

Revista Română de Filosofie Analitică

Romanian Journal of Analytic Philosophy

Volumul XII, Nr. 1-2, 2018



EDITURA UNIVERSITĂȚII DIN BUCUREȘTI
BUCHAREST UNIVERSITY PRESS

2023

Revistă semestrială publicată de Societatea Română de Filosofie Analitică,
cu sprijinul Facultății de Filosofie, Universitatea din București
Revue semestrielle publiée par le Société Roumaine de Philosophie Analytique,
soutenue par la Faculté de Philosophie, L'Université de Bucarest
Biannual journal published by the Romanian Society for Analytic Philosophy,
supported by the Faculty of Philosophy, University of Bucharest

Revista Română de Filosofie Analitică
este fosta Revistă de Filosofie Analitică (2007-2011)

© 2007, Societatea Română de Filosofie Analitică
www.srfa.ro

Ediția on-line: <http://www.srfa.ro/rrfa/>

DIRECTOR / DIRECTEUR / DIRECTOR

Mircea Dumitru, Universitatea din București

REDACTOR-ȘEF / ÉDITEUR-EN-CHEF / CHIEF-EDITOR

Constantin Stoenescu, Universitatea din București

REDACTORI / ÉDITEURS / EDITORS

Andrei Mărășoiu, Universitatea din București

Bianca Savu, Universitatea din București

Andreea Popescu, Universitatea din București

Bogdan Burunția, Universitatea din București

EDITOR ON-LINE

Lavinia Marin, Katholieke Universiteit Leuven

CONSILIUL ȘTIINȚIFIC / COMITÉ SCIENTIFIQUE / SCIENTIFIC BOARD

Marin Bălan, Universitatea din București

Radu J. Bogdan, Universitatea Tulane, New Orleans

Romulus Brâncoveanu, Universitatea din București

Cristian Calude, Universitatea din Auckland

Mircea Flonta, Universitatea din București

Mircea Dumitru, Universitatea din București

Radu Dudău, Universitatea din București

Kit Fine, Universitatea din New York

Adrian-Paul Iliescu, Universitatea din București

Hans-Klaus Keul, Universitatea din Ulm

Ilie Pârvu, Universitatea din București

Gabriel Sandu, Universitatea din Helsinki

Emanuel Socaciu, Universitatea din București

Wolfgang Spohn, Universitatea Contanz

Constantin Stoenescu, Universitatea din București

Ion Vezeanu, Universitatea Pierre-Mendès-France, Grenoble

REDAȚIA / SECRÉTARIAT / OFFICE

Universitatea din București, Facultatea de Filosofie

Splaiul Independenței nr. 204, București, 060024

E-mail: redactia@srfa.ro

ISSN (ediția electronică): 1843-9969

ISSN (ediția tipărită): 1844-2218



EDITURA UNIVERSITĂȚII DIN BUCUREȘTI
BUCHAREST UNIVERSITY PRESS

Bd. Mihail Kogălniceanu, nr. 36-46,
Cămin A (curtea Facultății de Drept),
Corp A, Intrarea A, etaj 1-2, Sector 5
050107, București – ROMÂNIA
Tel. + (4) 0726 390 815
E-mail: editura.unibuc@gmail.com
www.editura-unibuc.ro

LIBRĂRIA EUB-BUP
(Facultatea de Sociologie și Asistență Socială)
Bd. Schitu Măgureanu nr. 9, sector 2
010181 București – ROMÂNIA
Tel. +40 213053703

TIPOGRAFIA EUB-BUP
(Complexul LEU)
Bd. Iuliu Maniu nr. 1-3,
061071 București – ROMÂNIA
Tel.: +40 799210566

REDACTOR
Remus Tite

DTP / COPERTĂ
EUB-BUP

Revista Română de Filosofie Analitică

Vol. XII

Ianuarie-Iunie 2018

Nr. 1

CONTENTS

MIRCEA FLONTA, Cunoaşterea și valorile	7
CONSTANTIN STOENESCU, Translation and Linguistic Relativism. An Epistemological Approach	35
TUDOR MĂRGINEAN, Naturalism Reloaded: How Do We Construct Our World? – Review of Joseph Rouse, <i>Articulating the World. Conceptual Understanding and Scientific Image</i> (University of Chicago Press, 2015, 416 pp.)	53
*	
RRFA Vol. XII, 2018, Nr. 2	63

CUNOAȘTEREA ȘI VALORILE

MIRCEA FLONTA¹

Rezumat: Prin raportare la folosiri curente ale expresiilor *cunoaștere* și *valoare*, cât și la unele elaborări ale acestor concepte în literatura filosofică, se argumentează în favoarea unei abordări în care distincția cunoaștere-valori este gândită ca opoziție polară, ca distincție între condiționat și necondiționat, relativ și absolut.

Sunt evidențiate și discutate unele consecințe ce rezultă din această abordare pentru înțelegerea relației dintre cunoaștere și valori. Sunt evaluate, din această perspectivă, orientări tradiționale sau actuale ale gândirii, cum ar fi metafizica clasică, teologia dogmatică, variante ale scientismului și ale naturalismului etic.

Termeni-cheie: cunoaștere, valoare, reprezentări ale valorilor, viață, bine, frumos.

Ceea ce îmi propun prin considerațiile care urmează sunt clarificări ce pot oferi un spor de înțelegere.

În genere, abordarea relației dintre cunoaștere și valori – o temă perenă a gândirii filosofice – a depins în mare măsură de semnificația acordată acestor termeni. Chiar dacă aceasta este o constatare comună, cred că poate constitui un punct de plecare într-o discuție asupra raportului dintre cunoaștere și valori.

În lumea cercetătorilor, expresia *cunoaștere* este folosită în mod curent pentru enunțuri sau sisteme de enunțuri care sunt controlabile și

¹ Prof. univ. dr. Mircea Flonta este membru al Academiei Române și profesor emerit al Universității din București. Baza scrierii acestui text a constituit-o expunerea cu același titlu prezentată la 15 februarie 2021 la seminarul de cercetare al Departamentului de Filosofie Teoretică al Universității din București, în colaborare cu CELFIS, susținută la invitația moderatorului ei, dr. Andrei Mărășoiu. Țin să mulțumesc moderatorului pentru invitație și tuturor participanților la discuție pentru observații și sugestii. Ele mi-au fost de mare folos în redactarea finală.

întrunesc, în consecință, acordul tuturor celor ce examinează în mod imparțial faptele și raționează în mod corect. Este sensul în care susținerile ce reprezintă cunoaștere nu depind de dorințele, înclinațiile sau interesele cercetătorilor. Ele vor întruni consensul celor competenți și de bună credință. În cazul multor susțineri din viața cotidiană, spre deosebire de cazul afirmațiilor cercetătorilor, buna credință va fi, de multe ori, suficientă pentru obținerea consensului. Să ne gândim, bunăoară, la modul în care se obține consensul asupra faptelor în tribunale și în multe alte domenii ale activității cotidiene a oamenilor. Nu este însă mai puțin îndreptățită și observația că practica cercetării științifice moderne ne ajută să înțelegem mai bine cum se justifică pretențiile de cunoaștere, cum anume distingem, în afirmațiile și reprezentările noastre, ceea ce este cunoaștere de ceea ce nu este cunoaștere.

Se va remarca, fără îndoială, că în această caracterizare a cunoașterii, spre deosebire de multe dintre cele care au fost propuse de filosofi, lipsește referirea la adevăr.² Poate fi justificată o asemenea omisiune? În sprijinul unui răspuns afirmativ pot fi invocate două remarci. Prima este aceea că referirea la adevăr nu pare să fie indispensabilă pentru a distinge enunțuri ce reprezintă cunoaștere de alte enunțuri, adică pentru a arbitra între pretenții de cunoaștere care sunt în competiție. În sprijinul renunțării la adevăr pentru caracterizarea cunoașterii s-ar putea invoca, așadar, principiul desemnat prin expresia „briul lui Occam”. A doua observație este aceea că *adevărul* este un concept foarte controversat în filosofie. Au fost și sunt în circulație concepții diferite și ireconciliabile asupra adevărului. Abordarea numită *deflaționistă* va putea fi caracterizată drept o reacție față de rezultatele puțin concludente ale confruntărilor dintre cele mai influente concepții asupra adevărului. Din această perspectivă, o caracterizare a cunoașterii care nu se raportează la adevăr va putea fi apreciată drept un câștig.³

² O ilustrare, între altele, a locului central pe care îl acordă mulți filosofi adevărului în caracterizarea cunoașterii o reprezintă discuțiile asupra „definiției clasice a cunoașterii” din literatura aceluiași domeniu care este desemnat astăzi, în limba engleză, prin termenul *epistemology*.

³ Altminteri, eu nu am niciun fel de rezerve față de utilizarea atributului *adevărat* pentru susțineri din viața de fiecare zi sau pentru rezultate ale cercetării. Folosirea expresiilor *adevăr* și *adevărat* nu este, ce-i drept, de multe ori independentă de angajarea față de o

O delimitare a conceptului cunoașterii ce poate fi apreciată drept relativ neproblematică este, așadar, posibilă. Nu același lucru se va putea spune cu referire la conceptul *valoare*. Există folosiri foarte diferite ale expresiei în vorbirea curentă, ca și în scrierile filosofice. În limbajul comun, expresia apare drept termen de referință într-o mare varietate de judecăți ce exprimă aprecieri, estimări ale unor realități, ființe, reprezentări sau opere ale oamenilor. Merită amintit că *evaluarea* este o expresie sinonimă cu termenii *apreciere* sau *estimare*. S-ar putea spune, desigur, că asta ne interesează mai puțin, deoarece în discuție este valoarea ca temă a gândirii filosofice. Putem avea rezerve față de o asemenea observație. La fel ca în cazul altor concepte filosofice, examinarea relațiilor lor cu folosirea comună a cuvintelor poate să fie instructivă.

Dacă ne raportăm la elaborările și discuțiile filosofice, atunci ceea ce ne poate atrage atenția este asocierea frecventă și strânsă a conceptului cu distincția dintre *judecăți despre fapte* și *judecăți de valoare*. Aceasta este o constatare ce nu va surprinde, de vreme ce *valoarea* reprezintă termenul de referință al *judecăților de valoare*. Valorile ar fi acele entități la care ne raportăm atunci când ne formulăm aprecierile sau estimările pe care le exprimă judecățile numite *judecăți de valoare*. Judecățile de valoare par să fie avute în vedere adeseori atunci când unii filosofi își propun să răspundă la întrebarea "Ce este valoarea?"⁴ În literatura filosofică de limbă engleză,

anumită intuiție, cea a realismului gândirii comune, adică de presupunerea că acele afirmații despre fapte pe care le calificăm drept *adevărate* ar exprima caracteristici proprii faptelor ca atare. Cu toate acestea, de multe ori atunci când calificăm diferite afirmații drept *adevărate*, ceea ce avem în vedere este doar că orice persoană onestă, care este în cunoștință de cauză, va avea bune temeiuri pentru a le accepta.

⁴ Cea mai elaborată și sistematică cercetare asupra acestei teme, publicată în limba română, a rămas până astăzi cartea lui Ludwig Grünberg, *Axiologia și condiția umană*, București, Editura Politică, 1972. Valoarea este caracterizată aici drept o relație între subiect și obiect, relație prin care se exprimă prețuirea unor realități sau înfăptuiri ale oamenilor, prin raportare la capacitatea acestora de a satisface trebuințe și aspirații ale indivizilor și comunităților omenești. Autorul afirmă, de exemplu, că "...valoarea se exprimă întotdeauna în sentimente și judecăți cu caracter *imperativ* prin care se desemnează nu ceea ce *este*, ci ceea ce un individ sau o colectivitate, în condiții date, consideră că *trebuie să fie*, că este demn a fi dorit, prețuit, căutat, cucerit." (*op. cit.*, p. 33). Se menționează că judecățile de valoare își găsesc adesea expresia în norme adoptate și acceptate în comunități omenești. Pentru prezentarea și comentarea considerațiilor

discuția s-a concentrat de multe ori asupra examinării critice a distincției dintre *este și trebuie*, o distincție consacrată deja de David Hume. Analizele unor filosofi cunoscuți, dintr-o perioadă mai recentă, și-au propus să arate că distincția fapte/valori nu este o distincție polară, ci una relativă. S-a argumentat că din relativitatea distincției judecăm despre fapte/judecăm de valoare decurge și relativitatea distincției dintre fapte și valori.⁵

Dacă tot ceea ce apare drept termen de referință în judecățile de valoare va fi socotit *valoare*, atunci această expresie va căpăta un domeniu de referință extrem de larg, de cuprinzător. Vorbirea curentă, în care circulă expresii ca *valori economice, administrative, juridice, politice, ale sănătății, educative, intelectuale, ale cercetării, artistice, literare, sau chiar ale divertismentului și sportive*, ilustrează foarte bine spectrul larg de întrebuițări pe care le poate căpăta expresia *valoare* dacă ea va fi utilizată pentru a indica etalonul la care se raportează aprecierile sau estimările formulate în cele mai diferite domenii ale vieții și activității oamenilor. Se va recunoaște și sublinia, desigur, că în orice comunitate omenească există anumite valori care sunt socotite fundamentale și că pe acestea le au în vedere filosofii și intelectualii în genere, atunci când ei vorbesc despre valori. Totuși, dacă conceptul valorii va fi strâns asociat cu ceea ce numim în mod curent *judecată de valoare*, atunci trasarea graniței ce desparte asemenea valori fundamentale de subiectul multor altor judecăm de valoare se va dovedi dificilă.

Există o linie de gândire diferită care poate fi urmată în delimitarea conceptului valorii. Angajarea pe această linie este încurajată de promisiunea delimitării clare a lumii valorilor și a lumii faptelor, cu consecințe semnificative în ceea ce privește înțelegerea relațiilor dintre ele. Această linie de gândire poate fi prezentată după cum urmează.

Există o ierarhie a străduințelor și a aspirațiilor, a năzuințelor oamenilor, atât ale indivizilor cât și ale colectivității. Multe obiective ale acestor străduințe și năzuințe sunt, cum se știe foarte bine, subordonate altor obiective, socotite mai înalte. În desfășurarea acestei succesiuni există însă o limită, reprezentată de obiective ce nu mai pot fi subordonate altor obiective. Aceste obiective pot fi calificate drept obiectivele ultime,

formulate de Grünberg, vezi și *Misterul judecăm de valoare* în Adrian Miroiu, *Între logică și etică*, Iași, Institutul European, 2020, pp. 129-161.

⁵ Vezi, de exemplu, Hilary Putnam, *The Collapse of the Fact/Value Dichotomy*, Harvard University Press, 2002.

obiectivele supreme. Expresia *valoare* poate fi folosită, și a fost folosită, pentru a desemna asemenea obiective și numai asemenea obiective. Anumite reprezentări ale unor astfel obiective au apărut în comunități omenești atunci când dezvoltarea acestora a ajuns pe trepte mai înalte. Am în vedere reprezentări cum sunt cele ce susțin, de exemplu, învățăături morale cu o bază religioasă sau altele care nu au o asemenea bază, cum a fost budismul sau ceea ce propovăduiau Confucius și Socrate.

Voi menționa, în cele ce urmează, exprimări ale unor autori care au gândit valorile ca țeluri ultime. Pentru ei, distincția dintre lumea valorilor și lumea faptelor, a cunoașterii noastre despre fapte, era distincția dintre ceea ce este absolut, necondiționat, și ceea ce este relativ, condiționat. Adică o distincție polară, de natură, nu una graduală. Gândite drept sisteme de referință absolute la care sunt raportate strădaniile și aspirațiile comunităților omenești și ale indivizilor, valorile se situează în raport cu cercetarea faptelor și cu activitățile bazate pe cercetarea faptelor în termenii unei opoziții polare. S-ar putea spune că, în tradiția filosofică, diferite încercări de a elabora distincția dintre relativ și absolut, de exemplu în termenii opoziției dintre o lume sensibilă și una inteligibilă, au trasat cadrele în care poate fi gândită relația dintre valori și fapte, dintre valori și cunoașterea faptelor. Exemplare pentru o asemenea abordare pot fi socotite, mi se pare mie, reflecții care au fost formulate, cu un secol în urmă, de către doi cunoscuți autori de limbă germană, Max Weber și Ludwig Wittgenstein.

În scrierile sale, Weber revenea asupra observației că știința, cunoașterea prin excelență, nu poate să ofere orientarea fundamentală comunităților omenești deoarece ea nu ne poate spune spre ce trebuie să năzuim, cum ar trebui să trăim. Cunoașterea științifică este, ce-i drept, extrem de importantă atunci când este vorba de identificarea căilor potrivite pentru atingerea multor obiective, inclusiv pentru indicarea unor obiective ale activității noastre, dar ea nu ne spune nimic cu privire la scopurile ultime, la țelurile supreme. Într-o conferință susținută în anul 1918, Weber compara valorile cu zeii, iar lupta pentru supremație între valori e comparată cu înfruntările dintre zei, remarcând că „peste luptele lor domnește destinul, în nici un caz « știința ».”⁶ Pe cei văzuți drept purtători ai valorilor unor mari comunități

⁶ Max Weber, *Știința – profesie și vocație*, traducere de Ida Alexandrescu, Editura Humanitas, 2011, p. 64.

omenești, Weber îi numea „profeți” sau „mântuitori”, scriind: „Iar acum, dacă punei din nou problema în chip tolstoian și întrebați: « Cine răspunde, de vreme ce știința nu o face, la întrebarea ce trebuie să facem și cum trebuie să trăim? » sau dacă punei întrebarea « Pe care dintre zei în luptă trebuie să-i slujim? Sau poate ar trebui să slujim un alt zeu, dar care anume? », atunci vă spun: adresați-vă unui profet sau unui mântuitor.”⁷ Cei pe care îi avea în vedere Weber când vorbea de „profeți” sau „mântuitori” erau personaje, cum sunt cele pe care Karl Jaspers le-a numit „oamenii care dau măsura”: Socrate, Buddha, Confucius sau Isus.⁸ Să reținem că doar unul dintre aceștia, Isus, a fost propovăduitorul unei religii, în sensul strict al cuvântului. Toți au fost în căutarea a ceea ce am putea numi „binele suprem”, menit să călăuzească viața comunităților omenești.

În anul în care Weber a susținut conferința amintită, Wittgenstein încheia scrierea lucrării sale *Studiu logico-filosofic*, publicată mai târziu sub titlul *Tractatus logico-philosophicus*. Una dintre temele ei centrale este opoziția dintre lumea valorilor și lumea faptelor. „Lumea ca întreg” este opusă faptelor ce constituie obiectul cercetării. În opoziție cu cercetarea care se interesează de fapte, reflecția moral-religioasă și creația artistică se raportează la lume ca întreg, la valori. În exprimarea autorului, la ceea ce „nu se poate spune”, dar „se arată” prin reflecția morală, religioasă, filosofică și prin capodopere ale creației artistice. Valorile sunt calificate de Wittgenstein drept „ceea ce este mai înalt” și delimitate strict în raport cu lumea, caracterizată drept „totalitate a faptelor” (1.1). În paragraful 6.41, autorul exprimă astfel contrastul dintre fapte și valori: „Sensul lumii trebuie să se afle în afara ei. În lume totul este cum este și totul se întâmplă cum se întâmplă; nu există în ea nicio valoare și dacă ar exista nu ar avea nicio valoare. Dacă există o valoare care are valoare, atunci ea ar trebui să stea în afara oricărui eveniment și a unui anumit fel de a fi. Căci orice eveniment și orice fel de a fi sunt întâmplătoare.”⁹

Este clar că opoziția fapte/valori este gândită de acești autori în termenii opoziției polare relativ/absolut, condiționat/necon condiționat, trecător/etern.

⁷ *Ibidem*, p. 70.

⁸ Vezi Karl Jaspers, *Oamenii de însemnătate crucială*, traducere de Alexandru Al. Sahidian, București, Editura Paideia, 1996.

⁹ Ludwig Wittgenstein, *Tractatus logico-philosophicus*, traducere de Mircea Dumitru și Mircea Flonta, Humanitas, 2012, ediția a II-a, p. 185.

Adică cu totul altfel decât atunci când valoarea este raportată la ceea ce numim în mod obișnuit *judecăți de valoare*. Această linie de gândire poate fi urmărită în multe scrieri și însemnări ale lui Wittgenstein, îndeosebi în *Tractatus* și în *Conferința despre etică*. Subliniind opoziția dintre valori și fapte, Wittgenstein insista asupra observației că limbajul este apt să descrie faptele, adică ceea ce este relativ și condiționat, dar inadecvat pentru exprimarea a ceea ce este absolut, necondiționat, adică a valorilor. Orice încercare de a vorbi despre valori așa cum vorbim despre fapte, sublinia el, generează *nonsensuri*. Ceea ce se poate vedea foarte bine în cazul limbajului religios, un limbaj în care se încearcă să se vorbească despre valori. Bunăoară, Dumnezeu va fi descris drept o ființă ale cărei puteri și capacități sunt nelimitate, ceea ce este un nonsens.

În cele ce urmează, voi încerca să explorez implicații și consecințe ale însușirii acestui concept al valorii.

Dacă valoarea va fi gândită drept ceea ce este absolut, necondiționat, atemporal, drept reperul ultim în orientarea strădaniilor ființelor și comunităților omenești, atunci prima întrebare care se va pune este ce anume am putea identifica în mod îndreptățit drept *valoare*. Întrebarea este cu totul firească. Cel ce distinge valorile drept ceea ce este absolut, necondiționat, va trebui să se întrebe ce anume ar putea fi caracterizat în acest fel în orizontul vieții omenești. În cele ce urmează, voi încerca să răspund la această întrebare.

