a very close cousin, caused epidemics rather than a pandemic so we knew there was potential there. The problem was, of course, that it's much more infectious than SARS 1. It has a lethality not as high as SARS 1 but very significant. And we had absolutely nothing to protect ourselves with except social distancing and lockdown: simple measures to prevent infection. It's much less problematic than HIV from an immunological point of view, as it is a straightforward coronavirus and antibodies to the spike protein can be neutralising and highly protective. So standard vaccination can work. Now, there are some problems. It's an RNA virus: it mutates and can adapt very rapidly just like flu. This is a problem, but it's not insoluble.

### **JOE: Following on from that, what do you think the risk is that the virus will adapt to our defensive measures?**

ALAIN: It will continue to adapt and it will adapt, probably, to whatever we throw at it. It can certainly adapt to vaccination by mutating. It can almost certainly adapt to any drugs that will become available just as flu has done. But, I have great faith in the ability of science to solve these problems, one at a time if necessary. Vaccines are coming that can deal with many strains at once. Drugs given in combination where the virus has to develop multiple mutations at once to survive will become available. These problems are not insoluble, and I have great faith in the ingenuity of humankind to come to grips and to wrestle this thing to the ground. That will happen, but what we mustn't do is become overconfident because we haven't completed the job yet.

### **JOE: When you "say wrestle it to the ground", what does that mean?**

ALAIN: That means lower the mortality rate below the background rate and make sure that we don't have a heavy burden of non-mortal illness: so-called long COVID. We don't want a situation, for instance, where we have control of the virus,

in the sense that vaccination has prevented mortal illness, but people still get infected and get long COVID. So far what we're seeing, both in the animal models and in humans, is that in certain circumstances these vaccines do not prevent infection. The level of pathology might be less so that death doesn't occur, but other things might. There hasn't been enough time to watch it in long term studies.

**JOE: Do you think that social distancing measures might be necessary going forward?**

ALAIN: The difficulty is the conflicting risks and benefits of the different procedures. One thing for absolute certain is that strict social distancing and lockdown prevents infections and prevents deaths. We've seen how successful it can be when it's really strictly applied. Half measures don't really work; the tiering system, in my opinion, didn't really work. Now, in principle, if you locked down right at the very beginning as things take off, you can suppress COVID completely. Test and trace can then, in principle, trace every case. In some countries that has actually been done, but it requires an infringement of personal freedoms to levels which are almost intolerable in Western democracies. Part of me wishes we remained locked down last summer, developed test and trace to a point where it was working perfectly, and then avoided the terrible winter that followed. Looking forward I hope we do what our leaders have been saying, which is to unlock very, very cautiously.

**JOE: Do you think we are seeing the start of seasonality, and what would make a virus seasonal?**

ALAIN: One of the great temptations and dangers of being an experimental scientist is making predictions; you immediately enter into extremely dodgy territory. SARS 1 didn't become seasonal, but we do have seasonal coronaviruses. About 1/3 of common colds are caused by relatives of SARS CoV 2 and they are sort of seasonal. So, COVID might become seasonal or, once the availability of infective people drops below a certain level, it might just disappear. This family of viruses have a nasty habit of persistence, so it may well stick around. But if it does, we will vaccinate whenever we need to vaccinate, and the efficiency of the vaccines will improve. It's a bit like, unfortunately, war - in that research drives forward at an incredibly fast pace - and some of the research that has been done in the last year is absolutely phenomenal. I believe out of this we will develop new generations of vaccines that would be much better for flu also.

**JOE: I heard long ago, probably in a virology lecture, that there's selection pressure for viruses to become less deadly over time as deadly viruses have fewer opportunities to transmit. Do you think this is something that will play out for SARS CoV 2?**

ALAIN: Well, it's a very general theory. The pattern that I think you see very often, is that a virus which has achieved equilibrium in one species, so in this case

probably bats, that has crossed into humans where it's got an open field needs to find a new equilibrium. In some model systems, if you adapt a virus to a new species, it can become more virulent not less virulent. Usually if it's gone through several cycles of epidemics in a population, and the remaining people have had experience of that virus, there's a strong underlying basic immune response. And so, you can end up with a situation where most people get a mild illness from the new variant and there's this uneasy equilibrium like we have with flu. I don't think you can make too many general rules about how viruses end up in terms of their relationship with us. I have no idea what this thing is going to do. It might get nastier, or it might just fade: there's many scenarios all of which I think are possible.

**JOE: If you were to put money on any of these outcomes, what would it be on the basis of experience with pandemics?**

ALAIN: Whenever I hear a scientist making predictions, I worry. You can model the past and then say: "if the future behaves like the past this is what you might see". But in practice, the future never behaves like the past, something always changes. One thing which has been particularly pleasing, is to see that very quickly we've developed three types of vaccine, all of which are highly effective, and I think that will continue. So, if I were to make any predictions, I would say vaccination is going to be able to control it, at least based on what we've experienced so far. The results have been much better than many of us expected.

**JOE: 1 in 10 people have symptoms over 12 weeks. Is that something that's particularly unique to SARS CoV 2? Or maybe are we only measuring it because we've had such an enormous surge of virus in the population?**

ALAIN: That's a very good question. I do hear people questioning how many people who have flu have symptoms longer than a month. I don't know. But long COVID really does seem to be different to that. There are very specific syndromes that seem to be appearing. Without doing a really sophisticated prospective long-term cohort study, I don't think we're going to know. But that is happening, I think the NIHR have spent 18.5 million financing a big cohort study which I think would be very, very worthwhile, and that will give information revealing both what it is, in terms of pathology, and how best to manage it. There are lots of unknowns, but it definitely exists, and it's unpleasant. What worries me is that it doesn't seem to be related to the severity of the actual first illness. So, people with quite mild illness can end up with quite persisting long term symptoms. These can be cognitive problems, or depression, or breathlessness, or general fatigue, and these can be very, very disabling. Already we're building up a big health burden with these cases. So that worries me a lot because we know nothing about it, and we really don't have much of a clue about its extent or how to treat it either.

**JOE: Before COVID, there was a lot of concern about a deadly flu pandemic. Looking forward, do you think flu is still the virus to watch for the next pandemic?**

ALAIN: I think there's no way of telling. One thing is for sure, an absolute certainty, that one day we will have another flu pandemic, whether it's going to be next year or in 100 years. And that is why finding a flu vaccine which can cross protect against all comers is such a priority. And I have to say that the work on COVID has accelerated that. There are very exciting papers coming out on ways of doing that. I think a flu pandemic is inevitable at some stage, and the key preparation is having the vaccine. Unfortunately, the viruses become resistant to drugs almost immediately. Drugs will be useful, but resistance will become a problem very quickly unless we find combinations of drugs where the virus would need to mutate in multiple sites at once to become resistant; like with the treatments for HIV.

**JOE: You've been working on some quite exciting projects recently, would you be able to talk us through that?**

ALAIN: I'm basically a flu-ologist, all my research life flu has played a central part, either as a model system to understand how the immune system works in the early days, or contemplating various ways of immunising against flu that might have a deeper level of protection than just the next seasonal virus. There are endless aspects to it that make it a very interesting virus. It also happens to be very adaptable to doing experimental work. More recently we've been working with Mark Howarth in biochemistry because he's developed a very beautiful piece of work. With modern vaccines, if you create your vaccine as a particle that looks like a virus - it seems rather obvious when you think about it - it's likely to work. A group in Washington under David Baker have been playing for many years now with ways of forming self-assembling particles. That work is absolutely beautiful. It's beautiful from the structural point of view: their structures of their particles are as varied as snowflakes. You can then use those structures to display antigens, so you end up with what looks like a virus. Linking an antigen to this particle is not so simple to do.

Now what Mark has worked out is a really simple way of doing this based on their studies on *Streptococcus Pyogenes*. A protein in *Streptococcus Pyogenes* has an internal peptide bond; it's very rare in proteins. To cut a long story short they analysed that protein and developed a way of dividing it into two components, one of which has a small peptide tag, and the other has the so-called capture domain. The catalysis of the peptide bond is primed within the capture domain, so you just mix the two together and the bond forms. So, you can then put the capture domain on your particle, and you can put your tag on any protein you like, and then literally mix them together. That gives you a universal way of making particles with any antigen you want. I heard about this quite by accident because I was going down to scrounge

some enzymes off Mark and, over breakfast (there is a very good restaurant in the Biochemistry Department) he told me all about this. The scales fell from my eyes: this is an immensely powerful and universal way of making protein vaccines.

We've been working with Mark over two or three years to link flu proteins to these particles. We we're working on this when COVID came along. COVID binds to a protein in you called ACE-2, but the way it does so is much simpler than the way that influenza binds to you. COVID has a separate little domain (called the Receptor Binding Domain or "RBD") which has the binding sites on it, and that domain behaves like a sperate protein: you can lop it off and refold and it works because it retains its structure. Whereas, with flu, it is much more difficult. So, the RBD was an obvious Achilles heel on the virus that would be very important to stick onto Mark's particles. We did that straight away and made a design that seemed logical and it just worked. So, we made the particles, the particles immunised mice and pigs very, very successfully, and we're working with a company called Ingenza to try and turn it into a practical vaccine. It's still in the lab at the moment, but we're very much hoping to get funding to develop and go forward to our phase one trial. The key is that Ingenza can make the whole thing in microorganisms that should make it very cheap to produce – they are aiming for 50p per dose, and it is easily adaptable to all the new SARS-CoV-2 Variants of Concern.