Se poate susține că o valoare este ființa umană însăși. Și aceasta deoarece păstrarea și conservarea vieții reprezintă un țel care nu va putea fi subordonat vreunui alt țel și, în acest sens, va putea fi caracterizat drept obiectiv ultim, suprem sau final. Este tocmai ceea ce pare să fi avut în vedere Kant, în una dintre formulările pe care le dădea imperativului categoric, ca obligație absolută, necondiționată. Afirmând că natura rațională este "un scop în sine", Kant scria: „Acționează astfel încât să folosești umanitatea, atât din persoana ta, cât și din persoana oricui altcuiva, de fiecare dată totodată ca scop, niciodată numai ca mijloc.”¹⁰

¹⁰ Immanuel Kant, *Întemeierea metafizicii moravurilor*, traducere de Valentin Mureșan *et al.*, Humanitas, 2006, p. 75. În această scriere, Kant va caracteriza păstrarea și prelungirea vieții atât drept o înclinație a oamenilor cât și ca o datorie. Iată un pasaj semnificativ în acest sens: "În schimb, a-ți conserva viața este o datorie și, în plus, orice om are o înclinație nemijlocită pentru asta. Dar grija, adesea plină de teamă, pe

Ceea ce constituie, de asemenea, un obiectiv absolut, necondiționat, este coeziunea colectivităților omenești. De coeziunea colectivității depinde în mod hotărâtor atât ocrotirea, cât și calitatea vieții indivizilor care o compun. Această coeziune, susținută atât de înțelegerea rațională a dependenței individului de ceea ce îi oferă colectivității mai restrânse sau mai cuprinzătoare, cât și de acele sentimente de afecțiune și compasiune pentru semenii care pot fi atât de puternice încât să susțină uneori mari sacrificii personale, se exprimă în tipare de comportare care, dincolo de expresiile și motivațiile lor foarte diferite, au un caracter universal. O mare diversitate de obiceiuri, datini, tradiții ale comunităților omenești, de imperative și norme, reprezintă tot atâtea modalități de a promova înțâietatea intereselor supraviețuirii și prosperității colectivității. Prin toate acestea se exprimă, în moduri dintre cele mai diferite, recunoașterea binelui comunității drept finalitate ultimă, drept țintă supremă. Nu pare să existe vreo reprezentare cu autoritate în comunități omenești cu privire la ceea ce trebuie acceptat drept datorie pentru fiecare dintre membrii lor care să nu ținască, în mod direct sau mijlocit, promovarea și susținerea bunăstării colectivității. În procesul socializării, acel proces prin care fiecare ființă omenească devine membru al unei comunități, are loc tocmai formarea și consolidarea conștiinței datoriei morale, a cărei bază o constituie recunoașterea primatului binelui colectivității. Autoritățile invocate pot fi și sunt foarte variate, în locuri și epoci diferite, dar tâlcul ultim al obligației morale poate fi apreciat ca fiind pretutindeni același.

Este de remarcat că istoricilor din vechime, ca și moraliștilor de mai târziu, le-a atras în mod deosebit atenția marea diversitate a reprezentărilor despre bine și rău, a datinilor, tradițiilor și obiceiurilor, a regulilor și normelor urmate de comunități omenești distanțate în spațiu și timp. Ceea ce nu va surprinde dacă ținem seama de faptul că, inclusiv

care majoritatea oamenilor o au pentru acest lucru, nu are totuși o valoare lăuntrică...Ei își conservă viața, ce-i drept, *conform datoriei* dar nu și *din datorie*. În schimb, atunci când necazurile și o mâhnire fără speranță i-au luat unui nefericit tot cheful de viață și când el, având un suflet puternic, se dovedește mai mult indignat de soarta sa decât descurajat și resemnat și își dorește moartea, dar totuși își conservă viața fără a o iubi, nu din înclinație sau frică, ci din datorie, abia atunci maxima sa are un conținut moral." (*op. cit.*, pp. 34-35)

în societăți care sunt înrudite cultural, lista celor mai prețuite virtuți morale va putea fi destul de diferită. De asta ne putem convinge citind, de exemplu, scrieri ale lui Platon și Kant. Marea varietate a reprezentărilor și tradițiilor morale nu poate ascunde însă faptul că, dincolo de toate diferențele, prin aceste tradiții se exprimă conștiința că binele comunității ar trebui să stea deasupra altor obiective, aspirații și îndatoriri ale ființelor omenești, că acest bine reprezintă un obiectiv ultim. Dacă păstrarea și prelungirea vieții pot fi apreciate drept obiective necondiționate, atunci *binele*, înțeles în acest fel, va trebui să fie apreciat drept un obiectiv căruia trebuie să i se acorde același rang. Este ceea ce putem înțelege dacă avem în vedere că păstrarea și prelungirea vieții fiecărei persoane, la fel ca buna ei stare, depinde pretutindeni, înainte de orice altceva, de puterea și coeziunea comunității în care trăiește. Iată de ce contribuția la creșterea acestor valori va constitui etalonul universal în aprecierea intențiilor și acțiunilor indivizilor. Dincolo de tot ceea ce deosebește comunitățile omenești, se impune constatarea că ființa umană nu poate supraviețui și nu se poate împlini decât în colectivitate. Tocmai din această perspectivă putem înțelege mai bine vechea sentință că omul este o ființă socială.

Temeiurile invocate în sprijinul caracterizării păstrării și prelungirii vieții drept obiectiv necondiționat, absolut, vor fi, așadar, în egală măsură valabile cu referire la promovarea coeziunii colectivității. Sunt două ținte ale străduințelor oamenilor deopotrivă îndreptățite să fie calificate drept obiective ultime, necondiționate. Adică drept *valori*.

Merită făcută, fie și în trecut, observația că în viața oamenilor situațiile normale sunt cele în care menținerea și prelungirea vieții indivizilor, pe de o parte, și binele comunității, pe de altă parte, sunt ținte ce pot fi armonizate. Există însă situații în care ele pot intra în conflict. Sunt situații ce pot fi calificate drept excepționale. În asemenea cazuri, obiective ca păstrarea vieții sau promovarea fericirii unor persoane devin incompatibile cu binele comunității. Este un conflict al valorilor, adesea înfățișat în mitologie, literatură și artă. Este acel conflict ce poate fi calificat drept o sursă majoră a tragicului. Subordonarea deplină a vieții individului cauzei colectivității, împinsă uneori până la jertfa vieții, are măreția tragicului.

Cum vom raporta credințele religioase la acest concept al valorii? Este clar că multe dintre reprezentările religioase care au dominat mintea oamenilor pot fi greu armonizate cu concluzia că păstrarea vieții și calitatea ei, ce depind în mod hotărâtor de puterea și coeziunea comunității, ar reprezenta obiective ultime, necondiționate. Adepții unor foarte influente tradiții religioase, cum sunt creștinismul sau islamismul, cred în supraviețuirea sufletului și în viața după moarte. În mod firesc, pentru ei obiectivul necondiționat va fi fericirea în viața veșnică, iar absolutul va fi sacru, divinul. În raportarea la sacru vor căuta ei sursa și temeiul tuturor lucrurilor cărora le conferă semnificație și valoare. Ceea ce ține de viața comună va fi valorizat din această perspectivă. Bunăoară, existența ființelor omenești, deoarece este dată și ocrotită de divinitate, iar servirea comunității, deoarece reprezintă răspunsul dat poruncilor acesteia.

Gânditori religioși din trecut și din epoci mai recente au formulat o mare varietate de considerații ce converg spre concluzia că sacru este valoarea în sine, sursa și temeiul a tot ceea ce are semnificație și însemnătate majoră în viața oamenilor. Atrag atenția îndeosebi acele analize care urmăresc să sublinieze distincția dintre credința religioasă și conștiința morală, precum și primatul absolut al credinței. Accentul cade pe evidențierea naturii iraționale a trăirii religioase.¹¹ Dincolo de tot ceea ce desparte reflecțiile formulate de diferiți autori asupra acestei teme, ceea ce îi apropie este susținerea că prezența și trăirea sacruului ar fi o permanență a conștiinței umane. Este ceea ce au afirmat și unii cercetători ai religiilor, de exemplu Mircea Eliade. Religiozitatea, ca trăire a sacruului, ar fi o structură ultimă a condiției umane. Înțelegând în acest fel, religiozitatea nu va dispărea, chiar dacă reprezentările și

¹¹ Rudolf Otto, de exemplu, afirma că esența credinței religioase, care s-ar fi exprimat cel mai bine în religii arhaice, o constituie trăirea sacruului. Ceea ce are în vedere sunt, înainte de toate, sentimente de dependență față de ceea ce credinciosul resimte drept o putere absolută, nemărginită. Pierderea de sine și „cutremurarea lăuntrică” a creaturii în fața puterii supreme ar putea fi exprimată prin cuvintele „Eu sunt nimic. Tu ești totul”. Este sensul în care, pentru cel credincios, mânia acesteia nu are nimic de-a face cu condamnarea morală. Însemnătatea unei tradiții religioase ar fi dată, înainte de toate, de modul cum pune ea în valoare trăirea sacruului. Din această perspectivă, Otto aprecia creștinismul drept o religie mai desăvârșită decât altele. (Vezi Rudolf Otto, *Sacru*, traducere de Ioan Milea, Cluj Napoca, Editura Dacia, 1996.)

practicile religioase cunoscute până acum nu ar mai fi prezente. Eliade insista asupra observației că omul profan al zilelor noastre, pentru care existența este desacralizată, ar avea totuși reacții și comportări prin care s-ar exprima conștiința existenței sacrului și trăirea acestuia.¹²

Cercetarea nu oferă temeuri pentru susținerea ipotezei că sursa și temeiul credințelor în existența sacrului ar fi natura umană însăși. Totodată, cu greu s-ar putea contesta că totuși credințele religioase, ca sisteme de reprezentări și practici, răspund unor nevoi și aspirații ale comunităților omenești. Și că puterea și influența lor asupra vieții oamenilor va fi cu atât mai mare cu cât lipsesc mai mult căi alternative de satisfacere a unor nevoi și aspirații fundamentale ale ființelor omenești. Nu există, totuși, temeuri principiale pentru o susținere ca aceea că credința religioasă constituie baza conștiinței obligațiilor morale, că acestea ar exista și ar dispărea odată cu religia. Cu deosebire evoluții mai recente din societatea occidentală nu confirmă afirmația, adesea reluată, că religia ar oferi singurul suport sigur al moralității în genere, că conștiința obligațiilor oamenilor față de semenii lor nu ar putea supraviețui în lipsa credinței religioase. Și că, în genere, fără religie viața oamenilor ar fi lipsită de sens. Căci dacă suportul moralității ar fi credința religioasă, atunci existența în zilele noastre a unor comunități omenești cu o viață mai bună decât a multor altora din trecut, comunități formate din persoane care în majoritatea lor au un mod de a simți, a gândi și a acționa lipsit de motivații religioase, ar fi inexplicabilă. Atât cercetări asupra a ceea ce este constitutiv naturii umane în genere, cât și multe informații istorice nu susțin ideea mereu reafirmată conform căreia credințele religioase ar constitui fundamentul conviețuirii în comunitățile omenești.¹³ În societățile occidentale, la fel ca în alte societăți antrenate într-un proces rapid de modernizare, nu pot fi identificate manifestări semnificative ale conștiinței sacrului, ale distincției dintre sacru și profan, nu doar în trăirile și comportările persoanelor nereligioase, dar nici măcar în cele ale marii mase a

¹² Pentru dezvoltarea acestei teme, vezi Mircea Eliade, *Sacrul și profanul*, traducere de Brîndușa Prelipceanu, Editura Humanitas, 1995.

¹³ Pentru o discuție amplă asupra acestei teme, vezi Daniel C. Dennett, *Breaking the spell. Religion as a natural phenomenon*, Penguin Books, 2006.

oamenilor convențional religioși, adică a acelor care nu ar putea fi caracterizați drept *religioși* într-un sens mai strict al cuvântului.

S-a afirmat adesea, nu fără bune temeiuri, că multe credințe religioase ar merita respect deoarece ele susțin și promovează conștiința îndatoririlor morale ale oamenilor. Cu greu s-ar putea contesta că în cazul celor care cred în răsplată și în pedeapsă în viața după moarte, cu deosebire în pedepsele înfricoșătoare care i-ar aștepta pe cei ce încalcă percepțe morale fundamentale ale comunității, credința religioasă va fi în măsură să promoveze moralitatea, cel puțin în anumite direcții. Chiar dacă, așa cum s-a remarcat adesea, această constatare nu susține cea mai reconfortantă imagine asupra demnității persoanei umane. Oricum ar sta lucrurile în această privință, este un fapt că o mare varietate de precepte și imperative, cuprinse atât în învățături ale unor religii universale, cât și în cele ale sectelor religioase au susținut și susțin opțiuni și activități ce pot promova binele comunității drept îndatoriri religioase. Iar capacitatea credințelor religioase de a promova asemenea activități și opțiuni va trebui apreciată drept una considerabilă dacă ținem seama de faptul că pentru cei care le împărtășesc ele vor apărea drept porunci ale puterii supreme. De urmarea lor, și numai de urmarea lor, va atârna mântuirea, obiectivul suprem al credinciosului. Multe dintre reprezentările și poruncile religioase din epoci diferite și locuri diferite și îndepărtate converg, dincolo de ceea ce le deosebește, prin contribuția lor la promovarea coeziunii comunităților omenești.¹⁴ Aceasta pare să fie una dintre sursele viabilității lor. Cu greu pot fi trecute cu vederea sacrificiile aduse în slujba ajutorării semenilor de atâția oameni din diferite epoci și culturi, precum și multe alte fapte nobile, săvârșite sub impulsul unor convingeri religioase. Numeroase

¹⁴ Este interesantă observația că personalități ferm angajate în promovarea binelui general, a căror educație a avut o bază religioasă, nu au trebuit să-și revizuiască valorile morale atunci când și-au pierdut credința religioasă. Din acest punct de vedere este semnificativă mărturia lui Rudolf Carnap. El primise, în familie, o educație religioasă convențională. Carnap scria, cu referire la momentul în care a încetat să mai nutrească convingeri și sentimente religioase: "Evaluările mele morale au fost după aceea în mod esențial aceleași ca și mai înainte." (*Intellectual Autobiography of Rudolf Carnap*, în (ed.) A. P. Schilpp, *The Philosophy of Rudolf Carnap*, Open Court, 1963, p. 9).

observații asupra vieții unor comunități religioase și asupra comportării unor persoane cu convingeri religioase confirmă asemenea constatări. Este un adevăr că pentru aceste persoane obligațiile de ordin moral, înțelese ca porunci ale divinității, capătă o putere de convingere și o autoritate incomparabilă cu cea ce le poate fi conferită de către înclinații altruiste sau reflecții raționale. Este ceea ce au subliniat, nu o dată, cercetători ai comunităților omenești, etnologi, antropologi sau sociologi. Și afirmația, atât de des reluată, că dacă Dumnezeu nu ar exista, atunci totul ar fi permis, ar putea fi mai bine înțeleasă din această perspectivă.

Pot fi formulate, totodată, obiecții cu mare greutate împotriva valorizării globale a credințelor și practicilor religioase. O mare varietate de informații istorice arată că autoritatea credințelor religioase și a slujitorilor acestora a fost de multe ori utilizată pentru promovarea unor inițiativă și activități ce contraveneau în mod flagrant cerințelor ocrotirii vieții și promovării binelui comunităților omenești. Temeiuri de natură religioasă au fost invocate adesea pentru a justifica exterminarea sau oprimarea unor populații, pentru legitimarea unor stări și relații sociale profund nedrepte, a autorității arbitrare, ca și pentru condamnarea și reprimarea celor ce acționau în vederea înlăturării acestora. Tradițiile religioase și interdicțiile legate de ele au fost invocate de multe ori pentru a zădărnici o mare varietate de inițiativă și activități constructive. Într-o perspectivă istorică largă, examinarea conduitelor și a activităților promovate de credințele religioase, a inițiativelor și deciziilor autorităților religioase, în relație cu ocrotirea vieții și cu binele comunității, ni se înfățișează drept un *Janus*, drept o figură cu două fețe. Iată de ce pledoariile *pro* sau *contra* credințelor religioase, din perspectiva aceluși concept al valorii ale cărui contururi încerc să le schițez, vor putea, desigur, să apară drept convingătoare. Ele nu vor putea câștiga, însă, adeziunea deplină a unui judecător imparțial de îndată ce acesta va ține seama de tot ceea ce se poate spune în sprijinul sau, dimpotrivă, în defavoarea a ceea ce a fost inițiat, promovat și susținut în unele credințe religioase, de-a lungul unei îndelungate istorii, în comunități omenești răspândite pe întregul glob.

Un candidat pentru titlul de valoare este *frumosul*, într-o semnificație largă a termenului, și anume frumosul ca afirmare a vieții și a puterii vieții. Există bune temeuri pentru a susține că frumosul, în această

accepție, reprezintă o valoare, adică o țintă finală, necondiționată, a aspirațiilor și străduințelor oamenilor. În societăți dintre cele mai diferite, în epoci mai îndepărtate sau mai apropiate, oamenii și-au dorit frumosul și au cultivat frumosul drept scop în sine, adică drept un obiectiv care nu este subordonat vreunei alte finalități.

Ca afirmare a puterilor vieții, frumosul capătă o mare varietate de expresii atât în fenomenele naturale și cosmice, cât și în creațiile oamenilor. În societățile dezvoltate, evoluția și diversificarea reprezentărilor frumosului au loc, în primul rând, sub influența creației artistice. Operele artistice pot fi considerate drept tot atâtea transfigurări idealizate ale puterii vieții. Tocmai deoarece frumosul este afirmarea vieții și a puterii ei, sensibilitatea pentru frumos este înrădăcinată în natura umană. Puterea de atracție irezistibilă a frumosului exprimă reacția sensibilității față de tot ceea ce reprezintă manifestare și afirmare a puterii vieții. Căci dacă facem abstracție de rafinările și distilările pe care le-a cunoscut sensibilitatea pentru frumos în culturile înalte, putem constata că, la nivelul de bază, distincția dintre frumos și urât este susținută de reacția spontană diferită a ființei sensibile față de ceea ce este viu și mort, față de ceea ce este în creștere, față de ceea ce își afirmă puterea și față de ceea ce este în descreștere, în declin, în degradare. Semnificativă în acest sens mi se pare a fi și observația că toate acele lucruri care ni se par frumoase sunt resimțite ca și cum ar fi însuflețite.

Binele și frumosul împărtășesc cu viața atributul absolutului, al necondiționatului. Prin raportare la distincția între ceea ce este absolut, necondiționat, și ceea ce este condiționat și relativ, binele comunității și frumosul pot fi înfățișate drept două tulpini a căror rădăcină comună este viața.

Înclinațiile ce susțin promovarea binelui comunității ca și aspirația spre frumos, înclinații înrădăcinate în natura umană, se exprimă și se afirmă printr-o mare varietate de reprezentări ale valorilor. Sunt reprezentări ce diferă destul de mult de la o cultură la alta. De fiecare dată când vorbim de valorile unei colectivități omenești, ne referim nu la valori ca atare, ci la reprezentările despre valori ale unei anumite comunități. Iar ceea ce desemnăm adesea prin expresia *judecată de valoare* sunt aprecieri formulate prin raportare nu la valori ca atare, ci la reprezentări despre valori ale diferitelor comunități și grupuri de oameni. Este vorba despre toate

acele sentimente, idei, atitudini și comportări prin care se exprimă o valoare. Astfel, acea conștiință ce susține și promovează binele comunității se exprimă într-o mare varietate de reprezentări privitoare la relațiile interumane, cum sunt cele despre iubire, prietenie, mărinimie, compasiune, empatie, milă, fidelitate, loialitate, respect, devotament, sinceritate, dreptate, echitate, cinste, corectitudine și altele, care diferă adesea mult de la o cultură la alta. Marile culturi ne înfățișează o diversitate copleșitoare de reprezentări și expresii ale frumosului. Nivelul de dezvoltare atins de o anumită cultură va putea fi apreciat, între altele, prin raportarea la bogăția și varietatea reprezentărilor pe care le capătă valorile binelui și ale frumosului. Sunt acele reprezentări ce susțin idealurile morale sau estetice care orientează și însuflețesc viața indivizilor și îi apropie unii de alții. Variabilitatea istorică a acestor reprezentări nu poate să fie îndeajuns subliniată. Bunăoară, în societățile democratic-liberale din zilele noastre, reprezentările valorii morale par să fie dominate de cuplul *libertate - dreptate*. Primul termen, *libertatea*, evidențiază însemnătatea autonomiei persoanei, ca ființă rațională responsabilă. Al doilea, *dreptatea*, importanța orientării inițiativelor și a realizărilor personale de către obiectivul care este binele comunității. Ca reprezentări ale valorii morale, libertatea și dreptatea se înfățișează ca cerințe și imperative necondiționate. Preocupările pentru armonizarea intereselor personale cu cele ale comunității se exprimă, atât în opinia comună, cât și în elaborări teoretice, în reprezentări asupra dreptății. Se poate spune că relația libertate - dreptate a reprezentat axa centrală a reflecției morale moderne, de la Immanuel Kant la John Rawls.

Apropierea dintre bine și frumos a fost remarcată și subliniată nu o singură dată. Confucius, bunăoară, spunea că frumosul nu este frumos fără a fi și bun, iar binele nu este bun dacă nu este și frumos. Mai categoric s-a exprimat în această privință Wittgenstein. În *Tractatus* (6.421) întâlnim observația: „Etica și estetica sunt *unul și același lucru*”. Nu este prea clar ce anume avea în vedere autorul. Poate că el dorea să semnaleze distanțarea sa față de toți cei ce nu apreciază ceea ce am putea numi „frumusețea binelui”. Sau poate dorea să-și afirme dezacordul cu gânditori influenți pe atunci în cultura germană, cu cei care apreciau frumosul, ca afirmare a puterii vieții, drept unică valoare. Reprezentative pentru o asemenea poziție sunt multe din reflecțiile lui Nietzsche. Supraomului,

ființă ce se exprimă prin joc și prin creația artistică, binele comunității îi este indiferent. Prin glorificarea supraomului, Nietzsche se opunea unor influente tradiții ale gândirii occidentale, în primul rând mesajului central al culturii elenistice și iudeo-creștine. Afirmarea vieții prin creație era pentru el singura valoare. Iubirea aproapelui, solidaritatea colectivității le vedea ca opuse afirmării vieții și le respingea pe temeiul unei viziuni care susține afirmarea neîngrădită a voinței de putere. Prin reflecții mereu reluate asupra acestei teme, Nietzsche a scris o pagină aparte în analele gândirii elitiste. Sublinierea apropierei frumosului de bine va putea fi, așadar, înțeleasă și ca o reacție față de orientări ale gândirii care posedă o putere de atracție ce ar putea fi cu greu tăgăduită.¹⁵

La întrebarea dacă cunoașterea reprezintă o valoare, prin raportare la acel concept al valorii pe care l-am înfățișat și explorat în considerațiile anterioare, chiar titlul textului meu sugerează un răspuns negativ. Căci cunoașterea și valorile sunt calificate drept „lumi polare”.