Another area that we've been working on is to produce a simple, inexpensive test for detecting COVID antibodies (<http://www.nature.com/articles/s41467-021-22045-y>). In the well-funded West, we have very sophisticated tests for measuring antibodies. But this is expensive, costing maybe 2 or 3 pounds per shot. If you're working in a tent hospital in the middle of a very isolated district of the world, you will not have this. A colleague in Toulouse, Etienne Joly, wrote to me and asked if anybody was thinking about how to make a simple test that doesn't require any sophisticated equipment. If you work with flu, the way you do this is use what's called haemagglutination: if you mix influenza with red blood cells it forms clumps and you can see this by eye. In the 1950s and 1960s, most of the techniques used to measure viral antibodies were using these sorts of red blood cell clumping tests. The trick with these kinds of tests is that you've got to link the viral antigen to the red cell. The way we've done this is to use some work from a group in Paris who used an antibody against the glycoporphin A molecule on the red cell. We just linked that antibody domain to the receptor binding domain of SARS CoV 2. That molecule then binds to red cells and labels them. If antibodies against the SARS CoV 2 receptor binding domain are present, they'll bind to the receptor binding domain that's now on the red cells and cross link them and you'll see the clumping by eye. It's very cheap. We've made 10 million tests worth, and I manage a small charity which is funding that with dedicated donations. So far, we've sent the test out to 21 countries, and we're still



inviting people to join. We're sending out lots of 10,000 tests, and they can then do at least some local work on positivity rates and learn for instance, particularly for healthcare workers, who has antibodies and who doesn't. You can do this test at the bedside with a fingerprick and it takes an hour to develop.

**JOE: It's been a pleasure speaking with you Alain, do you have any wisdom to leave us with?**

ALAIN: One thing I would add, and I think this is very, very important, is that the collaboration between groups, both in Oxford and elsewhere, has been tremendous. We've combined data and helped each other where we can: that has been heartwarming. Of course, as time goes by, things get more competitive, they're bound to, but this period has been rather special. 🌟

## Papers Published 2020–21

This is the fifth year in which *The New Collection* does not only publish its own articles, but also a list of articles by New College MCR members that are accepted by other peer-reviewed journals since the previous edition. We are happy to present and congratulate the following entries<sup>†</sup>:

- Christopher Archibald, 'Remembering the St. Bartholomew's Day Massacre in Elizabethan England', *Studies in Philology* 18.2 (Spring 2021), 242-283. DOI: <https://doi.org/10.1353/sip.2021.0009>.
- **R. Buchanan**, M. Camurri and M. Fallon, "Haptic Sequential Monte Carlo Localization for Quadrupedal Locomotion in Vision-Denied Scenarios," 2020 *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 3657-3663, doi: 10.1109/IROS45743.2020.9341128.
- **R. Buchanan**, Bednarek, J., Camurri, M. et al. Navigating by touch: haptic Monte Carlo localization via geometric sensing and terrain classification. *Auton Robot* (2021). <https://doi.org/10.1007/s10514-021-10013-w>
- **Hajonides, J. E.**, Nobre, A. C., van Ede, F., Stokes, M. G. (2021). Decoding visual colour from scalp electroencephalography measurements. *NeuroImage*, 237, 118030.
- **Hajonides, J. E.**, van Ede, F., Stokes, M. G., Nobre, A. C. (2020). Comparing the prioritization of items and feature-dimensions in visual working memory. *Journal of vision*, 20(8), 25-25.
- Yang, W., **Huang, R.**, Li, Y., Li, H. (2021). Training teacher-researchers through online collective academic supervision: Evidence from a postgraduate teacher education programme. *Journal of Computer Assisted Learning*.

---

<sup>†</sup>Referencing styles vary by discipline and are recorded as submitted by the author(s). New College members are given in bold where there are multiple authors.