Ce-i drept, cunoașterea va putea fi socotită o valoare dacă ea va putea fi înțeleasă și cultivată ca scop în sine. Adică dacă ea ar fi cultivată așa cum sunt frumosul și binele. Cunoașterea înțeleasă în acest mod a fost ceea ce au avut în vedere autori din vremuri mai îndepărtate. Ceea ce aveau ei în vedere era cunoașterea înțeleasă drept dezvăluire a principiilor existenței, acea cunoaștere care a fost numită, după Aristotel, *metafizică*. Rezultatele gândirii speculative nu reprezintă însă cunoaștere în sensul conceptului modern, restrictiv al cunoașterii obiective, pe care l-am prezentat mai sus.¹⁶ Unii cercetători ai naturii au afirmat, ce-i drept, că

¹⁵ Seducția pe care o pot exercita reflecțiile elitiste o vor resimți puternic și cititorii multor pasaje din scrierile unor autori ca Emil Cioran. Mărturisind că „mișcarea vieții” a fost divinitatea tinereții sale, Cioran scria: „Pe cel care n-a cunoscut, până la 30 de ani, fascinația tuturor extremismelor, mă întreb dacă trebuie să-l admir sau să-l disprețuiesc, să-l consider un sfânt sau un cadavru... Lipsindu-i și dorința și voința de a distruge, el este suspect, căci l-a învins pe demon, sau, mai grav, nu a fost posedat niciodată. A trăi cu adevărat înseamnă să-i refuzi pe ceilalți; ca să-i accepți, ar trebui să știi să renunți, să te stăpânești, să acționezi împotriva propriei tale naturi, să te debilitezi; libertatea o concepi doar pentru tine însuși; aproapelui o cedezi doar cu prețul unor eforturi epuizante.” (Emil Cioran, *Istorie și utopie*, traducere de Emanoil Marcu, Humanitas, 1992, pp. 7-8)

¹⁶ Este acea cunoaștere pe care o oferă, în primul rând, rezultatele cercetării științifice, o cunoaștere al cărei mesaj a fost promovat de orientări filosofice empiriste și

obiectivul ultim al străduințelor lor ar fi atingerea unei cunoașteri ce nu servește atingerea unor obiective exterioare, mai mult sau mai puțin practice. De exemplu, Einstein a formulat unele observații semnificative în acest sens, în scrierile sale despre știință adresate publicului larg. Astfel, el aprecia ceea ce numea „religiozitate cosmică” drept impulsul cel mai puternic și mai nobil al cercetării științifice. Ceea ce avea în vedere Einstein prin această expresie era „uimirea extaziată față de armonia legității naturale”. Țelul suprem al cunoașterii științifice ar fi pătrunderea a ceea ce el numea „grandoarea rațiunii încarnate în existență”.¹⁷ Cu alte cuvinte, pătrunderea a ceea ce am putea numi „ordinea fundamentală a lumii”. Este îndoielnic că atingerea unui asemenea țel, reprezentat de o „teorie finală”, este un obiectiv ce va putea fi atins. Einstein însuși a eșuat în toate încercările sale îndreptate în această direcție. Merită să ne amintim că, începând cu a doua jumătate a secolului al XIX-lea, cercetători ai naturii de cel mai înalt rang, bunăoară Max Planck și Werner Heisenberg, au fost preocupați să schițeze o imagine generală a lumii bazată pe sinteza informațiilor oferite de descoperirile științifice, ceea ce s-a numit în germană „das wissenschaftliche Weltbild”. În măsura în care asemenea imagini erau generalizări de mare anvergură, generalizări care nu puteau fi controlate în mod strict, prin raportare la datele experienței, ele nu reprezentau cunoaștere în acel sens strict al termenului care a fost schițat mai sus. Asigurarea obiectivității cunoașterii, ca și orice alt lucru care este important, are un preț mare. Acesta este limitarea așteptărilor și pretențiilor noastre. Cunoașterea obiectivă este restrânsă și modestă. Ea

pozitiviste ale gândirii moderne. La confluența mesajului acestor orientări cu reflecții ale unor mari cercetători ai naturii asupra identității cunoașterii științifice a căpătat contururi clare punctul de vedere că toate cunoștințele autentice, atât cele obținute prin cercetări curente din viața cotidiană, cât și prin cercetarea științifică specializată vizează formularea unor răspunsuri controlabile la întrebări clare, bine delimitate. În această accepție, cunoașterea nu oferă adevăruri eterne, deoarece ea nu răspunde la întrebări privitoare la natura ultimă a realității. Oamenii pot nutri, desigur, diferite reprezentări în această privință. Desemnarea acestora prin același termen cu rezultatele cercetării obiective a faptelor va fi apreciată însă drept derutantă și, în consecință, drept nerecomandabilă.

¹⁷ Vezi Albert Einstein, *Cum văd eu lumea*, traducere de Mircea Flonta, Ilie Pârvu, Dragan Stoianovici, Humanitas, 2015, îndeosebi pp. 256, 268 și 289.

este incapabilă să satisfacă mari ambiții, să răspundă acelor așteptări care au fost nutrite, nu o dată, de grandioase reflecții asupra cauzelor și temeiurilor ultime a tot ceea ce există. Multe dintre acestea aparțin trecutului. Astăzi, nimbul lor nu mai este atât de strălucitor. Nu lipsesc, totuși, încercări de a reînvia și de a întreține asemenea așteptări.

Tocmai în virtutea obiectivității lor, cunoștințele obținute prin cercetare se pot dovedi utile pentru înfăptuirea unei mari varietăți de obiective. Suntem conduși astfel la concluzia că acea reprezentare despre cunoaștere ca scop în sine, care a susținut și susține caracterizarea ei drept valoare, pare să fie o himeră. În lumea de azi, înalta considerație de care se bucură cercetarea științifică se exprimă, înainte de toate, în conștiința rolului ei în întreținerea și ameliorarea tuturor activităților de care depinde viața și prosperitatea comunităților omenești. Pe de altă parte, cu greu s-ar putea susține că ceea ce recunoaștem drept cunoștințe prin excelență ne-ar oferi resurse pentru a propune sau a întemeia valori. Căci cunoașterea nu indică țelurile acțiunilor noastre, ci căi potrivite pentru a le atinge.

Dacă ne punem întrebarea de unde provine fascinația pe care a exercitat-o și o mai exercită și astăzi asupra minții multor oameni promisiunea cunoașterii absolutului, un răspuns posibil este acela că o asemenea cunoaștere ar oferi baza necesară pentru fundamentarea și asigurarea definitivă a reprezentărilor despre valori ale comunităților omenești. O sugestie interesantă în acest sens oferea Rudolf Carnap, discutând acea aspirație a cunoașterii absolutului care a însuflețit metafizica clasică, într-o lucrare de tinerețe. Metafizica – scria Carnap – „izvorăște din nevoia de a exprima sentimentul vieții, atitudinea față de viață a oricărui om, poziția acestuia în ordinea afectivă și volițională față de lumea înconjurătoare, față de semenii.”¹⁸ Autorul făcea observația că arta, și mai ales muzica, reprezintă un mijloc adecvat de exprimare a acestui sentiment al vieții, spre deosebire de metafizică. El remarca de asemenea că autori ca Nietzsche, a căror concentrare asupra discuției valorilor este manifestă, nu recurg la elaborări teoretice ci preferă să se exprime în mod artistic.

¹⁸ Vezi Rudolf Carnap, *Depășirea metafizicii prin analiza logică a limbajului*, în Rudolf Carnap, *Vechea și noua logică*, traducere de Alexandru Boboc, Editura Paideia, 2001, pp. 75-76.

Din toate considerațiile pe care le-am formulat cu privire la cunoaștere și la valori se desprinde concluzia că lumea faptelor (și a cunoașterii faptelor) și lumea valorilor sunt lumi polare. Relația dintre ele ar putea fi comparată cu relația dintre lumea fenomenelor și lumea *noumenelor* în filosofia lui Kant. Dincolo de deosebirile dintre societăți și culturi, în toate comunitățile reprezentările despre valori fixează obiective ultime, necondiționate ale aspirațiilor și străduințelor oamenilor și oferă astfel orientarea fundamentală. Cunoștințele determină căile prin care se încearcă atingerea unor țeluri dependente, în cele din urmă, de reprezentări ale valorilor. Fiind neutre în raport cu valorile, resursele oferite de cunoaștere, în primul rând de cercetarea științifică, vor putea fi utilizate atât pentru a putea susține activități orientate de obiective în raport cu reprezentări dominante despre valori, cât și activități cu obiective incompatibile cu orientările ce rezultă din aceste reprezentări. Pe temeiul acestor corelații, lumea valorilor și lumea cunoașterii ni se înfățișează drept lumi nu numai polare, dar și interdependente. De aderența sau lipsa aderenței indivizilor sau grupurilor de oameni la reprezentări despre valori dominante într-o comunitate va depinde în mod hotărâtor utilizarea cunoștințelor, precum și a tuturor capacităților și resurselor generate de progresele cunoașterii. Prin raportare la reprezentări despre valori care diferă adesea mult de la o cultură la alta, se va distinge între ceea ce comunitățile omenești califică drept *activități constructive*, respectiv *activități distructive*. Sunt activitățile ale căror obiective și rezultate sunt în acord sau în contradicție cu ceea ce membrii unei comunități ar trebui să facă urmând imperatiivele ce rezultă din aceste reprezentări. Cunoașterea este prima și cea mai importantă resursă de care a depins și depinde în măsură tot mai mare eficacitatea acțiunilor indivizilor, grupurilor și colectivităților. În societățile moderne, ea este produsă în primul rând de cercetarea științifică. A spune că întreaga cunoaștere a unei comunități, în particular cunoașterea obținută prin cercetare științifică, este neutră în raport cu valorile înseamnă a afirma că ea poate fi utilizată pentru a asigura, în egală măsură, eficacitatea acțiunilor pe care le-am numit *constructive* sau *distructive*. Modul cum a fost folosită până acum puterea pe care o oferă cunoașterea, îndeosebi prin mijlocirea tehnologiei, ca și evaluarea posibilităților folosirii cunoștințelor noastre în viitor, ne avertizează în această privință. Nu voi intra în discuția acestei teme, mult invocate și

intens discutate. Voi spune doar că vechea sentință – știința fără conștiință înseamnă pieire K este mai actuală decât oricând.

Există și o altă dimensiune a relației dintre lumea cunoașterii și lumea valorilor, a reprezentărilor comunităților omenești despre valori, dimensiune ce poate fi apreciată drept fundamentală. Acele reprezentări ale valorilor care modelează modul de a gândi și de a simți al unei comunități omenești sunt, sub multe alte aspecte, influențate de starea cunoașterii. Cu greu s-ar putea contesta că modul de a gândi și de a simți este în multe privințe foarte diferit în cazul unor comunități a căror stare a cunoașterii se deosebește foarte mult. Bunăoară în cazul comunităților arhaice ce mai supraviețuiesc astăzi pe glob, în raport cu societățile cele mai avansate din punct de vedere tehnologic. Viața de fiecare zi a oamenilor din aceste societăți este în multe privințe foarte diferită, ceea ce nu poate să nu influențeze, sub multe aspecte, configurarea reprezentărilor dominante despre viață, despre bine și frumos. Deosebiri vor fi atât de mari și de izbitoare încât un observator mai superficial s-ar putea întreba ce au în comun membrii unor asemenea comunități în afara unor asemănări de ordin biologic, ce anume îi apropie ca ființe culturale.

Considerațiile pe care le-am formulat cu privire la cunoaștere și valori, la relația dintre lumea cunoașterii și lumea valorilor, conduc la unele observații despre orientări ale gândirii, influente în trecut sau în zilele noastre.

Prima dintre ele privește o mare tradiție de gândire, tradiția metafizicii clasice, inaugurată de eleați și de Platon. Ne putem întreba care este miza pretenției ce a susținut această venerabilă tradiție, pretenția că rațiunea omului oferă o cunoaștere apreciată drept supremă, cunoașterea principiilor a tot ceea ce există. De ce au năzuit atâția filosofi, de-a lungul atâtor generații, spre o asemenea cunoaștere? De ce li s-a părut ea atât de importantă? Putem presupune că ceea ce i-a preocupat a fost să susțină stabilitatea unei anumite ordini sociale. Se poate presupune că ei au crezut că doar cunoașterea absolutului ar fi în măsură să asigure autoritatea acelor reprezentări despre valori care susțineau cultura lumii în care trăiau. Reflecții cum sunt reflecțiile lui Platon despre bine, despre locul binelui în lumea ideilor, îndeosebi caracterizarea binelui drept idee supremă, par să susțină această presupunere. Din aceeași perspectivă pot fi privite și sistemele metafizice elaborate de gânditori din secolele

XVII-XVIII, cum au fost Descartes, Spinoza, Leibniz sau Wolff. Sunt construcții ce pot fi caracterizate drept tot atâtea încercări de a oferi un fundament stabil reprezentărilor despre valori ale lumii occidentale din acea vreme. Pentru omul de rând, prestigiul reprezentărilor curente asupra vieții omenești este asigurat, înainte de toate, de autoritatea tradiției, în primul rând a tradiției religioase. Pentru cei ce afirmau, însă, supremația rațiunii, adică pentru oamenii învățați ai vremii, asigurarea acelor reprezentări prin fundamentarea lor pe ceea ce se pretindea a fi cunoașterea principiului suprem al existenței capătă o însemnătate eminentă. Chiar dacă evoluțiile ulterioare ale gândirii filosofice au slăbit tot mai mult încrederea în posibilitatea unei cunoașteri a absolutului, ceea ce s-a exprimat în abandonarea ambițiilor grandioase ale metafizicii clasice, această încredere a fost reafirmată, cu o consecvență neslăbită, în acea tradiție raționalistă care rămâne vie și puternică în teologia catolică până în zilele noastre.¹⁹

Tocmai prezentarea principiilor învățaturii creștine drept cunoașterea cea mai înaltă, drept cunoașterea supremă, a făcut ca incompatibilitatea

¹⁹ Reprezentativă în această privință este cuvântarea lui Joseph Ratzinger, *Glaube, Vernunft und Universität*, ținută la Universitatea din Regensburg în anul 2006, când era papă cu numele *Benedict al XVI-lea*. Vorbitorul deplângea declinul încrederii în posibilitatea de a atinge, prin rațiune, o cunoaștere a „Întregului”. Elogiind „acordul profund” dintre încrederea în rațiune a tradiției filosofice grecești și credința în Dumnezeu, întemeiată pe Biblie, care și-ar fi găsit expresia în sentința evanghelistului „La început a fost Logos-ul și Logos-ul este Dumnezeu”, Ratzinger critica ceea ce a numit „des-elelizarea creștinismului”, adică pierderea încrederii în posibilitatea întemeierii valorilor prin cunoașterea absolutului. El indica trei valori ale acestei evoluții pe care o deplângea. Primul val ar fi fost inaugurat de Reformă și bine ilustrat de gânditori cum este Kant, care contestă posibilitatea cunoașterii absolutului, a lui Dumnezeu. Al doilea val ar fi fost reprezentat de teologia liberală protestantă a secolelor XIX-XX. Iar al treilea, ultimul – de viziunea instrumentală asupra cunoașterii promovată de ideologia democratic-liberală din zilele noastre. Vorbitorul își încheia mult-comentata lui conferință cu următoarele reflecții: „Vestul este amenințat de această aversiune față de întrebările fundamentale ale rațiunii sale și poate, prin asta, să îndure doar o mare pierdere. Curaj în afirmarea amplitudinii rațiunii, nu refuz al măreției ei – acesta este programul cu care o teologie angajată față de credința biblică pășește în disputa contemporană.” Filosoful Jürgen Habermas a apreciat această luare de poziție drept un răspuns negativ la întrebarea dacă teologia creștină ar trebui să răspundă la provocările rațiunii moderne, pe care el o caracteriza drept „rațiune post-metafizică”.

între reprezentări de natură cosmologică ale acestei învățături, susținute și promovate de teologia și filosofia scolastică și neoscolastică, și concluzii derivate din rezultate ale cercetării științifice să apară drept un conflict între pretenții de cunoaștere incompatibile. Imaginea creaționistă și antropocentristă a universului a fost atât de strâns asociată cu reprezentările despre valori promovate de tradiția creștină încât autorități bisericești și civile au putut să creadă că cele din urmă vor fi primejduite de îndată ce acea imagine a universului ale cărei fundamente erau astronomia geocentrică și fizica aristotelică ar fi fost zdruncinată. Iată de ce nimic nu le-a apărut mai important decât apărarea acestor reprezentări. Acestea au fost socotite ca fiind mai presus de orice. Pare instructiv să privim și să examinăm, din această perspectivă, decizii cum au fost condamnarea lui Giordano Bruno sau a lui Galileo Galilei.

Aceasta este o perspectivă care nu s-a bucurat de prea multă atenție din partea istoricilor. Ea este adusă în lumină în piesa lui Bertolt Brecht, *Viața lui Galilei*. În Scena a VIII-a a piesei, intitulată *O convorbire*, Galilei discută cu un tânăr călugăr, versat în noua știință matematică a naturii. Interlocutorul îi atrage atenția marelui cercetător asupra consecințelor acelor descoperiri științifice care conduceau la distrugerea cosmosului, adică a universului închis și ierarhizat, ca întreg finit și bine ordonat. Călugărul îi explică lui Galilei cum înțelege el „înțelepciunea Decretului Sfântului Oficiu”, documentul care condamnase învățătura copernicană. Decretul – afirma călugărul – „mi-a dezvăluit primejdiile pe care le ascunde în sine o cercetare mult prea neînfrânată și am decis să renunț la astronomie”. Prezentându-i lui Galilei temeiurile deciziei sale, el insista asupra consecințelor nedorite pe care le-ar putea avea zdruncinarea încrederii într-o lume antropocentrică, acea lume în care tot ce se întâmplă are un sens și o finalitate, pentru oameni simpli și necăjiți, așa cum erau părinții săi, țărani din Campania: „Au fost asigurați că ochii lui Dumnezeu sunt îndreptați spre ei, cercetători, aproape stăpâniți de grijă; că întregul teatru al lumii din jurul lor a fost construit astfel încât ei, actorii, să se poată afirma în rolurile lor, mari sau mici. Ce-ar spune oare ei dacă ar afla de la mine că se află pe un mic bulgăre de piatră care, învârtindu-se neîncetat în spațiul gol, se mișcă în jurul unei alte constelații, una dintre multe altele, destul de lipsită de însemnătate! La ce ar mai fi necesară sau bună o asemenea răbdare, un asemenea

asentiment cu suferința lor? La ce mai e bună Sfânta Scriptură, care a explicat și întemeiat totul drept necesar, sudoarea, răbdarea, foamea, supunerea, dacă acum s-a găsit că ea este plină de erori? Nu, eu văd că ei se uită speriați... că se simt trădați și înșelați. Nu stă o privire asupra noastră, spun ei... Nu există niciun sens în suferința noastră, foamea nu înseamnă decât a nu fi mâncat, oboseala înseamnă doar a te apleca și a trage ceva după tine, nu un merit. Înțelegeți oare că eu citesc în decretul Sfântului Oficiu o nobilă pasiune maternă, o mare bunătate sufletească?"²⁰ Mesajul este clar. Cercetarea are un impact direct asupra acelei imagini a universului de care sunt legate reprezentări dominante despre valori ale unei colectivități dintr-o anumită epocă. Iată de ce angajarea față de aceste reprezentări ne va cere să apărăm cu orice preț acea imagine de amenințarea pe care o reprezintă noi rezultate ale cercetării.

Sursa a ceea ce, de la începuturile epocii moderne și până în zilele noastre, a apărut drept un conflict între știință și credința religioasă ne apare astfel drept un conflict între noi rezultate ale cercetării și acea imagine generală a lumii care susține reprezentări despre valori ale credinței creștine. Devenise clar că dacă aceasta imagine este subminată, atunci și acele reprezentări vor fi în mod iremediabil zdruncinate. Inflexibilitatea poziției autorității bisericești, în epoci mai îndepărtate sau mai apropiate, va putea fi mai bine înțeleasă, chiar dacă nu justificată, din această perspectivă. Suportul pare să fi fost supoziția că reprezentările asupra valorilor sunt susținute de cunoașterea cea mai înaltă, cunoașterea principiilor a tot ceea ce există.

Atât metafizica tradițională, cât și teologia dogmatică au întreținut încrederea în existența unei asemenea cunoașteri. Iar în toate culturile dezvoltate s-a încercat întemeierea reprezentărilor despre valori pe pretenții despre cunoaștere de acest gen. Din perspectiva conceptului restrictiv al cunoașterii obiective, a acelei cunoașteri ce constituie obiectivul cercetării atât în științe, cât și în viața curentă, cunoașterea absolutului ne apare însă drept o contradicție în termeni, prin raportare la conceptul cunoașterii schițat mai sus. Încrederea în posibilitatea de a sprijini valorile prin raportare la acest concept va fi astfel în mod inevitabil zdruncinată. Adoptarea conceptului restrictiv al cunoașterii

²⁰ Bertolt Brecht, *Leben des Galilei*. 8. Ein Gespräch.

obiective conduce la o separare clară și netă a cunoașterii de lumea valorilor. Din perspectiva unui asemenea concept, devine limpede că, oricât de mult ar progresa cunoașterea, ea nu va fi în măsură să indice și să fundamenteze țelurile ultime ale strădaniilor indivizilor și ale comunităților omenești. Una dintre exprimările curente ale acestei concluzii este aceea că întrebările cele mai importante la care trebuie să răspundă până la urmă ființe omenești ce gândesc în mod independent, întrebările privitoare la țelurile ultime, supreme, sunt cele la care cercetarea științifică nu este capabilă, în principiu, să ofere răspunsuri.

Cum bine se știe, lumea în care trăim este în măsură tot mai mare structurată și schimbată de progresele cunoașterii științifice. De aceea, și distincția dintre sublinierea importanței capitale a progresului cunoașterii științifice și afirmarea supremației acesteia devine tot mai greu de trasat. Ceea ce numim *scientism* este tocmai tendința de a atribui cunoașterii științifice competențe pe care ea nu le are. Iar una dintre manifestările curente ale scientismului o constituie tendința de a caracteriza și evalua credințe și reprezentări tradiționale din societăți arhaice sau moderne exclusiv din perspectiva cunoașterii științifice. Această tendință este bine ilustrată de Richard Dawkins, un binecunoscut cercetător și popularizator al biologiei evoluției, în cartea sa despre religie, intitulată *The God Delusion*.²¹ Autorul califică afirmarea existenței lui Dumnezeu drept o „ipoteză”. Constatând că această ipoteză nu este susținută de datele experienței, el trage concluzia că existența lui Dumnezeu a fost „infirmată”. Supoziția tacită a unui asemenea demers este evident aceea că fundamentul credinței religioase ar consta în susțineri de același gen cu cele ce constituie obiectul cercetării științifice.²² Ceea ce se pierde din vedere este că perenitatea și puterea de atracție a credințelor religioase a fost asigurată, în trecut ca și astăzi, de acele reprezentări despre valori pe care le promovează ele, și nu de afirmații despre fapte, afirmații care, cel puțin în principiu, pot fi susținute sau infirmate prin cercetare științifică. Dacă acceptăm, însă, că credințele religioase sunt legate de întrebări

²¹ Vezi și traducerea românească de Victor Godeanu, sub titlul *Himera credinței lui Dumnezeu*, București, Editura Curtea Veche, 2007.

²² Merită semnalat, fie și în trecut, că propaganda ateistă din Uniunea Sovietică și din țările blocului sovietic pornea de la aceeași supoziție. Nu întâmplător denumirea ei oficială era „Propaganda ateist-științifică”.

privitoare la orientarea vieții, întrebări ce se situează într-un orizont diferit de cel al cercetării, atunci vom fi conduși la concluzia că rezultatele celei din urmă sunt tot atât de puțin apte să zdruncine convingerile religioase pe cât sunt de puțin capabile să le susțină. Căci ceea ce oferă ele oamenilor nu este cunoaștere, ci îndrumarea vieții. Este ceea ce s-a străduit să arate, încă mai demult, Lev Tolstoi, bunăoară în scrierea sa *Biblia pe scurt*. Caracteristică pentru credința religioasă, susținea scriitorul rus, este o afirmație ca aceea că „adevărata viață este în afara timpului”.²³

Distincția dintre cunoaștere și cea îndrumare a vieții pe care o oferă reprezentările despre valori este subliniată de figura cercetătorului cu convingeri religioase. Este vorba de un personaj ce poate fi întâlnit și astăzi în mediile științifice. Prezența lui nu va fi greu de înțeles dacă ne vom gândi că cercetarea ne spune tot mai multe despre ceea ce este dar nimic despre ceea ce trebuie să fie. Iată de ce mulți oameni, inclusiv în lumea cercetătorilor, se vor îndrepta, atunci când caută răspunsuri la întrebări privitoare la orientarea vieții, spre credințele religioase, ale căror răspunsuri sunt simple și accesibile. În atașamentul lor față de asemenea credințe se va exprima adesea legătura strânsă cu tradiția familiei și a comunității în care s-au născut. Viabilitatea și vitalitatea credințelor religioase primește astfel o explicație simplă. Iată ce scria Ludwig Wittgenstein, referindu-se la distincția dintre „probleme științifice” și „probleme ale vieții”: „Noi simțim că până și atunci când toate problemele științifice *posibile* primesc un răspuns, problemele noastre de viață nu sunt câtuși de puțin atinse.”²⁴

Reprezentative pentru tendințele de a afirma supremația gândirii științifice sunt încercări de a întemeia drepturi ale omului pe rezultate ale cercetării faptelor. Este vorba de acele cercetări care conduc la concluzia că există caracteristici universale ale nevoilor și ale aspirațiilor

²³ Vezi Leo Tolstoy, *The Gospel in brief*, traducere de Isabelle Hapgood, University of Nebraska Press, Lincoln and London, 1997. În *Introducerea* cărții sale, Tolstoy caracteriza învățătura creștină drept “cea mai pură și mai completă doctrină a vieții și cea mai înaltă lumină pe care mintea omenească a atins-o vreodată, o doctrină din care derivă în mod instinctiv cele mai nobile activități ale omenirii în politică, știință, poezie și filosofie.” (p. 32)

²⁴ Ludwig Wittgenstein, *op.cit.*, par. 6.52, p. 188.

ființelor omenești. Apărarea vieții, nevoia de securitate, îngrijirea urmașilor, solidaritatea familială, aspirația de a avansa într-o ierarhie, dreptatea ca reciprocitate sunt unele dintre acestea. Ele conferă contururi noi conceptului tradițional de natură umană. Unii autori contemporani susțin că drepturile omului, acele drepturi care sunt afirmate și promovate de instituțiile democrației liberale, ar putea fi întemeiate pe trăsături universale ale ființelor omenești, trăsături identificate prin cercetare. Sunt susțineri ce ilustrează o poziție care ar putea fi caracterizată drept o variantă a naturalismului etic.²⁵

Încercarea de a legitima instituțiile democrației liberale, la fel ca așteptarea ca acestea să devină universale, pe rezultate ale cercetării științifice, este susținută de supoziții în mod evident incompatibile cu acea înțelegere a cunoașterii și a valorilor, precum și a relației dintre ele pe care am schițat-o mai sus. Concluzia la care suntem conduși, dacă aderăm la această înțelegere, este că drepturile omului, în accepția care li se dă în lumea occidentală contemporană, pot și trebuie să fie întemeiate nu pe caracteristici ale naturii umane ce constituie obiectul cercetării, ci pe reprezentări dominante despre valori. Pe de altă parte, ar fi greu să se conteste că prin reprezentările diferitelor comunități despre țeluri absolute, necondiționate, ale aspirațiilor și străduințelor omenești se exprimă trăsături ale naturii umane, trăsături ce pot să devină și devin obiect al cercetării. În lumina considerațiilor de mai sus, asemenea trăsături ar putea fi caracterizate drept condiții de posibilitate ale reprezentărilor diferite și schimbătoare despre valori, acele reprezentări ce deosebesc și despart culturile comunităților omenești. Putem presupune că ipotetice ființe raționale și sociale diferite de cele din specia *homo sapiens* ar forma comunități cu nevoi și aspirații sub anume aspecte diferite. Că asemenea ființe și-ar putea reprezenta altfel sensul și țelurile ultime ale existenței. Au fost și pot fi formulate speculații interesante pe această temă. Dacă susținem că în reprezentările comunităților omenești despre valori se exprimă și trăsături ce disting natura umană în genere, atunci va trebui să admitem că aceste

²⁵ Vezi, în acest sens, Roger D. Masters, *Beyond Relativism. Science and Human Values*, London, University Press of England, 1993; Larry Arhart, *Darwinian Natural Right: the biological ethics of human nature*, New York State University of New York Press, 1998; și îndeosebi Francis Fukuyama, *Our Posthuman Future*, Profile Books, 2003.

reprezentări ar putea fi sensibil diferite la ființe sociale și raționale care s-ar deosebi sensibil de cele din specia noastră.

Observație finală. Putem admite că orice încercare de delimitare a unor concepte centrale ale gândirii prezintă avantaje și dezavantaje, implică deopotrivă câștiguri și pierderi. Am asumat această supoziție în discuția de față asupra implicațiilor și consecințelor unei anumite înțelegeri a conceptelor *cunoaștere* și *valoare*.

BIBLIOGRAFIE

- Arhart, Larry. *Darwinian Natural Right: the biological ethics of human nature*, New York: State University of New York Press, 1998.
- Brecht, Bertolt. *Leben des Galilei*. Frankfurt am Main: Suhrkamp Verlag, 1998.
- Carnap, Rudolf. *Depășirea metafizicii prin analiza logică a limbajului*, în *Vechea și noua logică*, ed. Rudolf Carnap, traducere de Alexandru Boboc. București: Editura Paideia, 2001.
- Carnap, Rudolf. *Intellectual Autobiography of Rudolf Carnap*, în *The Philosophy of Rudolf Carnap*, ed. Paul Arthur Schilpp. La Salle, Ill.: Open Court, 1963.
- Cioran, Emil. *Istorie și utopie*, traducere de Emanoil Marcu. București: Humanitas, 1992.
- Dawkins, Richard. *Himera credinței lui Dumnezeu*, traducere de Victor Godeanu. București: Editura Curtea Veche, 2007.
- Dennett, Daniel. *Breaking the spell. Religion as a natural phenomenon*. London: Penguin Books, 2006.
- Einstein, Albert. *Cum văd eu lumea*, traducere de Mircea Flonta, Ilie Pârțu, Dragan Stoianovici. București: Humanitas, 2015.
- Eliade, Mircea. *Sacrul și profanul*, traducere de Brîndușa Prelipceanu. București: Humanitas, 1995.
- Fukuyama, Francis. *Our Posthuman Future: Consequences of the Biotechnology Revolution*. London: Profile Books, 2003.
- Jaspers, Karl. *Oamenii de însemnătate crucială*, traducere de Alexandru Al. Șahighian. București: Editura Paideia, 1996.
- Kant, Immanuel. *Întemeierea metafizicii moravurilor*, traducere de Filotheia Bogoiu, Valentin Mureșan, Miki Ota, Radu Gabriel Pârțu. București: Humanitas, 2006.

- Masters, Roger. *Beyond Relativism. Science and Human Values*. London: University Press of England, 1993.
- Miroiu, Adrian. *Între logică și etică*. Iași: Institutul European, 2020.
- Otto, Rudolf. *Sacrul*, traducere de Ioan Milea. Cluj Napoca: Editura Dacia, 1996.
- Putnam, Hilary. *The Collapse of the Fact/Value Dichotomy*. Cambridge, Mass.: Harvard University Press, 2002.
- Ratzinger, Joseph (Benedict al XVI-lea). *Glaube, Vernunft und Universität*, Universitatea din Regensburg, 2006. Text accesat în data de 17 octombrie 2021: <https://epub.uni-regensburg.de/406/>
- Tolstoy, Leo. *The Gospel in brief*, traducere de Isabelle Hapgood. Lincoln and London: University of Nebraska Press, 1997.
- Weber, Max. *Știința – profesie și vocație*, traducere de Ida Alexandrescu. București: Humanitas, 2011.
- Wittgenstein, Ludwig. *Tractatus logico-philosophicus*, traducere de Mircea Dumitru și Mircea Flonta. București: Humanitas, 2012, ediția a II-a.

TRANSLATION AND LINGUISTIC RELATIVISM. AN EPISTEMOLOGICAL APPROACH¹

CONSTANTIN STOENESCU²

Abstract: The purpose of this study is to provide a reconstruction of the philosophical discussions generated by the issue of translation from one language to another. Modern philosophers have already observed that language influences the way we think.

The hermeneutic tradition was followed by the establishment of a linguistic research tradition whose first doctrinal thesis was to notice the relativistic consequence of the plurality of languages. Later, epistemological relativism also underwent a linguistic turn. Exploratory concepts such as radical translation, indeterminacy of translation, paradigm and incommensurability, conceptual scheme, translation and interpretation were discussed.

Keywords: "Linguistic turn", linguistic relativism, translation, interpretation, meaning and truth.

1. Looking for the starting point. Two research traditions

The story is well-known by everyone. We find it in *Genesis*, 11, 4-9:

4. Then they said, 'Come, let us build ourselves a city, with a tower that reaches to the heavens, so that we may make a name for ourselves; otherwise we will be scattered over the face of the whole earth.'

¹ This paper was written as a support for a lecture given at a workshop organized in Luxembourg by the Directorate-General for Translation, June 2017.

² Constantin Stoenescu is Professor and Director of the Department for Theoretical Philosophy at the University of Bucharest. Contact: constantin.stoenescu@filosofie.unibuc.ro

5. But the Lord came down to see the city and the tower the people were building. 6. The Lord said, 'If as one people speaking the same language they have begun to do this, then nothing they plan to do will be impossible for them. 7. Come, let us go down and confuse their language so they will not understand each other.'

8. So the Lord scattered them from there over all the earth, and they stopped building the city. 9 That is why it was called Babel[c] – because there the Lord confused the language of the whole world. From there the Lord scattered them over the face of the whole earth.³

This biblical story teaches us that language is the cement of social cooperation, and that the diversity of languages could lock or close different communities and the channels of communication and common understanding. But we can overcome these difficulties if we are able to translate our languages into each other so that mutual understanding becomes possible. Moreover, to be able to speak another language and to translate from one language into another presupposes a language. This became a cultural virtue at the end of the Renaissance. Many intellectuals began to write their works in national language and not in Latin. Marcilio Ficino wrote *Sopra l'amore*, a dialogue in platonic fashion, in Italian. The new trend at the dawn of modernity was to use national languages in all domains. In France, as an exemplary case, the change was very fast, from Descartes who has made the transition from Latin to French to Voltaire who had already begun to think of a comparative perspective between Shakespeare and Molière. The Bible was also translated into national languages and the need to talk about translation and interpretation was recognized.

Linguistic relativism was expressed as a clear hypothesis by German Romantic philosophers beginning with the end of the eighteenth century. They proposed the concept of *Volkgeist*, the idea that every national or ethnic group has some characteristics which are in a causal or a determinative relation with a spiritual moving force. J. G. Herder, one of the leaders of *Sturm und Drang* movement, published the book *Ursprung*

³ See *The Bible, Biblica*. The International Bible Society. New International Version, <https://www.biblica.com/bible/niv/genesis/11/>, accessed on March 5th, 2017.

der Sprache in which he argued that language shapes the frameworks in which each linguistic community thinks. Wilhelm von Humboldt asserted that language can't be reduced to a set of sounds and signs, but is even more than a view of the world, namely, the primary place where our thoughts are born. The diversity of languages became a reason that helped explain the diversity of nations and their identity in different forms, from cultural identity to political identity as a national state. This cultural approach led to the birth of hermeneutics as a new domain of research which continues nowadays. Translation is seen as an alteration through which original meanings are transformed under the impact of another cultural framework, yet there is no other way because human understanding is nothing but translation and presupposes a language: "inside or between languages, human communication equals translation. A study of translation is a study of language" (Steiner 1998, 49).

The other research tradition, formed more recently, at the beginning of last century, has its roots in cultural anthropology and linguistics and it gradually led to a cognitivist approach that has raised questions with epistemological relevance. The constitutive moment is represented by the so-called Sapir-Whorf hypothesis regarding the principle of linguistic relativism. In the "Preface" of his book *Language. An Introduction to the study of speech*, Sapir mentions that his main purpose is to offer an explanation regarding the variability of language in time and place and its relations to other fundamental human interests (see Sapir 1939, iii). According to Sapir, "Culture may be defined as *what* a society does and thinks. Language is a particular *how* of thought" (Sapir 1939, 233). He claims that it is difficult to find the causal relations between our cultural experience and the manner in which these are expressed by language as a social and historical product.⁴

⁴ Sapir explains this difference in terms of his own theory: "The drift of history, another way of saying history, is a complex series of changes in society's selected inventory – additions, losses, changes of emphasis and relation. The drift of language is not properly concerned with changes of content at all, merely with changes in formal expression. It is possible, in thought, to change every sound, word, and concrete concept of a language without changing its inner actuality in the least, just as one can pour into a fixed mold water or plaster or molten gold. If it can be shown that culture has an innate form, a series of contours, quite apart from subject-matter

Moreover, Sapir worked with the distinction between cultural content and linguistic form and he clearly express the idea that the structures of the two aren't isomorphic because language isn't the only determining factor. Therefore, if we conceive two communities which share a common language we have to accept the possibility that they can't share the same thought because other determinants are different. For example, if we take into account our perceptions of things from the external world, we'll discover that it is possible to have one word or many words for a perceived thing or even to have none. This means that we are free to propose different linguistic descriptions of the world. Language is just a condition, not the only determining causal factor of our descriptions of the world.

This weak form of linguistic relativity proposed by Sapir was challenged by a strong one based on the recognition of a determining relation between language and thought proposed by Benjamin Whorf. The strong version is based on the idea that the given structure of language constrains us to describe the world in a certain way and, as a result, it also shapes our thought and our cognition of the world. We usually describe the world in terms of substances and properties because the elementary structure of assertions, based on two elements, the subject and the predicate, determines our conception about the world:

We cut nature up, organize it, into concepts, and ascribe significances as we do, largely because we are parties to an agreement to organize it in this way – an agreement that holds throughout our speech and is codified in the patterns of our language. (Whorf 1956, 214)

It is important to note that Whorf has studied Native American languages and he was interested to reveal the differences between European languages

of any description whatsoever, we have a something in culture that may serve as a term of comparison with and possibly a means of relating it to language. But until such purely formal patterns of cultures are discovered and laid bare, we shall do well to hold the drifts of language and of culture to be non-comparable unrelated processes. From this it follows that all attempts to connect particular types of linguistic morphology with certain correlated stages of cultural development are vain." (Sapir, 1939, 233-234)

(the so-called Standard Average European – SAE) and indigenous language. He discovered that in Native American Languages there are many terms which correspond to a single term in SAE. The well-known example was that regarding more than twenty words for “snow” in Eskimo (Inuit) language (similarly in Sami language, in Scandinavia). Or, another example, in the case of drinking water, there are two different words in Hopi language, one for natural sparkling water, another for water which is put in a container. Moreover, Whorf discovered that in Hopi language there aren’t nouns for units of time (one day, two years and so on) because they treat times as a single process which can’t be cut in countable instances or sequences. One of the preferred examples was that of two languages which use different terms for colors. The conclusion was that if the two languages are so different then the members of the two linguistic communities will have difficulties to understand one another because the translation can’t be completed when the vocabulary and background linguistic structures place the speakers in different worlds.

My aim in this paper is to follow this second research tradition which was developed starting from the Sapir-Whorf hypothesis about linguistic relativism and to offer a reconstruction of the debate in an epistemological framework which takes into account the relation between language, translation, knowledge and truth. I think that such an approach should consider some of the main theories which contain at least some elements of linguistic relativism.

2. The place of linguistic relativism among relativisms

How shall we understand linguistic relativism in epistemological terms? How should we explain the truth of an assertion in relation with the language in a relativistic manner? A general taxonomy of different cognitive relativisms was proposed by Mandelbaum. He makes a distinction between subjective, objective⁵ and conceptual relativisms.

⁵ Mandelbaum takes the expression “objective relativism” from Arthur E. Murphy. See below.

Subjective relativism holds that “any assertion must be viewed in relation to the beliefs and attitudes of the particular individual making the assertion” (Mandelbaum 1982, 35). Epistemologically speaking, this means that the truth or falsity of an assertion is relative to the epistemic subject who made the assertion, to his/her subjective interests, attitudes and biases. As a consequence, “true” is replaced with “true for:” this means that things are for an epistemic subject just as they seem to be for that subject. Moreover, the possibility of disagreement is dissolved because the distinction between correct and incorrect judgments can’t be supported any longer since all judgments we believe in will be correct or true just for the reason that we believe them. In European thought, this tradition started with Protagoras’ doctrine of humans as measure of all things and continued to be mentioned more as a possible philosophical standpoint. Many forms of subjectivism, and I include methodological relativism among them, don’t include the relativist thesis. It is obvious that linguistic relativism can’t be identified as subjective relativism given that language is intersubjective by its very nature.

Objective relativism of a knowledge relation is based on the principle that there is always a personal reason for any assertion which was made, or that the person who has made the assertion occupies a particular position in that epistemic situation, or that any assertion is able to refer only to some of the aspects of the object with which it is concerned. Consequently, the truth of an assertion is dependent on the context in which the assertion was made. Unlike the subjective relativist, the objective one

would deny that what is taken to be true or false is primarily a function of the beliefs and attitudes of the particular person making the assertion; rather, it is relative to the nature of the total context in which the assertion is made. (*ibidem*)

The difference between the two relativisms is that in the first case our knowledge becomes personal knowledge and is subjective because is relative to our own system of beliefs, whereas in the second case our knowledge remains objective because the components of the context

which produce our beliefs are independent from the purposes of the epistemic subject.

The third form, conceptual relativism, holds that our assertions have to be interpreted with reference to the cultural context in which they are embedded, namely, not in their relations with the objects, as in the case of objective relativism, but in their dependence on the intellectual or conceptual background. Wittgenstein's later work, Whorf's linguistic relativism, Kuhn's theory about knowledge based on paradigms, and Rorty's idea of contingent vocabularies are usually considered good samples of conceptual relativism. This enumeration already contains a theoretical place for linguistic relativism as a case of conceptual relativism.

In Mandelbaum's view, conceptual relativism is culture bound and there is no way to ground it in data which aren't culture bounded. Therefore, someone who supports conceptual relativism makes claims which contain the so-called "self-excepting fallacy," namely "the fallacy of stating a generalization that purports to hold of all persons but which, inconsistently, is not then applied to oneself."⁶

Let us return to objective relativism as a preliminary step to a better understanding of linguistic relativism.

The term "objective relativism" was proposed by Arthur E. Murphy in his article "Objective Relativism in Dewey and Whitehead" (Murphy 1927). The two philosophers suggest that events and relationships, and not objects, are the ultimate constituents of what there is. What are the epistemological consequences of this position?

Mandelbaum mentions some difficulties of objective relativism under three headings:

first, with respect to the role of interest or purpose in judgments concerning matters of fact; second, with respect to the influence of the standpoint of the observer on the judgments he makes; and, third, the consequences which follow from the fact that any judgment is selective, dealing only with particular features or aspects of the object or situation judged. (Mandelbaum 1982, 35)

⁶ Mandelbaum (1982, 35). See also Mandelbaum (1962).

We are easily ready to accept that every judgment about facts stands in a relation to a situation in which the epistemic subject making the judgment is bound to some interests or purposes which relate to the content of that judgment. But such interests and purposes can interfere in two ways. On the one hand, the epistemic subject is instrumentally interested about an object of knowledge because he would like to bring about or to avoid a state of affairs. On the other hand, the epistemic subject may be interested in an object for the simple reason that it interests him. In this case, the purpose is to explore, to understand or to explain the state of affairs or the object. The two ways aren't mutually exclusive. The objective relativist claims that any epistemic subject who knows something is trapped at least in one of the two situations, and there is no escape from them.

In an instrumental or pragmatic way we can introduce the same approach based on the idea that a standpoint is the framework in which the subject develop its organic functions (just as the Darwinian theory of evolution asserts) and its mental capacities (as in Dewey's theory of education). A standpoint can be conceived in two ways, temporally or spatially. Those objective relativists who are concerned with historical knowledge put the relativity of our judgments in connection with the moment when they were made. Those objective relativists who are concerned with sense perception lay emphasis on the fact that different observers look at the same object from different points of view. Certainly there isn't a sharp line between the two ways.

In the case of historical knowledge, relativists can use two ways in order to understand the influence of temporal factors on historical judgments of the past. Each way depends on selections and interpretations; this is something proper to the writing of history. The first (and less radical) is the claim that selection and interpretation are made starting from present interests. This means that we focus our attention on some events and they are seen as continuous with the present. If the present changes, the interpretations of the past will change also. The second and radical claim is that "the past itself undergoes significant change through later developments" (Mandelbaum 1982, 40).

If we turn to the hard core of this kind of relativistic argument springing from Whorf's hypothesis, then I think that we can conclude

that the way the world appears to be to the user of a language depends on the implicit metaphysics of that language. In Whorf's terms, this means that:

The background linguistic system (in other words, the grammar) of each language is not merely a reproducing instrument for voicing ideas but rather is itself the shaper of ideas, the program and guide for the individual's mental activity, for his analysis of impressions, for his synthesis of his mental stock in trade. (...)

The categories and types that we isolate from the world of phenomena we do not find there because they stare every observer in the face; on the contrary, the world is presented as a kaleidoscopic flux of impressions which has to be organized by our minds – and this means largely by the linguistic system of our minds. (Whorf 1956, 212-213)

3. The “linguistic turn” of cognitive relativism

Would this mean that the worlds in which we live will be more or less similar depending on the language we speak or will we be able to understand each other through communication and translation? We agree that even if languages are so different from each other and our minds are determined or structured by them, we can understand each other as members of different linguistic communities. We can express the same ideas in different languages and we can cooperate even if we can't speak the same language. We can translate one language into another and obtain the same practical effects. Moreover, members of a scientific research community communicate with each other and share a common vocabulary of the scientific discipline in which they work. But what is the relationship between the research community and the linguistic background on which the social network operates? Does language influence community structuring?