- **Huang, R.**, Yang, W., Li, Y. and Li, H. (2021) Training teacher-researchers through collective supervision, *Researching Education* 2(4). doi.org/10.5281/zenodo.5121316
- New, K. A., 'On Intercultural Communication in Diachronic Translation', *Functional Aspects of Intercultural Communication. Translation and Interpreting Issues*, vol. 7, RUDN University, Institute of Foreign Languages, Moscow, 2021. Pp. 352-359.
- New, K.A., 'The Reception of Horace's Ode 3.30 in Russian poetry'. In: *In Lieu of Duration: Spatiotemporal Excursions in Literature*. Ed. M. Stasiowski. London: Interdisciplinary Discourses, 2021. Pp.44-64.
- New, K.A., 'The Language of Praise in Russian 18th Century Ode'. In: *Music, Poetry and Language*. Ed. K. Gunesch. London: Interdisciplinary Discourses, 2021. Pp.133-47.
- Ashley J. Redman, **Gabriel Moise**, Sabine Richert, Erin J. Viere, William K. Myers, Michael J. Therien, and Christiane R. Timmel, *J. Phys. Chem. C* 2021, 125, 21, 11782–11790
- **Andras Sandor**, Mark D Fricker, Verena Kriechbaumer, Lee J Sweetlove, InterResting structures: formation and applications of organized smooth endoplasmic reticulum in plant cells, *Plant Physiology*, Volume 185, Issue 3, March 2021, Pages 550–561, <https://doi.org/10.1104/pp.20.00719>
- Tzavaras, Andrew. "Cosimo's Crusades 1559-1563: Tuscany, the Military Order of Santo Stefano, and the Nature of Early Modern Crusades," in "Crusading Ideas and Fear of the Turks in Late Medieval and Early Modern Europe" [Collection Meridiennes Série Croisades tardives]. Toulouse: PUM Presses, forthcoming 2021, pp. 165-180.
- **Bingjun Wang**, Hao Ye, Moritz Riede, and Donal D. C. Bradley. Chain conformation control of fluorene-benzothiadiazole copolymer light-emitting diode efficiency and lifetime. *ACS Appl. Mater. Interfaces* 2021, 13, 2919–2931.
- Jingjing Xu, Lauren E. Jarocha, Tilo Zollitsch, Marcin Konowalczyk, Kevin B. Henbest, Sabine Richert, **Matthew J. Golesworthy**, Jessica Schmidt, **Victoire Déjean**, Daniel J. C. Sowood, Marco Bassetto, Jiatae Luo, Jessica R. Walton, Jessica Fleming, Yujing Wei, Tommy L. Pitcher, **Gabriel Moise**, Maike Herrmann, Hang Yin, Haijia Wu, Rabea Bartölke, Stefanie J. Käsehagen, Simon Horst, Glen Dautaj, Patrick D. F. Murton, Angela S. Gehrckens, Yogarany Chelliah,

Joseph S. Takahashi, Karl-Wilhelm Koch, Stefan Weber, Ilia A. Solov'yov, Can Xie, Stuart R. Mackenzie, Christiane R. Timmel, Henrik Mouritsen P. J. Hore, Magnetic sensitivity of cryptochrome 4 from a migratory songbird. *Nature* **594**, 535–540 (2021). <https://doi.org/10.1038/s41586-021-03618-9>

- Anjuli Shere, 2021: “Legal tiptoeing: the Integrated Review, Data-Sharing, and the EU”, for Wavell Room, <https://wavellroom.com/2021/06/22/legal-tiptoeing-the-integrated-review-data-sharing-and-the-eu/>
- Anjuli Shere, 2021: “COVID-19: IMPACTS ON MEDIA FREEDOM - Examining the effects on journalists and media organisations of the pandemic”, in AIB Member-only Media Intelligence Briefing
- Anjuli Shere, 2020: “Investigating the Information Commissioner’s Office: Is It Fit for Purpose?” (co-written with Miranda Melcher), published in RUSI Newsbrief, <https://rusi.org/publication/rusi-newsbrief/investigating-information-commissioner-office-it-fit-purpose>
- Anjuli Shere, 2020: “The Inaugural ACE Winter School” and “Hunted during a DPhil” in University of Oxford CDT in Cyber Security Yearbook 2020, <https://www.cybersecurity.ox.ac.uk/site-resources/uploads/2020/11/2020-yearbook-web.pdf>
- Anjuli Shere, 2020: “Now You [Don’t] See Me: How have the GDPR and a changing public awareness of the UK surveillance state impacted OSINT investigations?”, in the *Journal of Cyber Policy* (volume 5, issue 3), <https://www.tandfonline.com/doi/full/10.1080/23738871.2020.1832129>
- Anjuli Shere, 2020: “Security should be there by default”: Investigating how journalists perceive and respond to risks from the Internet of Things”, at The 5th European Workshop on Usable Security (EuroUSEC 2020), <https://eusec20.cs.uchicago.edu/eusec20-Shere.pdf>





Volume 15 | 2021  
ISSN 1757-2541

S. Nagler, *editor-in-chief*

M. Courtney, M. Grossi, N. Han, B. Hoover, A. Ibáñez Núñez, A. Laffere,  
J. Milton, J. Paek, O. Peters, J. Pitaluga, *editors*