I will further argue that Thomas Kuhn's relativism is the best example of such an approach. He starts his *The Copernican Revolution* with some remarks about the observations made by two astronomers, a Ptolemaic one and a Copernican one, to notice the differences between them and their relativistic cognitive commitments. When the two

astronomers look at a sunrise, they will see different things just because the Ptolemaic believes that the Sun moves around the Earth while the Copernican believes that the Earth moves around the Sun. They will use the same statements about what they perceive, but their observations are influenced by the previously mentioned beliefs, so that they will think that they are talking about different things.

This idea was developed by Kuhn in *The Structure of Scientific Revolution*, where he used the concept of paradigm in order to explain the differences between two scientific communities which share different views about the world, and the concept of incommensurability in order to explain the relations between two different paradigms.

Let's consider two examples. The first is about Lavoisier and Priestley, and their attempt to explain the phenomenon of burning. Lavoisier saw oxygen and talked about "oxygen" where Priestley saw and talked about "dephlogisticated air." As a consequence, the two scientists saw different things and their descriptions of the world differed. The second example concerns Newton and Einstein. In their case, the word "mass" as it is used by Newton cannot be translated by "mass" as it is used by Einstein. Although the words are the same in material mode, as succession of letters, their meanings are different. Therefore, because the meaning of a word is given in a holistic mode, the two scientists work with different concepts of mass. The two concepts are incommensurable and a translation can't be made.

Kuhn explained and developed his ideas in "Second Thoughts on Paradigms" (1977), then, in "The Road since *Structure*" (2000) proposed that incommensurability has to be understood in terms of differences between taxonomies which are used as classificatory schemes. Moreover, Kuhn introduces the new concept of lexical network. We learn a language by learning some words in an ostensive manner. Let's imagine a small child on a walk with his father in a zoological garden. The child previously learned to recognize and to discriminate some species of birds, but that day he will learn to identify swans, geese, and ducks. Ostension is the best tool to learn something in these circumstances, because phrases like "all swans are white" may play a role, but we have no guarantee that they suffice for identification. The father sees a bird, points to it, saying "Look, there's a swan!" "A swan!",

the boy repeats, and adds something new in his taxonomy as a network of relations between words and objects. They continue their work and in a short time the child points to a bird, saying "Daddy, there's another swan!" But he hasn't yet learned what swans are and he must be corrected by his father: "No, that's isn't a swan, that's a goose". The next identification of a swan will be correct, but the next bird identified as a "goose" is, in fact, a duck. The child develops his lexical network adding the new word and better understanding the differences between a real swan, goose and duck. After a few such encounters and other corrections, the child will acquire the ability to identify these different species of birds.

This is just one of the possible stories about how is possible to learn a language by ostension and how to use a taxonomy. Different persons can learn a language in different ways and they can use different lexicons. For example, let's suppose that someone has only the word "bird" in their vocabulary. If they will correctly use this word for descriptions of their perceptual experiences when they sees a swan, a goose or a duck, the effect will be that they will use the language efficiently all the while avoiding some practical troubles caused by the poverty of their language. This means that our personal taxonomy could be different because we associate different cognitive and non-cognitive meanings to the words we use. For example, we associate some emotions and feelings to the word "earthquake" if we have experienced this kind of natural phenomenon. The problems that arise concern a) the possibility of different people being engaged in communication given that the taxonomies they use only partly overlap, and b) the possibility of translating one language or vocabulary into another.

A relatively similar "linguistic turn" was developed by Quine (1960) in his theory about the indeterminacy of translation, in relation to a behaviorist theory of meaning. Quine proposed the so-called *gavagai* thought-experiment. Let's suppose that a linguist tries to find out the meaning of "gavagai," an expression used by a native speaker of an unknown language. The ostensive way is the best to capture the meaning: if the speaker points to a rabbit when he utters the term "gavagai," we'll conclude that this is its meaning and that the word "gavagai" has to be translated by "rabbit." But it is a mistake to think that we always have the capacity to compare a foreign language with

our own and that the background language and referential devices help us do this. It is easy to imagine that when the native speaker utters the word "gavagai" pointing to a real rabbit he refers to something else, for example, to undetached rabbit-parts, to a young rabbit or to rabbit-tropes. The conclusion is that it is better to work with several translation hypotheses even if the sensory stimuli and the behavioral data are the same for speakers of two different languages. There are many ways to make a translation fit the behavior of the speaker.

The difference between Quine's theory about the indeterminacy of translation and Kuhn's theory about different lexicons is that the first talks about the possibility of multiple partial translations, the second gives strong reasons for the impossibility to translate a language into another. So how do we explain the possibility that two speakers of different languages can understand each other? How do they overcome the inconveniences of translation?

4. Translation and interpretation

Let's start by going back to the distinction made by Frege between sense and reference.⁷ We can easily understand it with the help of an example. When we talk about the planet Venus we can use two alternative expressions, "Morning Star" and "Evening Star." The meanings or senses of the two expressions are different, but their reference is the same, the planet Venus. Therefore, we can speak about the same thing using different expressions which refer to that thing in different senses.

Therefore, in a translation it is important to preserve not only the reference, but also the sense. It is clear that in the case of the words which are rigid designators we can do this easily. The reference of the word "water" will be H₂O in all the native languages spoken on Earth and we can talk about it using different descriptions, such as "the most important liquid on the Earth, the liquid which covers the Earth" and so on.

⁷ See Frege (1949) for the English version. The original German version, "Über Sinn und Bedeutung" was published in the year 1892.

Generally speaking, we can use different descriptions for the same reference. For example, for the city of Luxemburg I can use different descriptions such as, "the capital of Luxemburg, the city I am currently visiting, one of the capitals of the European Union together with Brussels and Strasbourg, the seat of the European Court for Justice." The meanings of all these descriptions are different even if they have the same reference. But it is important to preserve the same sense if we want to preserve the initial thought or intention of the speaker. This request is very strict if we don't want to change the meaning through translation.

If we take into account again the ideal project of translation in a pure form as radical translation we'll assert that the difficult task for the translator is to preserve exactly the initial meaning of the words and sentences. The ideal task of the translator is to perfectly translate a sentence from one language into another sentence from a different language without any change in meaning. But, as I have said above, this task depends on the translator's prior linguistic knowledge.

Indeed, we can identify, with Quine, a case in which translation of a language is possible without any prior linguistic knowledge and solely on the basis of the observational knowledge (the observed behavior of the speaker and our acquaintance with the perceptual stimulations that give rise to that behavior). All the members of a linguistic community will be able to understand each other when they speak about basic perceptual stimulations. For example, when we see something, a tree, all of us agree that the word which have to be used is the word "tree". We suppose that all languages are basically learned starting from basic stimulations. Therefore, it will be easy to translate from one language to another if we speak about this kind of perceptual stimulations. But, if we remember the above example of seeing a rabbit, the so-called *gavagai* mental experiment, we have to admit that things are not so simple and safe even if these conditions of a radical translation are met.

This puzzle is a serious reason to look for an alternative. Davidson (1984) proposed a broader conception of the behavioral evidence available to a speaker/ translator/ interpreter, and he rejects Quine's idea about the special role of perceptual stimulations. He introduces the concept of "understanding a language" and claims that a theory of translation is necessary but not sufficient to ensure the understanding of

the translated language. For example, this is the case of a translation into a language which isn't understood by the speaker. As a result, the notion of "translation" is replaced with the notion of "interpretation".

"Radical interpretation" implies an interpretation of the linguistic behavior without any support from a speaker's prior knowledge. The domain of this prior knowledge should be comprised of a speaker's different beliefs about the world and the ways in which it is structured by a conceptual schema, as well as the accepted meanings of the speaker's different utterances. As a consequence, we can speak of which meanings are assigned to the speaker's utterances if and only if we have sufficient knowledge of what the speaker believes, and we can grasp these beliefs if and only if we know what the speaker's utterances mean. Is there a way out of this mess? Davidson solved the problem by stating the principle of charity, according to which we usually work with the presupposition that all the speakers of a different language are rational, they want to communicate with each other and their intention is to tell the truth.

For example, let's suppose that we travel to a country where a language entirely different from ours is spoken. Let's suppose that we are in Japan, in Sapporo City, in winter times. When a Japanese will take some snow in his hands and he will say a word looking at it, we'll suppose, according to the principle of charity, that all he wants to do is to give us the linguistic equivalent for the word "snow". We don't have any reasons to suppose anything different regarding his behavior.

But is such a behaviorist approach complete from an explanatory point of view? Or do certain mentalistic components remain, at least in terms of understanding, unexplained? Let's focus on the problem regarding the relation between a translation from one language to another and the capacity to understand this process as a mental activity. I'll adapt the so-called "Chinese Room" thought-experiment proposed by Searle (1980) to the case of translation. Some philosophers and scientists think that, in the future, artificial intelligence will be able to translate more accurately from one language to another. Let's suppose that we construct a computer that takes Chinese characters as input and, running the computer programme, it produces other characters in another language as outputs, say in Romanian. Are we warranted in thinking that the computer is a Chinese and Romanian speaker? Does

the computer literally understand Chinese or Romanian? Or does it merely stimulate the ability to understand Chinese and Romanian? Searle's argument is that, without understanding (or, in Searle's terms, intentionality), we can't say that the machine thinks. If we take into account the case of multiple translations we'll say that is obvious that the computer isn't a human translator (or a human mind) because it can provide only statistics of uses and not a certitude based on feelings, as in the case of the humans. We'll conclude that, in Searle's terms, any translation is epistemically subjective, and that it is always related to interpretation and in need of understanding.

5. Conclusion: the languages we speak, the worlds we live in...

But what are the consequences of such an approach that draws the contours of an inevitable relativism? To what extent does speaking different languages mean thinking differently about the world or living in different worlds? Davidson developed an idea that associates having a language with having a conceptual scheme. The relation is a very simple one: if conceptual schemes differ, so do languages, but if the languages differ, this does not mean that the conceptual schemes are also necessarily different:

speakers of different languages may share a conceptual scheme provided there is a way of translating one language into the other. Studying the criteria of translation is therefore a way of focusing on criteria of identity for conceptual schemes. (Davidson 1974, 6)

Therefore, we have to consider the possibility that more than one language may express the same scheme and this means that these languages are intertranslatable. But is the relation of translatability transitive? Davidson's answer is that some language, say Saturnian, may be translatable into English and that some further language, like Plutonian, may be translatable into Saturnian, while Plutonian is not translatable into English. Corresponding to this distant language would be a system of concepts altogether alien to us.

Therefore, we return to the basic idea that the two worlds are different because we talk about them in different languages which aren't translatable. We explained this incommensurability, according to Kuhn's view, with the help of his theory about paradigms (or traditions). But what might it mean to live, due to one's differing paradigms or traditions, in different worlds? Following Kuhn's theory, we'll agree that it is possible to imagine that there is only one world, our own, that is described from different points of view with the help of different languages. Strawson proposed a purely epistemic alternative to the kuhnian linguistic and ontological approach. He claimed that "It is possible to imagine kinds of worlds very different from the world as we know it" (Strawson 1966, 15). But is it possible to imagine that these different worlds are described from the same standpoint with the help of the same language?

The case of natural kinds or essences seems to be the most challenging in terms of the relationship between worlds and languages. Here is possible to only sketch what the bearing of a different thought-experiment might be. The thought-experiment of the so-called "Twin Earth", proposed by Putnam (1973), helps us to understand this puzzling problem. Let's suppose that there are two identical Earths. The difference in the case of water is that on Twin Earth "water" refers to something that has the same perceptual properties but isn't H₂O, having a different chemical structure, XYZ. Putnam proposed two theories about the meaning of the word "water":

1. "Water" has a meaning relative to the world, but constant. Therefore, "water" means the same thing in World 1 and World 2, but it is H₂O in World 1 and XYZ in World 2.
2. "Water" is H₂O in all possible worlds. Therefore, in World 2 we use the same word for a thing with the same properties, but the word "water" has in World 2 another meaning.

Putnam prefers the second theory and asserts that the word "water" is a rigid designator and it denotes the liquid H₂O in all possible worlds.

If we prefer the same second choice then we have only one step back to return to Davidson's proposal to define meaning on the basis of conditions of truth. This is a robust solution at least when it comes to

facts. Undoubtedly, the Romanian translation for "Snow is white" will be "Zăpada este albă".

REFERENCES

- Davidson, Donald. "On the Very Idea of a Conceptual Scheme." *Proceedings and Addresses of the American Philosophical Association* 47 (1974): 5-20.
- Davidson, Donald. *Inquiries into Truth and Interpretation*. New York: Oxford University Press, 1984.
- Frege, Gottlob. "Sense and Reference." *Philosophical Review* 57, no. 3 (1948): 209-236. Translated by Max Black.
- Kuhn, Thomas S. *The Copernican Revolution. Planetary Astronomy in the Development of Western Thought*. Cambridge, Mass.: Harvard University Press, 1957.
- Kuhn, Thomas S. *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press, 2nd edition, 1970.
- Kuhn, Thomas S. *The Essential Tension. Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press, 1977.
- Kuhn, Thomas S. *The Road since Structure: Philosophical Essays, 1970-1993, With an Autobiographical Interview*. Edited by James Conant and John Haugeland. Chicago: University of Chicago Press, 2000.
- Mandelbaum, Maurice. "Some Instances of the Self-Excepting Fallacy." *Psychologische Beiträge* 6 (1962): 383-386.
- Mandelbaum, Maurice. "Subjective, Objective, and Conceptual Relativism". In *Relativism. Cognitive and Moral*. Edited by Jack W. Meiland and Michael Krausz. Notre Dame & London: University of Notre Dame Press, 1982. 34-61.
- Murphy, Arthur E. "Objective Relativism in Dewey and Whitehead." *Philosophical Review* 36 (1927): 121-144.
- Putnam, Hilary. "Meaning and Reference." *Journal of Philosophy* 70 (1973): 699-711.
- Quine, Willard Van Orman. *Word and Object*. Cambridge Mass.: MIT Press, 1960.

- Sapir, Edward. *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace and Co., 1939.
- Searle, John. "Mind, Brains, and Programs." *Behavioral and Brain Sciences* 3 (1980): 417-457.
- Steiner, George. *After Babel. Aspects of Language and Translation*. Oxford: Oxford University Press, 1998.
- Strawson, Peter F.. *The Bounds of Sense*. London: Methuen, 1966.
- Whorf, Benjamin. *Language, Thought, and Reality. Selected Writings of Benjamin Lee Whorf*. Edited by John B. Carroll. Cambridge, Mass.: MIT Press, 1956.

NATURALISM RELOADED: HOW DO WE CONSTRUCT OUR WORLD?¹

TUDOR MĂRGINEAN²

Review of Joseph Rouse, *Articulating the World. Conceptual Understanding and Scientific Image*, University of Chicago Press, 2015, 416 pp.³

In the second half of the 20th century, naturalism became one of the main points of view embraced by philosophers in multiple fields, from metaphysics to philosophy of mathematics and philosophy of science. In philosophy of science, naturalism was often associated with the endeavor to scientifically explain our capacities for doing science and, at the same time, the epistemic normativity involved in sciences.

Joseph Rouse's book makes a step further by trying to eliminate the last remaining bastions of a transcendental, metaphysical or theological point of view regarding conceptual normativity. His book is hardly the only comprehensive attempt to articulate a naturalistic image of the world. He comes from a tradition which can be said to have started by Sellars's distinction between the manifest and the scientific image, which gave birth to different attempts to explain how the

¹ This review originally appeared on the "Let's talk about books" academic blog. It was accessed here: <https://letstalkaboutbooks.blog/2021/01/19/naturalism-reloaded-how-do-we-construct-our-world/> The editors thank both the author and the coordinator of the blog, professor Dana Jalobeanu from the University of Bucharest, for agreeing to reprint the review.

² Tudor Mărginean is a doctoral student within the Faculty of Philosophy at the University of Bucharest. Contact: <tudorov1997@gmail.com>

³ Unless otherwise noted, the page references below are to the book being reviewed.

scientific image can be reliably constructed as an objective way of describing the world, including us as epistemic subjects, starting from our standpoint as biological and historical beings. At the same time, naturalism received critiques from empiricist philosophers like Bas van Fraassen, but also from the so called "left-Sellarsians" like McDowell, Brandom, or Haugenland, who could not make sense of the conceptual normativity of the sciences (or, more accurately, could not make sense of how scientific authority can be justified and how humans, as rational subjects, could be held accountable to such norms). In order to make easier for the reader the understanding of the most important issues addressed in this book, let's expose a bit the meanings of "manifest image" and "scientific image". The first refers to us, as humans, as biological organisms constrained by our cognitive apparatus and biological purposes and as members of a society, embedded in social interrelations, and our ways of making sense of the world through knowledge and skillful manipulations of objects. The second one, the "scientific image", refers to the picture resulting from an accurate scientific description of the world (including ourselves, as subjects of knowledge), let's say, from a "God's point of view". The way this distinction is articulated is one of the main issues of this book.

Articulating the world. Conceptual understanding and the scientific image is divided in two parts, complementary to each other and proceeding from opposite directions. The first part proceeds from the scientific image and tries to explain our development as language-using rational beings capable of conceptual understanding, and the second part goes on from our standpoint as knowers and tries to explain the normativity involved in sciences. These two parts are complementary to each other, which may seem to be one of the faults of this book. One who wishes to attack naturalism as articulated by Rouse may reason in the following way: if in order to explain our status as beings capable of conceptual understanding we need to appeal to evolutionary biology and anthropology (which belong to the "scientific image"), the explicative power and normativity of which are to be explained starting from our capacities of conceptual understanding, then the entire project is stuck in circularity. While Rouse doesn't address this problem directly, and I think this is one of the reasons for us not being

constrained to accept his conclusions, I think he manages to show how this circularity is not as vicious as it may seem to be. Each of the two complementary parts makes sense on its own, and is consistent with our practices. Furthermore, circularity is not always avoidable: often, when we are trying to define a term, we are using other terms which are themselves definable through the *definiens*. What is important in this case, I think, is to avoid our conceptual construction to be a "frictionless spinning in the void", as Rouse often quotes McDowell while criticizing various attempts to account or scientific normativity without grounding it in the material world and actual practices.

One of the first questions Rouse tries to answer concerns how intentionality is coupled with conceptual normativity in humans. There are four main types of attempts to deal with this problem. One of them is encountered in authors like Husserl's structures of consciousness and Carnap's logical structure of language, who saw intentionality as being an operative process regarding ungiven or nonexistent objects. The second one takes intentionality to be operative, but about given objects. Here we have Dennett, Millikan or Dreyfus. The third one takes it to be rather normative, but regarding nonexistent objects (here we have Rorty of Davidson as representatives, with their views of linguistic meanings and normative rules as not being accountable to the world). Finally, the fourth one, that Rouse defends, views intentionality as being normative and about given objects (more precisely, given to intentionality, as anticipated or foreseeable in the future).

To understand the difference between operative and normative accounts for intentionality, take one of the most used examples in the book: chess games. A chess player knows the rules of the game and also which moves are better and give more advantages in the game. Ordinary players usually have to actually think about these rules and principles of the game while playing, but a grandmaster makes many of the moves automatically, without actually thinking about rules and principles. If we take intentionality as operative, then in many cases our actions fail to be able to be taken as accountable to rules and principles,⁴

⁴ Rouse, Joseph. *Articulating the world. Conceptual understanding and the scientific image*. Chicago: University of Chicago Press, 2015 (47-48, 63-64).

even if they were not actually propositionally formulated in the players' minds. In this second sense, conceptual normativity applies to actions as well as it applies to rules and principles.

Rouse spends much of the first part of the book attacking the second and the third accounts of intentionality. I will not enter into every detail regarding how he establishes his point of view as the most accurate one, but some critical steps should be highlighted.

The question that arises, Rouse says, is how our kind of normativity is constituted and how we differ from other animals from which we evolved. To be normatively constrained means to be able to make mistakes or to be wrong about something. That's why objects are not normatively constrained: they cannot make mistakes (they can only make mistakes as *our* instruments, regarding *our* goals). Do other animals make mistakes? Are they wrong about things? The example Rouse uses the most is taken from Haugeland:⁵ imagine a bird which avoids catching only yellow butterflies, which happen to be poisonous. If, for example, there is one species of yellow butterflies that are not poisonous, but the bird still refuses to hunt it, can it be said that the bird is wrong about that? The answer is no, because to be wrong about something means to be able to take it *as* something. The bird doesn't take the non-poisonous yellow butterfly as poisonous simply because it avoids yellow butterflies only due to a visual cue, with no knowledge about "poisonousness". The conclusion from this step is that, in order to be normatively accountable, something has to take things *as* being in some ways.

The other extreme is that of being able to take things *as* being somehow but not being able to hold them accountable to objects. That may be the case with Davidson's or Rorty's account of societies formulating rules as "frictionless spinning in the void": without being grounded in objects, the rules cannot have normative power, because in such a case we don't know when a rule is followed or not and also we don't seem to be *normatively* constrained by them (Rouse 2015, 68-69).

That being said, the first part ends with sketching a view which does justice both to how we take things *as* being somehow and to the objects themselves. Rouse (2015, 82-83) makes a distinction between

⁵ *Ibidem.*

what is at stake and what is at issue within a practice. What is at issue refers to how that practice is continued when some obstacle or problem arises, and what is at stake refers to what it means for the issue to be solved in some way (or, in other words, what is at stake refers to the larger goal of a practice). If we take an organism as being analogous to a practice, we can say it has the goal of maintaining its existence and its way of life in an environment, and in this regard it can be successful or not, but it can't be held accountable to norms regarding "mistakes" it cannot make: the bird which doesn't catch yellow butterflies might have had a more developed apparatus allowing it to distinguish between different yellow butterflies, but that apparatus might have been too costly in other ways with respect to survival.

The other main problem of the first part is that we seem constrained by a dual normativity: both by what is at stake and what is at issue. The trait that seems to do the job is language. Not only language, of course: using equipment, dancing, painting or singing are manifestations of conceptual understanding too. It can be argued, though, that the acquisition of language was the crucial step, and, it seems, a very difficult one. Rouse (2015, 91) gives the example of a bonobo, Kanzi, who could understand and even compute expressions remarkably well, which is a "proof" of the fact that maybe the brain was "ready" for language acquisition in our ancestors, but who could not use those expressions to communicate anything other than things connected to their immediate surroundings. The formation of the capacity for "symbolic displacement", that is, for the ability to use clues (gestures, sounds, graphic symbols) in order to express something disconnected from the surroundings (for example when I say "I found some source of fresh water) is very unlikely. That's why biologists are talking about a cognitive trade-off: we had to "give up" some capacities in order to be able to use symbols to communicate abstract information. Symbolic displacement is, in most cases, something very costly in terms of survival, unfavorable and counter-selective, because it makes immediate responses to the surroundings more difficult.

Our species most probably acquired symbolic displacement once our ancestors had to work in groups in order to avoid predators and find sources of food, after they left the forests for the savannah. You can find a more comprehensive explanation in the third chapter of the book.

The second half of the book, as I said earlier, proceeds from our standpoint as beings capable of conceptual understanding in order to show how we are constructing the scientific image. Well, it may be improper to talk about a scientific "image". A scientific image presupposes a unitary picture and, moreover, it seems to presuppose a representationist schema of science as a set of propositions. Rouse questions all of these assumptions. There are a few things which must be said in order to sketch Rouse's characterization of science. Firstly, science is not retrospective, as philosophers of science often describe it, but prospective: it doesn't consist in a set of sentences already established which form *the* body of scientific knowledge. Rather, previous scientific discoveries and established knowledge stand for future discoveries and are so understood by practitioners. The relevant scientific facts are those which allow for the discovery⁶ of new facts. Secondly, scientific practice matters: as conceptual understanding is normative, not operative, it is involved in all sorts of actions, and is not a property of mental activity only. Skills can succeed or fail in being in accordance with conceptual norms.

An important problem which needed to be solved is that of the applicability of scientific models. For other philosophers of science as Ian Hacking or Nancy Cartwright, scientific models or scientific laws apply only in very specifically determined cases. As Ian Hacking observes,⁷ phenomena which are studied by scientists do not exist in nature as such, but must be created in the laboratory. If the theoretic model is constructed in order to explain or describe the phenomena, then there is a sort of fitting between them such that the model does not apply outside the range of phenomena which were especially designed for the model. Or, according to Cartwright, it applies to other phenomena only if they are in accordance to the model constructed for the laboratory-

⁶ The term "discovery" may be problematic if it is seen as a commitment to scientific realism. Rouse is neither a realist nor an anti-realist in the classical sense of these terms. He doesn't presuppose that there is a set of facts ready to be discovered, because the facts depend not only on "the world" but on our interests and our practices too: science goes on some path depending on many factors (that is, what is "at stake"), including what is "interesting" or "important" for the practitioners.

⁷ Hacking, Ian. *Representing and Intervening*. Cambridge: Cambridge University Press, 1983.

created phenomena.⁸ But in this case, as Rouse says, the model (or concept, or law) applies only when it applies, which is a tautology. Of course, tautologies are hardly interesting when it comes to describing how science works.

According to Rouse, a scientific concept applies, let's say, inductively. It is designed from the start to apply in various cases, but it comes down to our skillful manipulation of experimental equipment to decide whether a new phenomenon can be modelled through a concept or not. In other words, concepts are articulated in such a way as to allow further articulation through observation, experiment and theoretical work. Consequently, experimental skills and theoretical modelling are mutually accountable: scientific practices are theoretically driven and are held accountable to norms prescribed by concepts, while concepts are accountable to nature, such that every theoretic model is defeasible. Of course, theoretic models or theories are revisable and resilient at the same time, such that the further acceptance or rejection of a theory after recalcitrant phenomena are observed depends of a holistic schema and depends of what is seen as being at issue and at stake in a science.

That being said, what remains to do is to explain how science is constituted and what kind of patterns in nature are tracked by scientists. As anyone can observe, there is not a single science, but many sciences, and some of them, as most branches of physics, are called "fundamental", while others are called "special sciences". Usually what makes a difference is the supposed fact that fundamental sciences have laws, while special sciences exhibit only regularities (even if they are very strong regularities). This view was challenged in various ways: some philosophers, like Cartwright, attacked the concept of law, while others tried to show that even special sciences have some kind of laws. Of course, laws can be understood in many ways, from principles governing the nature to counterfactual invariance. Rouse, following Lange, adopts the latter view. Laws are described as counterfactual invariance, that is, facts which would remain constant if other facts were changed. Of course, this definition is not sufficient, because there may be contingent regularities which keep their constancy across possible worlds to a

⁸ Cartwright, Nancy. *The Dappled World*. Oxford: Oxford University Press, 1999.

greater extent than some laws. Since a detailed account of how to define laws to answer these issues would depart from the purposes of my review, I advise the interested reader to check Chapter 8 of the book. The main idea Rouse wants to propose is that every science has its "relevant context", and what is invariant in a science, once some facts are accepted, constitutes a law in that particular science. For example it would make no sense to say that had we evolved in another way, such and such facts about our bodies would have been different too. If we accept some facts about our evolution and the constitution of our bodies, then we can find the invariance required so as to be able to talk about laws.⁹

The last thing about the second part of the book that I want to highlight is that sciences cannot be otherwise but subjective in a specific sense. To talk about a "scientific image", according to Rouse, is a mistake because there is not such a global unity within sciences: every scientific domain is created by constructing theoretical concepts which can be further applied and held accountable to nature such as to produce interesting knowledge. The comprehensibility of sciences is limited by our context as biological entities on Earth with specific needs. That doesn't mean that Rouse is an advocate of the disunity of sciences either: there is always the possibility of creating new sciences at the boundaries of already existing sciences, using concepts from both.

After exposing the main claims and arguments of the book we may ask what constitutes epistemic normativity in science. Why should we believe what physicists are saying about such and such phenomena? If we were to accept the conclusions of the book, the answer would be that scientific practice and knowledge are not just a product of our way of life, but they are producing it by changing our environment and interrelations within our societies. Our practices are bound together, and science has no special status in this respect. It has a special status because it is held accountable to nature, and for our practices to continue (this is what is at stake) we should decide what is for science to continue

⁹ Here it might be said that in special sciences such as biology the proportion of "noise" across the supposed regularities is greater than in the case of physics. And, of course, what counts as "noise" in developmental biology or physiology is regular fact in evolutionary biology, which studies variation, while ordinary regularities in physiology are less interesting.

as a practice. And, as I already said, for sciences to continue is for us to accept results which are produced in specific ways as accountable to nature. I find the entire construction strong, without a little (or maybe not so little) exception: Rouse did not exert himself enough with respect to the normativity of logic and mathematics. If they are just greater counterfactual invariances, as Lange and Rouse seem to suggest, this needs a justification, because of their supposed apodictic character (if there is no apodictic necessity in logic and mathematics, and they are inductive and revisable instead, that has to be shown too).

Another complaint we might have is that Rouse accounts for a kind of normativity restricted to epistemic contexts, although he did not limit his pretensions explicitly. He rejected normativity as it was conceived by philosophers like Rorty, as being derived from socially accepted rules, because socially accepted rules do not bind us not to violate them. If we were to accept Rouse's account, then normativity binds us because we are engaged in some practices, and in order for a practice to continue we might make norms which are to be respected. This account, in my opinion, is a good justification for an instrumental normativity, or reducible to, as Kant would put it, an hypothetical imperative. Even in this last case, his account doesn't seem to be much stronger than Rorty's. Rouse doesn't claim that he limits himself to account just for that kind of normativity, but also he doesn't account successfully for stronger versions, such as moral normativity.

As a conclusion, despite the fact that some issues remain unsolved here and there, the entire project is well articulated, comprehensive, scientifically informed and strongly defended argumentatively. In my opinion, even though I have my personal reserves with respect to naturalism, the fact that Rouse pays a special attention to scientific practices and biological evolution makes his project to be the starting point of a promising path for successfully defending a naturalistic image of the world.

REFERENCES

- Cartwright, Nancy. *The Dappled World*. Oxford: Oxford University Press, 1999.

Hacking, Ian. *Representing and Intervening*. Cambridge: Cambridge University Press, 1983.

Hacking, Ian. *Scientific Reason*. Taipeh: National Taiwan University Press, 2009.

Rouse, Joseph. *Articulating the world. Conceptual understanding and the scientific image*. Chicago: University of Chicago Press, 2015.

Revista Română de Filosofie Analitică

Vol. XII

Iulie-Decembrie 2018

Nr. 2

CONTENTS

RRFA Vol. XII, 2018, Nr. 1 5

*

OCTAVIAN PURIC, *Ways of Seeming in Plato* 65

BOGDAN DUMITRESCU, *Freedom vs. Eternalism: Some Objections to the Inside
Out Perspective* 103

DAIAN BICA, *The Last Nail in the Coffin of Scientific Theories – Review of Steven
French, *There Are No Such Things as Theories* (Oxford: Oxford University
Press, 2019, 288 pp.)* 137

WAYS OF SEEMING IN PLATO

OCTAVIAN PURIC¹

Abstract: The purpose of this article is to present what I take to be the two main senses of seeming that we can find at play throughout Plato's work. These are what I have called the ontological sense and the genealogical sense.

I begin by introducing Plato's model-image metaphor. The model image relation will provide me with the elements necessary to illustrate the two main ways of seeming. I distinguish two senses in which we can read the image metaphor based on two types of objects the metaphor can refer to. When I want to refer to the particulars and their images, I will use the term "literal relata". When I refer to the Forms and the particulars I will use "metaphorical relata". I will call the models from the literal relata — the particulars — by the name of "relative models", while reserving the unqualified term of "model" to the models of the metaphorical relation, i.e., the Forms. Afterwards, I argue that there are two main types of seeming throughout Plato's work, the ontological seeming and the genealogical seeming.

On the one hand, I define ontological seeming as investing, either tacitly or explicitly, that which is ontologically an image with the role and function proper to the real model. Genealogical seeming, on the other hand, presupposes a difference between model and image, and consists in incorrectly identifying an image as being of a model rather than another.

I will maintain that identifying an image as being of a model is the basis on which Plato understands predication. I will further divide both types of seeming. I will call both ontological and genealogical seeming "perspectival" whenever their objects are particulars and when error is due to perspectival causes. With regards to ontological seeming, I will call it "radical" whenever it invests a relative model with the function of the true model. I shall call genealogical seeming "heuristic" whenever improper images of models are used to instill in someone an improper model.

Radical genealogical seeming, on the other hand, will consist in the application of unsound models in identifying particulars. Finally, I will argue that the radical ontological

¹ Octavian Puric is a doctoral student within the Faculty of Philosophy at the University of Bucharest. Contact at: octaviandan91@gmail.com

seeming and the genealogic heuristic seeming can be dispelled only by the use of dialectics, and thus constitute the target of Plato's actual philosophical concern.

Keywords: model, image, ontological seeming, genealogical seeming, perspectival seeming, heuristic seeming.

Introduction

In the Cave Allegory, Plato has Socrates depict man's grasp of reality by analogy with prisoners captive in a cave. These prisoners believed reality to be nothing more than the shadow show that they have been watching from birth. What they believed to be true reality, Plato suggests, only seemed to be so. The ascent of the freed prisoner out of the cave is depicted as an ascent from the seemingly real to the really real. On his descent back, the newly minted philosopher, having encountered reality itself, tries to convince his peers of the merely seeming nature of their world.

If we read this allegory as pointing directly toward what Socrates undertakes throughout the Platonic dialogues, as I believe we should, we can get a sense of the diversity of appearances the philosopher must fight against when he returns to the cave: sensible particulars seem to be the ultimate reality, there seems to be no difference between knowledge and opinion, virtuous acts could seem to be foolish, whereas vicious acts could appear beneficial, bodily pleasure could seem to be real pleasure, or the philosopher might seem to be a sophist, and *vice versa*. Yet, even though seeming occupies such a central place in Plato's thought, it is not at all clear how we should understand it.² As I will maintain, there is no single notion of seeming that can fully account for all the types of appearances that Socrates deals with.

Plato discloses in *The Sophist* through the words of the Eleatic Stranger the necessary condition for the possibility of falsehood. It is the

² The secondary literature usually treats seeming as if were a unitary phenomenon. I take this to be a mistake. If we rely on only the most general sense of seeming to account for all the diverse contexts in which it is used by Plato, we will surely fall prey to misunderstanding. For articles that deal in one way or another with the notion of seeming, see Nehamas (1982), Deleuze (1983), Silverman (1991), and Moss (2006, 2007, 2008, 2014).

double nature of any image, consisting of a weaving together of “that which is not [...] with that which is” (*Sophist*, 240c) that carries with itself the possibility of falsehood. Seeming and falsehood are deeply connected: when we say that “Y only seems to be X” without being so, we mean that the judgement “Y is X” is false. Yet, we mean more than that. That is, we also mean that there is something within “X” that lures us into believing that it is “Y”. We could suggest that seeming is a luring toward a falsehood. Thus, we can divide the problem of seeming into its constituent parts: (i) falsehood, the analysis of what falsehood is, and (ii) the lure, or the luring towards the falsehood, the analysis of what causes falsehood.

Thus, an analysis of the notion of seeming in Plato could focus either on the arguments for the possibility of falsehood that he developed in *The Sophist*, or on the way that he conceives the luring aspect of seeming, in other words, on its causes.³ Both have received more or less explicit treatment from Plato himself, and benefited from wide attention in the secondary literature. Finally, an analysis of the notion of seeming in Plato could focus on the different species of seeming, if indeed there are any types of seeming. In this paper, I will take the latter road and argue that throughout Plato’s work we can consistently identify two main species of seeming, what I call *ontological seeming* and *genealogical seeming*. The first, I will argue, consists in investing a generated being, i.e. a particular, or what Plato metaphorically calls an image, with the ontological status of its generative formal cause, i.e. the platonic Forms, or models. The second consists in wrongly identifying the character of a particular, thus essentially linking it to a different model than it is of. In order to get an intuitive grasp of what I am aiming at, think of the following situation. Let’s say Cebes sees Socrates in a reflection and says “That’s Socrates!”. If what Cebes meant was that what he saw was Socrates himself – and not a reflection of Socrates – then he would have fallen prey to an ontological seeming. If, on the other hand, Cebes assumed that what he saw was only a reflection, and he meant to identify whose reflection it was, but would have said “Theaetetus” instead of “Socrates”,

³ See Moss (2014) for the argument that both perceptual and value-based seeming are caused by the irrational appetitive soul described in the *Republic*. For the argument that a base part of the rational soul is the cause for perceptual judgement see Nehamas (1982).

then he would have fallen into a genealogical seeming. As we can see, both types of seeming can be understood in terms of the relation between a model and its image. Plato draws on this distinction when he has Socrates describe the knowledge proper to a philosopher: “And because you’ve seen the truth about fine, just, and good things, you’ll know each image for what it is and also that of which it is the image” (*Republic*, 520c-d). Thus, ontological seeming will amount to not acknowledging the image status of an image, while the genealogical type implies an error when connecting the image with its model.

Shifting our attention now strictly to the register of Form and particulars, we can in correlation to the two distinct senses of seeming presented above, determine two different ways in which the model, or the Form, is a standard of truth. First, in relation to genealogical seeming, i) it acts as a genealogical guide, or as a standard for true predication. This means that only by knowing the model can one identify which particulars possess the character of that model. To put it in a more Platonic formulation, only by knowing Virtue itself can one identify a virtuous act and a vicious one. Second, with regard to ontological seeming, ii) it acts as a standard of what is really real and truly true, i.e., of what it means to be “*F*” in the most proper sense. To illustrate, take the following propositions:⁴ a) “Socrates’ decision not to avoid punishment is just” and b) “Socrates’ decision not to avoid punishment is not really/truly just.” From the point of view of my analysis, both a) and b) can be true. While proposition a) refers to applying the predicate “just” to Socrates’ act in the genealogical sense, proposition b) is ontological in that it does not deny the justness of the act, but that his act, or more generally that justice-in-an-act is what Justice truly and really is in and of itself.⁵

⁴ It should count as no surprise that we can analyze propositions as expressing one type or another of seeming. Propositions express judgements, and some judgements may be the product of either ontological or genealogical seeming.

⁵ It is thus not a problem with Socrates’ act, but with the fact that justice is *in* an act as opposed to being in and of itself. By analogy, the same reasoning would apply whenever we want to separate between a certain configuration of black chess pieces that happen to instantiate the checkmate position, being as it is dependent on the white pieces’ configuration, and the rule of checkmate itself.

I will make a further distinction between two types of ontological and genealogical seeming. The first is a) *perspectival*, in the sense that the appearance is caused by perspective, taken in the widest possible sense, and deception follows from an imperfect access to the phenomenon. According to this first type, when “X” seems to be “Y”, it only does so from a specific point of view. Changing the point of view, be it in a literal sense, or by uncovering new information about the phenomenon, will uncover it as a simple appearance. Thus, essential to this type of seeming is that we have at our disposal an explicit or implicit criterion of verification, i.e., a model by which we can expose something as an appearance.⁶ The second type is b) *radical*, in the sense that the appearance is not caused by perspective and our imperfect access to the phenomenon, but by the implicit or explicit usage of a model that is bad, or of a bad definition.⁷ As such, one can use bad definitions in a correct way in identifying particulars, yet still be subjected to appearances. The act by which someone instills in another such bad models I will call *heuristic-genealogical seeming*.⁸ As we will see, while experience or measurement can dispel perspectival seeming, the latter radical type requires an altogether different type of measuring that can only be done through dialectic. In a nutshell, a) the perspectival type of seeming refers to bad or incomplete access to a phenomenon as a cause for error, while b) the radical type refers to the usage of a bad model for accessing phenomena as a cause for error.

I am aware that all this new terminology and plethora of distinctions can probably make the text hard to follow. Consequently, I propose to the reader the following schema of my paper:

⁶ Some examples of this include believing the painting of the cobbler seen from afar to be a real cobbler (*Republic*, 598b-c), taking the submerged stick to be really bent (602c) or being charmed by meter, rhythm and harmony into believing the contents of a poem are true or wise (413, 601a). In all the cases above, the cause of seeming is not essentially related to our understanding of what it means to be *F*, but by our access to *F*. This is why even a child could realize that he is dealing with a painting and not the real thing just by moving closer to it.

⁷ Some examples include believing gold to be what it means to be beautiful (*Hippias Major*, 289e), taking bodily pleasure as real pleasure (*Republic*, 586b), or believing that the principles of returning what is owed represents the nature of Justice (331c-d).

⁸ A good example of such heuristic-seeming is the poets’ description of the gods as changing and deceitful (*Republic*, 380d).

The main distinction: 1) Ontological seeming and 2) Genealogical seeming. The secondary distinction: a) Perspectival and b) Radical. Both elements of the main distinction will be separated in terms of the secondary distinction, giving the paper the following structure: 1a, 1b and 2a, 2b.⁹

At least to my knowledge, no clear articulation of these distinct senses of seeming has been made so far with regards to Plato's epistemology. Most of the times, authors use one sense or the other when interpreting some passage from Plato, but never in any consistent, methodical or explicit way. In order for that to be possible, one would require an analysis of the difference between these two main senses. I propose to offer such an analysis in this article. My account could benefit from a more in-depth look at the details of how the ontological is connected to the genealogical seeming, and also from an enlargement of the analysis to account for some possible objection not covered here.¹⁰ It could also be assisted by a more direct confrontation with Plato's text, and especially with his own account of falsehood from the *Cratylus*, *Theaetetus*, and *The Sophist*. As it stands, this article intends to provide only a general theoretical framework that can be later supplanted and refined.

The model-image metaphor

Before delving in the analysis of the different types of seeming, we need to take a quick look at Plato's image metaphor. By way of this metaphor, we will be able to refer to the ontological structure that grounds the epistemological issues we are dealing with here. The language of model and image, and the relevant relation of model to image (i.e., the imaging relation) is used by Plato to convey the relation of the sensible to the intelligible: in the same way in which an image is said to be of its

⁹ If the reader wishes for a more concrete taste of these distinctions, he is invited to read the conclusion of this paper first, where he will find a side-by-side application of these notions on a case study.

¹⁰ One such objection, for example, would state that the perspectival ontological seeming can be reduced to the perspectival genealogical type.

model, a particular is said to be of its Form. This is one of Plato's most prevalent metaphors.

From an epistemological perspective, the image was used in the *Sophist* to enable the possibility of falsehood in general. The image's ambivalence between truth and falsehood presupposes in turn a relation to the model taken as standard of truth. In other words, only images can be false, and their falsity can be acknowledged only in relation to the model. Model and images are thus related not only in the domain of ontology, but also in that of knowledge. If we are to understand how the model relates to the image with regard to seeming, we must first take a look at what the metaphor has to say with regard to ontology.

Taking the literal¹¹ sense of the model-image relation first, we can think of examples such as the relation between a tree and its shadow or its reflection in water, Socrates and a painting of Socrates, or an event and the verbal reproduction of that event. In the above cases, the latter element of the relation is the image,¹² while the former is the model.

¹¹ For the sake of clarity, I will call the commonplace relation of the model-image relation the *literal* sense of the relation, because it refers to what we commonly take as being models and images. It is on the basis of this literal sense that we are to understand the metaphorical use which Plato will employ. The philosophical use by which the Forms are introduced as the real models, I will call the *metaphorical sense*.

¹² As Patterson (1985) points out, besides εἶδωλον, we also find μίμησις, φάντασμα, ὁμοίωμα, or εἰκῶν as alternative words for image. Depending on whether Plato wants to underline the common ground between image and original we will usually find εἰκῶν / ὁμοίωμα / μίμησις, or in case he wants to highlight the difference, φάντασμα / εἶδωλον. We must bear in mind that sometimes Plato uses "image" in a technical way, as identifying a proper particular, and sometimes in a purely pejorative sense, as pointing out a fraud. The way I see it, based on the discussion in *Republic X*, the image as fraud is nothing but an image of the image in the technical sense, as proper particular. The meaning of "fraud" or "fake" is used whenever Plato wants to highlight that it has taken the place of that of which it is an image of, the same way a painting of a tree could be taken as a real tree. If I correctly understand Notomi (1999, 153-154), he uses a similar type of reasoning when he interprets φάντασμα in the *Sophist* as an imitation of εἰκῶν. If this is the case, I must disagree, for there can be no real analogy between the way the painting of a tree passes as a real tree, and the way a false account passes as a true account. Truthfulness is not something that can be imitated, the same way a tree is imitated. In this article, I will not provide an alternative reading, but only the concepts from which an alternative reading can be constructed.

Plato uses examples such as these to lend force to his metaphorical and philosophical use of the relation when he describes the world as an image of the Forms (*Timaeus*, 29b), drawn geometrical shapes as images of those shapes in themselves (*Republic*, 510-511), facts and words as images of the things they are about, in this case φύσις (*Phaedo*, 100a-b), the written word as an image of the living and ensouled (*Phaedrus*, 276), time as a moving image of eternity (*Timaeus*, 37d), or the material bed as an image of the Form of the bed (*Republic*, 596-597).

It is vital to note that the elements from the two model-image relations, the literal and metaphorical ones, are part of the same metaphysical picture. That which most often is a model for images in the literal sense – trees, beds, actions, and the like – are at the same time images when understood under the metaphorical relation. The concrete bed is a model for the painted one, yet, at the same time, it is itself an image of the Form of bed (510-511). The same relation holds between the drawn geometrical figures and their reflections in water, and those geometrical figures and their Forms (510d-e). Thus, for Plato, the phenomena that are models in the literal sense are models only relatively so: they are models only with respect to their own images. What makes them relative models is the fact that they themselves share with their own images the characteristics of an image, they are generated, derivative, and dependent in both nature (how they are) and identity (what they are), to that which generated them, to which they are dependent in regard to nature and identity, and whose identity is autonomous. These ontological characteristics should be had in mind as essential to Plato's usage of the term "image" and not the accidental ones that come to mind when one thinks of ordinary images. The latter, in an ironic twist, provides a paradigm case for these ontological characteristics which constitute the essential meaning of "image."

As we can see, the concrete particulars, the phenomena we encounter in our everyday lives, play a double role for Plato: from the standpoint of the sensible world, and for the non-philosopher they are the ultimate reality – autonomous models for derivative things such as paintings, shadows or reflections, written or spoken accounts. The latter, being images, are said to be *of* their model. Their identity and sometimes their

existence¹³ are dependent on it: it is by being *of something* that they are what they are, a shadow of a tree, a painting of a bed, an account of Socrates' just acts. The model thus constitutes the identity of the image. Yet from the standpoint of the metaphorical use of the relation, i.e., from the metaphysical standpoint, these relative models are themselves mere images, they are *of* their Forms in an analogous¹⁴ way to how their images were said to be *of* them.

If we take the model-image relation as constituting the structure of the real, we can see how my two main senses of seeming can be applied. The first, which I have called *ontological seeming* would amount to taking what is only an image, be it in the metaphorical or the literal way, as playing the role of that of which it is, or purports to be, an image of. The second, or *genealogical seeming*, would consist of taking an image as being of a different model than the one it actually is of.¹⁵ Both types of seeming, I will maintain, can be further divided along the lines of their objects: if the object of seeming is sensible, then the cause of seeming will be in some sense perspectival. I will call this species of seeming *perspectival seeming*, be it ontological or genealogical. If, on the other

¹³ This is true especially in the case of shadows or reflections. If the models cease to exist, so do the images. Paintings or sculptures, on the other hand, only depend on their model for their identity and not for their existence. For the argument that Plato has in mind in the first type of relation, see Allen (1960) and Lee (1964). For an argument for the second case, see Patterson (1985, 46-47, 171-180).

¹⁴ As crucial as it is to understand the genitive sense of the *being of* relation that binds an image to its model for understanding how Plato saw the relation of Form to particular, such an analysis cannot be accomplished here. It would require a research project of its own, one that as far as I am aware of has yet to be undertaken.

¹⁵ This would also account for seeming with regard to accidental properties, such as dimension, color, temperature and the like. The sentence "The stick is bent" could be taken to mean either that the particular stick is in part an image of Bentness itself or that the bentness in the stick is an image of Bentness. For the first analysis see Patterson (1985, 197-198), for the second see McPherran (1988, 533). With regard to the treatment of attributive statements as relational model to image statements: "What appear to be attributive statements are in fact relational or identifying statements, depending on the designation of their predicates. In derivative designation, to say of something that it is F is to say that it is causally dependent upon the F. Notice that "F" is here not strictly a univocal term, but a common name, applied in virtue of a relationship to an individual, the Form" (Allen 1960, 150).

hand, the seeming consists in investing an image with the role of a model I will call it radical ontological seeming. Radical genealogical seeming amounts to using unsound models in identifying particulars. When someone uses these models to teach others, I will call it heuristic seeming, which consist in describing a model¹⁶ in a false way. This latter sense is the most general in scope, as it can imply, but is not restricted to, radical ontological seeming. One could describe Justice by using examples of actually unjust acts, or one could take what is actually a just act, but say that that act itself is what it means to be Just. In the following sections we will take a closer look at each of these types of seeming.

1. *Ontological seeming*

Taking into account the ontological distinction between model and image, we can understand ontological seeming in terms of the following error: investing what is only an image with the identity, the role or the function of the model of which it is an image. Alternatively, we can formulate this in terms of an improper unqualified application of a name:¹⁷ if the name *F* belongs in a proper sense only to the Form *F*, an ontological radical seeming would consist in applying it unqualifiedly to a particular that poses it only relatively so, or through participation.

The Form of Bed, the concrete bed, and the painting of a bed, are all called by the name "Bed". Yet, the difference between them is of another kind as that between two particular beds, two different paintings of a bed, or between a bed and a chair, for instance. One, albeit partial, way of understanding this difference is through the notion of ontological dependence. While the Form of Bed is what it is without

¹⁶ One way to make this clearer is by imagining that the genealogical seeming implies a "downward" movement of identification, from a model to an image, while the heuristic seeming suggests an upward movement of description, from the image as example to the model as what is exemplified.

¹⁷ Only the Form is called "*F*" in an unqualified manner, while the particulars are called "*F*" only qualifiedly so, or in relation to the Form, the proper bearer of the name. For a detailed development of this approach see Allen (1960, 149-151) and Patterson (1985, 69-70).

reference to anything else, the concrete bed is recognized as being “Bed” only in the light of the Form of Bed, the same way the painting of the bed is labelled “bed” with reference to the concrete bed. Thus, we can speak, at least in this case, of a hierarchy between three different ontological levels in which we can talk about the bed. The way we differentiate between them is by taking away or adding to their reality or truth,¹⁸ which in turn is reflected in the way they deserve the *name* “Bed”, either in relation to something else, thus qualifiedly, or in and of themselves, or unqualifiedly. By contrast, two different concrete beds deserve the name “bed” in the same way with regard to reality and truth, and should probably be differentiated with regards only to their matter, shape, and other accidental qualities. On the other hand, a bed and a chair, for example, would differ concerning the Form they instantiate, or, in the terminology I employ here, they would differ as to the model they are images of.

Ontological seeming thus holds “vertically” in respect to ontological dependence, between that which can be understood as the generative formal cause and the generated particular. This type of seeming does not need to be an explicit assertion that “X” is what it means to be “Y”, but, as is most often the case, just an implicit behavior that naturally assumes the sensible reality or the particulars to constitute the ultimate real. The reason I call this type of seeming ontological lies in the fact that it invests that which is generated, derivative and dependent, with the function of that which generated it, and to which it is dependent in regard to nature and identity.¹⁹ To use a simile, this type of seeming would amount to taking the royal messenger as the king himself. By contrast, the genealogical seeming would amount to wrongly identifying who is the king’s messenger. If the latter type consists in a wrong attribution of the predicate “messenger” to a subject, the former amounts to investing the messenger with real, true and ultimate authority. This distinction, I

¹⁸ For an interpretation of what constitutes for Plato the criterion by which something is considered more real and consequently more true, cf. Heidegger’s analysis of being as presence (1997, 23) and (2002, 38).

¹⁹ This implies that the function of the model, even if not understood thematically, is always at play. If we understand the function of this model as what makes intelligibility possible, we can already get a sense of why there is a problem in investing that which is made intelligible with the function of making intelligible.

believe, is reflected in Plato's emphatic usage of terms like "really real", "truly true" or "truer". These terms are used to differentiate between the proper bearer of a name, the Form, and that by which, through participation, a particular also receives its nature and name. As I shall discuss in the final chapter, this use should not be taken to mean that it is not true to predicate "*F*" about some subject, because it is only apparently so, as would be the case of the submerged stick that only appeared bent! The latter case is a type of genealogical or predicative use of seeming that should be carefully separated from the ontological one, which only governs the right use of the unqualified name.

The two readings of the model-image relation thus offer us two ways of understanding the ontological seeming. Under the literal reading, the model will be a particular such as a tree, a house or an action, while their images would be things like paintings, reflections or written or spoken accounts of these models. In this case we can think of situations as when one takes what is only a reflection for the thing reflected, or when one mistakes the painting of the tree seen from afar with a real tree. These cases fall under the category of *perspectival* ontological seeming. These, as we shall see, are due mainly to an imperfect or partial access to the object, which I will try to understand through the notion of perspective.

On the metaphorical reading we have seen that the worldly phenomena are models only relatively so: their identity is autonomous and they are generative causes *only* in relation to their images. In and of themselves though, they share with their images the same relation to the Forms that their images have with regard to them. What I have called radical ontological seeming comes about whenever one believes that the relative models of our worldly experience are the ultimately real and autonomous being, and thus invests them with the role proper only to the Forms. Thus, if the perspectival mode of seeming was caused by an incomplete or improper access to some phenomenon, the radical type consists in having a corrupted tacit or explicit understanding of what it means to be a certain phenomenon.²⁰

²⁰ In other words, while the former refers to deception stemming from our improper access to the phenomena of our experience, the latter points to deception coming from bad "concepts" with which we access this experience.

1.a *Perspectival ontological seeming*

The perspectival type of ontological seeming holds only between what I have called the literal relata of the model-image relation, the phenomena of our everyday experience, and what we usually call images.²¹ As the name implies, this type of seeming is dependent on perspective and because of this the mistaking of the model for its image is only temporary, or accidental. The examples Plato furnishes for this type of seeming are usually based on illusionistic painting. The painting of a cobbler – when seen from the right distance – can seem to children or to foolish people to be a real cobbler (*Republic*, 598b-c). In the *Sophist* (234c-b) the Stranger offers a similar example in the case of a drawing that seen from afar can seem real to some people. Both examples serve Plato in illustrating how the poet in the first case, and the sophist in the latter could seem for the ignorant to be able to produce everything there is, and thus to have universal knowledge. One of the tricks by which the deception of the poet and that of the sophist operates can be seen through the benign example of painting: like the painter, they create only the images of phenomena. These in turn make the audience, who “judge by color and shape” (*Republic*, 601) believe that they are in contact with the phenomenon itself.²² Putting it in another way, the same way that for some children the visual aspect of a tree is enough to make them believe that what they are seeing is a real tree, so for the ignorant and young some aspect of virtue that shines through a discourse would immediately make them believe that they are witnessing true virtue. Yet, there is an important disanalogy here that we must be aware of. For

²¹ It is vital that we understand the product of imitation on the lines of the image and not on that of the copy. Briefly put, the copy and the model are the same type of things, e.g., a key and the copy of a key are both keys, while the image must necessarily not be the same kind of thing as its model in order to be an image. For Plato’s remarks on this subject, see *Cratylus* (432). For a development of the distinction between the model-copy relation and model-image relation, see Patterson (1985, 25-63).

²² This is far from the full picture of what happens in such cases. In my view, a proper analysis of the way the sophistic deception operates would require the careful deployment of the conceptual net I am trying to develop here. For this reason, in this article sophistic deception will not be itself rigorously analyzed so much as it will serve as an illustrative case for the concepts I am trying to present.

even children have a solid enough grasp of what it means to be a tree that they would be able to easily dispel the appearance were they to get closer to the painting. On the other hand, the audience of the poet or the sophist are so inexperienced in what virtue or wisdom are, that they will mold their understanding of these phenomena on the discourses they hear. What could be perceived as inadequacies in the discourse by someone who has real knowledge of virtue, for the ignorant would simply amount to another aspect of what virtue is. This is exactly the point the Stranger makes when he says that there is another type of expertise next to that of the illusionistic painter that “someone can use to trick young people when they stand even farther away from the truth about things” (*Sophist*, 234c-d). In this case, the trickery comes about not from the distance to the image, as was the case of painting, but from the distance to the “truth about things,” or in my terminology, from their model.²³ This point should be kept in mind, for the perspectival ontological seeming fits neatly only for the painter, but it is not enough to account for the deception of the sophist or the poet.

Before moving on, we should take a quick look at some of the characteristics that make up the perspectival type of seeming, in order to better distinguish it from the radical type. These characteristics apply both to the ontological and the genealogical type that I will discuss later, but for now I will illustrate it using an example of ontological perspectival seeming.

1) It is dependent on perspective, i.e., it holds sway only as long as the right perspective is in place. The optimal perspective for an illusion makes only the identity between the image and the model visible, while hiding their difference. We can think here of how a painted tree seen from the right angle can seem to be three dimensional, or how a scarecrow seen from a certain distance could seem to be a real man. In both cases the aim of the former was to look exactly like the latter in

²³ Notomi (1999, 139) gives a similar reading to this passage, while Benardete (1984, 106) while giving the same interpretation to the structure of the argument, interprets the “truth about things” as indicating deceptive life-experience as opposed to discourse. I see no reason for such an interpretation, for I, like Notomi, believe that the contrast intended in this passage is between ignorance and knowledge, and not, as Benardete seems to imply, between ignorance from words and ignorance from life-experience.

some respect, and there is a perspective that allows exactly that. To put it more concisely, what enables the possibility of confusion between a thing and its image is perspective. The right perspective,²⁴ or point of view, is what enables the possibility to hide the obvious difference between the image and its model, and lets only what they have in common be seen.

From this point we can infer the second characteristic of the ontological perspectival seeming:

2) It rests upon an implicit or explicit distinction between model and image. Changing the optimal perspective can instantly uncover the image character of the phenomenon, e.g., looking at the painted tree from the side rather than the front can make it obvious that we are dealing with a painting and not a real tree. This tells us that we are consciously or unconsciously in possession of a regulative idea with regards to what it means to be a tree, and that the painting of the tree, upon further verification, does not pass the test. This is why someone who deceives in this manner always takes perspective into account.

The model, be it relative or absolute, or the criterion of what it means to be "X," is thus developed enough to allow for differentiations not only between trees and rocks, but also between trees and images of trees. It is essential to this type of seeming that there be a model in regard to which the image's character can be brought to light as a simple image, following an investigation. The model needs to be formally distinct from what shows itself, and to function as an evaluative criterion for the manifestation's claim to be this or that thing. The model is that which enables us both to doubt a manifestation's claim to being,²⁵

²⁴ Perspective is what makes this type of seeming possible, but it is not a sufficient condition for it. In order for someone to fall prey to an appearance, he must give his assent to it. Yet he can withhold his assent, or otherwise correct the appearance by means of his knowledge of the world. For a full discussion of the relation between belief and assent in Plato, see Moss (2014). For an account of how background knowledge can alter the beliefs formed about the same perceptual phenomena, and thus how the educated and non-educated can have completely different beliefs about the same phenomenon, see Silverman (1991).

²⁵ I am thinking here of situations when something appears to be the case, but we withhold our assent. In this sense I am saying that some manifestation "claims to be" something. A cardboard apple might appear to be a real apple, and thus "claim to be" an apple. The model, i.e., our implicit or explicit understanding of what an apple

while also providing the criterion on which to test it. There are some characteristics that we find necessary for a thing to be what it is. For example, if we were to get a closer look at what we thought was a tree, and find out that it was a two-dimensional painting, we would not say that we found a different species of tree, but rather an image of a tree. Because this type of seeming refers to the literal model-image relation, one need not be a philosopher in order to realize that what he is seeing is not the real thing. The model he uses in distinguishing the image from the original is constituted by the phenomena of his everyday experience. The fact that we are able to correct ourselves and be aware of the possibility of false appearances indicates that we rely more or less tacitly on a separation between model and image.²⁶

The last characteristic of the perspectival ontological seeming is that:

3) It is possible only inside a medium. As Plato shows in the *Timaeus*, the notion of image presupposes that of a medium in which it can come about:

Then we distinguished two kinds, but now we must specify a third, one of a different sort. The earlier two sufficed for our previous account: one was proposed as a model, intelligible and always changeless, a second as an imitation of the model, something that possesses becoming and is visible. We did not distinguish a third kind at the time, because we thought that we could make do with the two of them. Now, however, it appears that our account compels us to attempt to illuminate in words a kind that is difficult and vague. What must we suppose it to do and to be? This above all: it is a *receptacle* of all becoming – its wet-nurse, as it were. (*Timaeus*, 48e-49b).

Plato then adds, in *Timaeus'* words, that the image should not be taken as something in its own right, being as it is split between its debt for what it is to the model, and for the possibility of instantiating that identity to the medium:

is, will operate as a criterion of verifying whether or not the cardboard apple's claim to be a real apple is justified or not.

²⁶ The lack of a distinction between model and image, as I will argue later, would amount to a Protagorean world, where images, and thus falsehood would be principally impossible.

That for which an image has come to be is not at all intrinsic to the image, which is invariably borne along to picture something else, it stands to reason that the image should therefore come to be *in* something else (*Timaeus*, 52c).

In other words, when something manifests itself to us through a medium, we are in contact with an image of the thing,²⁷ and not with the thing itself. I believe that for Plato all but the soul's contemplation of the intelligible realities constitutes fundamentally mediated contacts with phenomena.²⁸ The concrete bed of our everyday experience is for Plato only an image of the Form of Bed, or, seen from the other way around, the concrete bed is the Form of Bed as mediated by the Receptacle. Yet, the concrete bed can also itself be subject to mediation: we can come into contact with it through its images, such as through painting, reflections or shadows.²⁹

It must also have been – the image represents the model. The mirror might be straight and thus afford accurate representations, or it might be crooked and create inaccurate images. The example of the giant statue that is made disproportionate to compensate for perspective, and thus appear proportionate to the viewer (*Sophist*, 235-236d), proves that Plato was not only aware of the effects that a medium can have on how the image represents its model, but also that this effect can be predicted and used consciously. A second way in which the medium plays a determining role is by the fact

²⁷ This is true only for one type of medium, for example the Receptacle, but not for the light in *Republic* (508), which is a medium in a different sense.

²⁸ The language used in the *Republic* when describing the philosopher's grasping of the Forms as "whenever someone tries through argument and apart from all sense perceptions to find the being itself of each thing and doesn't give up until he grasps the good itself with understanding itself" (532a-b), seems to me to support interpreting only the soul's contact with the Forms in terms of un-mediation.

²⁹ It could be argued, based on (*Republic*, 598a-b) that the visual image one has of the bed is precisely what the painter copies when he tries to impart it on a different matter, e.g., on canvas and paint (see Nehamas 1982, 263). The fact that one perceives images of things does not imply that Plato was a representationalist, because the image is not a mental entity, but an objective thing that can be perceived or copied. To put it differently, we perceive something which in turn is called an image as a metaphor to highlight its ontological status. This should not suggest that we perceive the world as mediated by something like mental images.

that it determines how, and in what respect the image represents the model. Different mediums offer different possibilities of representation;³⁰ think of how depicting a human being in sculpture, in painting or in speech will affect what the respective image will be able to say about their model.³¹

We can thus define the perspectival ontological seeming as the putting in place of a perspective or point of view inside a medium, where the difference between the model and the image is hidden in favor of their identity. As we have seen, a mere change in perspective could deal the killing blow for any such seeming. Yet, in order for that to happen, I argued that there must be in place either a tacit or an explicit separation between model and image,³² where the model acts as a criterion for verifying the manifestation's claim to be the thing that it initially suggests to us that it is. The fact that Socrates says that "only children and foolish people" (*Republic*, 598c) could be deceived by the painter's illusionist painting, shows that for Plato this type of seeming was of no great concern. Rather, I believe the main reason he talks about it is the fact that, in this way, he can furnish an analogy for the type of seeming that befalls the prisoners in the cave, the radical ontological seeming.

1.b *Radical ontological seeming*

This type of seeming occurs whenever one takes the phenomenal reality as being ultimately real, and not itself dependent on and determined by the Forms. It is thus a seeming that takes place between what I have called the metaphorical relata of the model-image relation. I believe that

³⁰ Cf. *Statesman* (286) where Socrates says that *logos* is the only proper medium for images of abstract notions.

³¹ For a more detailed account see my 2017 article *The Platonic Receptacle: Between Pure Mediality and Determining Cause*.

³² A change in perspective can prove the deficiency of phenomenon only if we have at our disposal a model, i.e., a criterion, in the light of which something could appear as a deficiency in the first place. Otherwise, we would have to take the would-be deficiency as just another property of the phenomenon, e.g., if we lack any prior substantial knowledge of trees, seeing that a tree was made of plastic would in no way prove that we are dealing with a fake tree. Rather we would probably be tempted to think that this is what trees are made of.

for Plato all but the true philosophers are experiencing this type of seeming and remain entangled in it.

The reason I have called it *radical* lies in the fact that it takes the relative models of the phenomenal world, which ontologically are just images, as – or in some way as – absolute models. This in turn leads to investing the image, i.e., the concrete particulars, with the evaluative and prescriptive roles that are proper only to the absolute model. The radical aspect of this seeming comes from the fact that that which is taken as a standard for what is real, and by extension that by which the real is judged, is itself, in Plato's terms, not really real, just an image. This opens up an important question: if the model by which we judge what is real is itself an image, that is, it is itself in a sense unreal, how is it possible for anyone to uncover its relative unreality? It is this apparent circularity that affords it its radical character.

One of the more poignant and explicit formulation the Plato has to offer regarding this type of seeming is the following:

What about someone who believes in beautiful things, but doesn't believe in the beautiful itself and isn't able to follow anyone who could lead him to the knowledge of it? Don't you think he is living in a dream rather than a wakened state? *Isn't this dreaming: whether asleep or awake, to think that a likeness is not a likeness but rather the thing itself that it is like?*

I certainly think that someone who does that is dreaming. But someone who, to take the opposite case, believes in the beautiful itself, can see both it and the things that participate in it and doesn't believe that the participants are it or that it itself is the participants – is he living in a dream or is he awake? (*Republic*, 476c-d).

In the case of the perspectival ontological seeming, it was the irrational soul that fell prey to optical illusions and that had to be corrected by the measurements (*logismos*) of the rational soul.³³ It was the rational soul's job to decide whether the two-dimensional painting of a tree, even

³³ For a full analysis of the role the rational soul plays in dispelling appearances, see Moss (2008).

though it certainly looked like a tree from one side, has what it takes to be called a real tree rather than just an image tree. This procedure implies verifying the painted tree's claim to be a real tree upon the independent criterion of what it means to be a tree. The radical seeming on the other hand comes about when one ends up believing that the phenomena used as standards in the case of perspectival seeming are absolute models. It would be as if someone believed that the painting of a lyre is grounded and dependent for its identity upon a concrete lyre, but would not believe that the concrete lyre would need any further analogous grounding. In other words, he would accept that we recognize the meaning of the painting by reference to the concrete object, but would not extend the same relation of dependence to the concrete object and an intelligible Form.

Thus, it can be stated that the radical ontological seeming consists in applying the role of absolute models to phenomena which are only relative models. The perspectival ontological seeming was empirical, dependent on a point of view, and could be easily dispelled by a simple change in perspective. The radical ontological seeming, on the other hand, consists in taking as a criterion for what is real something that is actually an image. From this we can delineate two essential ways in which the radical seeming is different from the perspectival one.

First, in the case of the radical ontological seeming there is no independent criterion immediately at hand by which to dispel the seeming. If the model itself is imbued with the characteristics of the image, how are we to step outside of what it claims to be real, and judge it as unreal? This situation would at first hand seem as a case of one trying to jump over his own shadow. The second difference follows from the first: there is no possible change in perspective, no change within the properties of the pseudo-model which could show its inadequacy and as such expose it as an image, as long as we don't have any independent criteria of evaluation.³⁴ We can state this more

³⁴ Alternatively, this can be understood as a case of not separating between intension and extension. In this case, all changes within the extension would reflect in the intension, and *vice versa*. If one were to take the meaning of "hot" to be a particular hot object and use it as a criterion to identify other hot things, then if the original object cooled down, then it would accordingly change the criterion by which hot

forcefully in terms of the cave allegory's notions, in the following way: *no event from within the shadow world could ever show us that it is a shadow world.*³⁵

The paradigm case of this type of seeming can be found in the condition that befalls the prisoners in the cave allegory. Their reality, and consequently what they take as models for what is real, is made of mere images.

More concrete examples of radical seeming include Hippias' answer to the question of "what is Beauty?" with "gold" (*Hippias Major*, 289e), or believing the "friend of a friend" to be the friend itself – alternatively, confusing the means for the goal (*Lysis*, 219d). Socrates also warns in the *Republic* (597) that whoever were to take the bed produced by the carpenter as "completely that which is," instead of the Form of Bed, "would risk saying what isn't true". Also, the identification of true pleasure with bodily pleasures, which are described as "mere images and shadow paintings of true pleasures" (*Republic*, 586b) would constitute a common deception. Even the geometers can fall into the same kind of trap if they take, as Socrates indicates they often do, their hypothesis as real principles (533b). In the same vein, we can understand Diotimas' description of the journey of the soul from images of beauty, e.g., the beautiful body, or the beautiful soul, towards Beauty itself (*Symposium*, 210-211d), as a journey from the image towards the model. We can safely assume that if one were to voluntarily stop in his upward journey to one of these images of Beauty, he would do so only if he would

things are selected. This, of course, would be highly implausible with something like hotness, but not so much with non-sensible concepts like virtue, or justice.

³⁵ An account of the platonic solution to this problem would require a work on its own. For now, I can only suggest that for Plato the relative-model is laden with tension. By this I mean that at closer inspection the nature of the image taken to be a model and the meaning of the model it is taken as, will come to light as different and inconsistent. Socrates usually exploits this inner tension when criticizing his interlocutors' choice for models, or for what a thing truly is. One of the most common ways of refuting his interlocutors, especially in the early dialogues, was for Socrates to prove that you can have the supposed model (the image), without the properties or effects of the model that it is supposed to be, i.e., gold without beauty (*Hippias Major*, 289e), or you could respect the principle of always returning what was borrowed without bringing about justice (*Republic*, 331c-d) etc.

wrongly believe that there is nothing better to be found, in other words, that he has found Beauty itself.

What we must be clear about is that the seeming in these cases consists not in saying that gold is beautiful, or that bodily pleasure is pleasant, but in thinking that gold is *the* Beautiful and not just an image of Beauty, that bodily pleasure is *the* Pleasure and not just an image of Pleasure. This is analogous to the way it would be correct to say that a painting of Athens is of Athens,³⁶ but not that it *is* Athens, or that a checkmate position is an instantiation of checkmate, but not the checkmate itself.

Why is radical ontological seeming bad?

In order to understand the negative consequences that radical seeming brings about, it is necessary that we take into consideration two things: the function that the model plays, and the nature of the relative models. With regard to the first aspect, probably the most obvious role that knowledge of the Forms plays in the practical life of human beings for Plato is that of standard for true predication, or true genealogy. This can be rendered as the Socratic assumption³⁷ that in order to know which mode of life is virtuous one must first know what Virtue itself is. In this light we can understand more clearly Plato's dismay towards the unreflective confidence in one's knowledge that Socrates' interlocutors so often exhibit. For if one either tacitly or explicitly takes as the criterion for being just some just act, or type of act, i.e., an image of justice, then whatever is true of that image will creep into the meaning of Justice itself. This, as we will see, constitutes a problem because the structure of the particulars precludes them from acting as absolute models.³⁸

³⁶ The analogy is not perfect, though. A picture of Athens will always be a picture of Athens, while gold, or a fair maiden could cease to be an image of Beauty. Socrates makes Hippias concede both that the wooden spoon can be more beautiful than the golden one, and thus make the latter appear ugly, or that comparing the goddess with the fair maiden would make the latter appear no prettier than a monkey.

³⁷ Geach (1966) goes so far as to call it the "Socratic Fallacy".

³⁸ For an account of how radical ontological seeming comes about in the first place that takes into consideration the epistemology of Book V of the *Republic*, see Smith (2012).

To the lover of sights who “doesn’t believe in the beautiful itself or any form of the beautiful itself that remains always the same in all respects but who does believe in the many beautiful things”, Socrates presses the question:

[...] of all the many beautiful things, is there one that will not also appear ugly? Or is there one of those just things that will not also appear unjust? Or one of those pious things that will not also appear impious?³⁹ (*Republic*, 479a-b).

It is one of the defining characteristics of the particulars that we cannot say of “any one of them any more what we say it is than its opposite” (479b-c). The contrast between the changing nature of particulars⁴⁰ as opposed to the unchanging nature of the Forms is also emphasized in the *Phaedo*, when Socrates asks whether they “in total contrast to those other realities, one might say, never in any way remain the same as themselves or in relation to each other?” (78e).

³⁹ Against the approximation view that would suggest that a beautiful particular can appear ugly because it is imperfectly beautiful, see Nehamas (1975). I agree here with Nehamas that what makes the sensible world roll about between extremes is the fact that their being “X” is dependent both on relation to other things and on context. This interpretation allows that in a determinate context we can say with confidence that something is just rather than unjust. Yet even if some action is just in a given circumstance, that does not guarantee that it will be so in all circumstances. See Patterson (1985, 95-100) for a critique of what I too believe to be wrong with Nehamas’ position.

⁴⁰ I believe we can understand the ever-changing nature of the particulars in a twofold fashion. Taking first the relative possession of properties, one thing’s being small or large, hot or cold, beautiful or ugly, depends entirely on how it relates to the thing compared to. Thus, one and the same particular can have the same height and the same temperature, and still be dubbed large or small, hot or cold, depending entirely on what it is compared to. The other sense of changing relates to the determinate properties that something has, e.g., someone might be six-foot-tall and have a body temperature of thirty-six degrees Celsius. These properties are also liable to constant change. Thus, particulars not only possess their properties relatively, but the relative relations between particulars are themselves liable to constant change. This in no way precludes the possibility that there can be true predication about particulars as long as the statement is qualified. For a more detailed account of the changing nature of particulars see Fine (2004, 54-57).

The case of the misologues from the *Phaedo* makes an emphatic case for the consequences of demanding of that which is inherently unstable and always changing, i.e., the image, to act as a model, and consequently to act as criterion for what is real and for the nature of things:

Those who spend their time studying contradiction in the end believe themselves to have become very wise and that they alone have understood that there is no soundness or reliability in any object or in any argument, but that all that exists simply fluctuates up and down as if it were in the Euripus and does not remain in the same place for any time at all. (*Phaedo*, 90b-d).

This is a good example of how the lack of a clear distinction between the argument⁴¹ and the thing the argument is about ends up transferring the properties of the argument upon the thing the argument is of. This in turn bestows upon the practitioners of eristic a false type of wisdom. By analogy, it would be as if someone who saw different paintings of Athens at different times would end up believing that Athens itself was changing. An even more radical situation, and, I believe, the final stage of the eristic false wisdom, would have someone believe that there is no difference at all between the paintings and Athens.

Because for Plato the aim of politics is so intimately connected with justice, we can see why the ability to clearly separate model from image, an ability which distinguishes the philosopher from the common folk, also translates in the logic of the *Republic* as the criterion which separates the should-be ruler from the ruled:

Since those who are able to grasp what is always the same in all respects are philosophers, while those who are not able to do so and who wander among the many things that vary in every sort of way are not philosophers, which of the two should be the leaders in a city? (*Republic*, 484b-c).

⁴¹ The argument and the thing the argument is of constitute a case of the image-model relation. In the *Phaedo* Socrates later compares arguments with images: "I certainly do not admit that one who investigates things by means of words is dealing with images any more than one who looks at facts" (100a1-2).

One way,⁴² then, of understanding the negative impact of the radical ontological seeming, is that by investing an image with the role of the model, we end up doing and asking of it things that should be asked and done only with something that has the characteristics of a true model. Radical seeming weighs heavily especially upon practical questions as “What is virtue?” or “How should one live?” Thus, for example, someone, by seeing how actions that he at one time took not only *as* just, but as a model for justice, at another time appear unjust, could start believing that there is no stable nature to justice at all, that it is always changing and shifting. The Platonic insight against such tempting relativism comes, I believe, by way of an analogy: the same way you won’t judge a lighthouse to be unstable and ever-changing just because it appears to change in size as you move closer or further away from it, i.e., the same way in which you separate between images of the lighthouse and the lighthouse itself, the same should be done with Justice and just acts.

In order to get a more revealing look at the character of the radical ontological seeming, we must revert back to the relation of dependence that the image has to its medium. As the *Timaeus* (48e-52d) showed, the Forms needed a medium in which to imprint their character in order to give rise to the images. Accordingly, the image has a twofold origin. On the one side it is indebted to the Form for its character. On the other side, it is also indebted to the Receptacle for its existence.⁴³

If the image-nature of a phenomenon is hidden, i.e., if it is believed to be an unmediated showing of the true reality, then it follows that so is the presence and consequently the effect that the medium⁴⁴ has

⁴² If this argument relies on the practical consequences of radical seeming, another, probably more fundamental way for Plato of understanding the problem with radical seeming would appeal to the proper function and place of the soul. As we see most poignantly in the *Phaedo*, the soul’s contemplation and nearness to the Forms is good in itself.

⁴³ And, arguably, for all the characteristics that particulars have and which do not originate from the Forms, such as spatiality, visibility, composability, decomposability, being in flux and being perceptible.

⁴⁴ Hiding the image status of a phenomenon equates with hiding the medium, and *vice versa*. In the absence of a medium, we do not have the logical resources to talk about seeming as opposed to being. In other words, in the absence of a medium, of that

on the way it appears. If one were to believe that what he sees is not mediated in any way, he would not be able to use the word “appear” to indicate change, for instance, but rather only “is.” Looking at a stick that appears bent when submerged and then taking it out of the water, he would not be able to account for the change in aspect by saying that the stick merely appeared bent, while being straight all along, but rather that it was bent and then it straightened out. This is due to the fact that we usually distinguish between the actual properties that a thing has, and the apparent properties that are due solely to the influence of the medium in which we perceive the object. If we cannot find the medium accountable for the property “bent” that the stick took on when submerged, and thus construe it as an apparent property of the stick, then we are forced to take it as a real property. If one presupposes that he has an unmediated contact with some object, then he does not have the tools required to construe any of the changes the object suffers as apparent changes. Rather, whatever aspect the object takes will have to be construed as a real change in the object itself.⁴⁵

The relevant point here is that if one takes what is ontologically an image as a model then, because images are ever-shifting between opposites, he would be compelled to believe that the model itself suffers these changes, and consequently end up entertaining the same type of beliefs⁴⁶ towards Justice for example, as do the misologues with regard to the objects of argument: namely, that all “simply fluctuates up and

which mediates the model through images, we do not have the possibility of doubt: whatever presents itself to one cannot be separated from what is. If for example, we are not aware that we are watching a video projection of a locomotive coming towards us, we would have no resources to doubt that a locomotive is indeed heading our way. The fact that Plato was aware of this logical implication can be supported by the fact that in *The Sophist* only the image can carry falsehood, and so only the image can support the possibility of doubt.

⁴⁵ Consider the case of an object that constantly shifted colors. Think of how one would proceed in deciding whether the object actually changes color or if instead the colors are due to some source of light that is projected on it.

⁴⁶ This type of relativism must follow at least an active reflection on the subject matter, so it would not be a danger for the usual Athenian who, if we are to take the Socratic dialogues as reference, has difficulty in even understanding what Socrates means when he asks of them to give an account of the unitary aspect of a thing. Rather, I believe, this position is more closely related to the sophists.

down as if it were in the Euripus and does not remain in the same place for any time at all" (*Phaedo*, 90b-d).

One of the main functions of the model is that of stating how things ought to be in order to be recognized as images or instances of that model. If the model we use is not authentic, then all the genealogies that we will use it for run the risk of being wrong. If, for example, we take bodily pleasure as being what Good is, then by this one radical ontological seeming countless genealogical ones will follow. A crooked model of the Good will be used to wrongly identify what is pleasurable as what is good in any given situation, at the expense of what is truly good. Now we shall turn to one of the main consequences of the radical ontological seeming: genealogical seeming.

2. *Genealogical seeming*

One plausible interpretation of how Plato conceived of the way we identify the character of particulars is that it goes along the same lines that one would proceed when connecting an image to the model it is of.⁴⁷ Following this interpretative direction, whenever we get something wrong about a particular, either if we identify it wrongly or we predicate something false about it, we are committing what I called a genealogical error. If the identity of particulars is provided by the Forms through participation, then whenever we determine a particular in some way, either as being this or that, or as being in this or that fashion, we do so by identifying it as an image of some Form. When we say something like "That statue is proportionate," we take something *as* a statue, and also *as* being proportionate. Yet, in both moments of judgements we can err:⁴⁸ what we took as a statue could prove to be a painting, and what looked proportionate from a distance might seem disproportionate from a better point of view. Thus, this type of seeming occurs when what is

⁴⁷ See Allen (1960), Lee (1964), Patterson (1985), for authors that take the model-image metaphor as crucial for understanding the relation of Form to particular, and of predication.

⁴⁸ For an analysis of the relation between the structure inherent to judgement of taking "something as something," see Heidegger (1997, 416-417) and (2002, 220-221, 225).

ontologically an image is incorrectly identified, i.e., whenever it is put in a genealogical relation with, or as participating in the wrong model.

The main difference between this type of seeming and the ontological one can be put as follows: while the ontological referred to the act of collapsing the difference between model and image by placing the generated in the role of the generative, the genealogical on the other hand refers to the act by which we connect them in a wrong way.⁴⁹

In this section I shall talk about two types of genealogical seeming. The first is perspectival. It regards wrongly connecting an image to its model because of the cosmetic effects that the medium can have on the image, thus making it appear as of some other model than its true one. The second type I have called radical. This concerns wrongly connecting an image to its model due to the model having been defined in an improper way. Even though one can use a bad definition correctly, this will still not get him any closer to the truth.

We can think of situations like wrongly identifying a person when he is far away, misidentifying the subject of a painting, taking an object's reflection as being that of another, or, through some ingenious trickery on Theaetetus's part, taking him to be flying when he is actually just

⁴⁹ We should resist the temptation of reducing the radical ontological seeming to the genealogical one as still another case of predication. While the genealogical is concerned with identifying the character of something following a pre-established criterion of identification, the ontological concerns these criteria of identification themselves. While the former refers to rule following, the latter is a matter of rule setting. By analogy, in the case of chess a genealogical seeming would consist of wrongly identifying a position as checkmate, while an ontological seeming would be more akin to taking a certain checkmate pattern as being what checkmate is. Yet, if in this case the distinction is more poignant, it may seem a lot fuzzier between the two types of perspectival seeming. In this case it could seem that we lose nothing if we reduce the perspectival ontological seeming to a sub-species of the genealogical type. Mistaking a painting of a man as a real man seems to be structurally identical to mistaking a man for a tree. More so, depending on how we understand the elements that make up deception, we could turn the tables, and construe all genealogical seeming as following the fundamental structure of the ontological one. An argument of why I believe this distinction should be maintained would require a research into the causes and elements that make up seeming which is beyond the scope of this article. As it stands for now, perspectival ontological seeming can do for us what it did for Plato, that is, offer us an analogy by which to understand the radical type.

sitting. Closer to Plato's concerns, we can think of taking an unjust act as being just (as an image of Justice), an impious act as being pious, a virtuous way of life as being vicious. These errors can be caused either by our mediated and imperfect access to phenomena, in which case they are merely perspectival, or by using corrupted models, in which case they are radical. By contrast to the radical ontological seeming that takes a just or an unjust particular to stand for Justice itself, genealogical seeming amounts to identifying the act *as* just, as an image of Justice, when it is not so.

The ability to make a correct genealogy, i.e., to say to what model an image belongs to, is dependent on a prior knowledge of the model itself. The assumption that you cannot correctly identify the instances or images of a something if you do not know that thing in itself (an assumption specific to the Socratic dialogues) comes into play in the *Republic*, when Socrates expresses the condition of the philosopher who returns to the cave: "And because you've seen the truth about fine, just, and good things, you'll know each image for what it is and also that of which it is the image." (*Republic*, 520c-d).

In the same vein, when talking about the true meaning of musical and poetical education, Socrates asks "isn't it also true that if there are images of letters reflected in mirrors or water, we won't know them until we know the letters themselves?" (402b-c). He then goes on to say that no one can claim to be educated in these arts unless he has knowledge of the virtues and vices that manifest through them "and see them in the things in which they are, both themselves and their images [...]" (402c5-6).

The analogy between the blind and the ignorant is brought up in relation to the Guardians for the same reasons. They are to look, in the manner of painters, to the true models and to establish and preserve the conventions, that, as their images, reflect them the best:

Do you think, then, that there's any difference between the blind and those who are really deprived of the knowledge of each thing that is? The latter have no clear model in their souls, and so they cannot – in the manner of painters – look to what is most true, make constant reference to it, and study it as exactly as possible. Hence, they cannot establish here on earth conventions about what

is fine or just or good, when they need to be established, or guard and preserve them, once they have been established. (484c6-d3)

We can see instances of this applied philosophical knowledge both when Socrates distinguishes between “true falsehood” that resides in the soul, and falsehood in words, which he describes as “an image of it that comes into being after it and is not a pure falsehood” (*Republic*, 382b-c), and also when true justice, as the proper organization of the souls’ parts, is distinguished from “the principle that it is right for someone who is by nature a cobbler to practice cobblery and nothing else, for the carpenter to practice carpentry, and the same for the others”, of which he states that it “is a sort of image of justice” (443c).

Thus, we can see how the knowledge of Forms guarantees not only the bringing to light of the images’ character as mere images, that is, ontological knowledge, but also the unveiling of their identity, or, in other words, genealogical knowledge. Coming back to *Republic* (520c-d), the double knowledge that the philosopher possess regards the ability to know each “image for what it is”, i.e., merely an image, a generated and derivative being, but also “that of which it is the image of,”⁵⁰ thus genealogical or practical knowledge.

2.a *Perspectival genealogical seeming*

This type of genealogical seeming is due to the distortions that the medium in which the image manifests effects upon the image. We can think of optical illusions, like those that make the straight stick appear bent (*Republic*, 602c), or of great distances that can make the disproportionate statue appear proportionate (*Sophist*, 236). Language itself is a medium, and as a medium it can effect changes to the way the phenomena that are manifested through it come out on the other side; we can think here of rhetorical devices of all sorts, that make the weak

⁵⁰ One interesting thing is that not all images are equally hard to identify. In the *Phaedrus* (250b-c), Socrates says that while Beauty shines through its image, it makes a sensible appearance. On the other hand, the images of Forms like Virtue or Justice, which are more abstract and non-sensible, are very hard to identify.

argument appear strong (*Apology*, 18c), eristic tricks that create mere verbal contradictions (*Republic*, 454a), or poetical devices such as meter, rhythm, and harmony who charm the soul (*Republic*, 413, 601a). All these can be seen as analogous means by which to create illusions in the medium of *logos*, comparable to the perspectival tricks in the medium of sight. Socrates highlights the persuasive effects that poetical devices have on our judgment:

So great is the natural charm of these things — that he speaks with meter, rhythm, and harmony, for if you strip a poet's works of their musical colorings and take them by themselves, I think you know what be they look like. You've surely seen them. [...] Don't they resemble the faces of young boys who are neither fine nor beautiful after the bloom of youth has left them? (*Republic*, 601a-b).

In short, in all this cases the medium effects cosmetic modifications, so that the phenomenon reflected in it resembles another model than the genealogically proper one: the stick appears bent (an image of Bentness) when it is not so, the act appears virtuous (an image of Virtue) when in fact it is not.

2.b *Radical genealogical seeming*

If the perspectival genealogical seeming was caused by a distortion of the image by the medium, in this case the seeming is caused by the fact that one uses an improper model or criterion for identifying images. We can, nonetheless, in a formally correct way, connect an image to its model, but if the model is badly constructed, then we are going to make only a seemingly true genealogy. This would be the case whenever one would correctly identify an act as being pious according to a certain understanding of what is pious, but would get the definition all wrong.

Before moving on, I should clarify the difference between the ontological and genealogical type of radical seeming. One way of putting it is that the ontological seeming describes the state of one's soul in terms of *hexis*, as affected by radical ontological seeming, while the genealogical one is

a description of the same state seen from the actuality of knowing, and refers to the effect that affection or *hexis* has on knowing.⁵¹ The first describes the cave allegory's prisoners' unawareness of the shadow nature of their reality, the second the effect this unawareness has on the way they judge something to be this or that.

In order to keep close to Plato's own concerns, I will present the notion of radical genealogical seeming through its heuristic causes. Heuristic genealogical seeming refers to an act of teaching by which a model is described by way of images not proper to it. This is one way someone ends up with bad definitions of models, and it's the one on which we will focus here. For example, we can think of a situation where one would describe Socrates to another person who does not know him, using characteristics that are not his own, as having long blond hair, and a sharp nose, for example. Or, more in tune with Plato's concerns, we can think of a false teacher of virtue who describes and in effect teaches what virtue is, in ways not proper to its character. It is thus a matter of using a corrupt model to identify particular instances of it, which in turn are used in a pedagogical manner to describe and instill that unsound model in the student.

I believe that it is on heuristic grounds that Plato launches his attack on the poets and the way they represent the Gods as ever-changing and deceitful in the *Republic* (379-386). Socrates compares their accounts of the gods to the works of a bad painter: "When a story gives a bad image of what the gods and heroes are like, the way a painter does whose picture is not at all like the things he's trying to paint" (377e). Following such bad descriptions, a corrupt model will be formed inside people's minds that will lead them to use it for bad genealogical practices: identifying which characteristic or behavior is godly, divine, or not, basing such identification on a wholly corrupted criterion.

⁵¹ By analogy, we say of an eye that it has myopia by looking at its inner structure, but we can also call someone's vision myopic. In the latter case we do so either for a) referring to one way of unclear vision – the myopic type –, or b) for the purpose of indicating the structure of the eye as a cause for the unclarity of vision. In this analogy, the ontological stands for the myopia of the eye, while the genealogical refers to a) the unclarity of the vision that derives from it, while at the same time it can be used to point out to b) its structural cause, the configuration of the eye.

Referring to poetical images Socrates states: "All such poetry is likely to distort the thought of anyone who hears it, unless he has the knowledge of what it is really like, as a drug to counteract it" (595b).

In the same way that only the one who knows the truth about some event first hand can have a sure way to identify false accounts about it, so it is in that only an unmediated contact with the Form offers someone the possibility of being uncorrupted by ignorant or deceitful accounts.

Throughout the *Republic*, the art of measurement ought to hold the key for verifying genealogies and dissipating mere seeming. Yet, measurement implies that we take the measure of the model first, going past the image, to the thing itself. Only after this procedure is finalized, i.e., only after we get the measure of the model,⁵² are we able to measure each image's claim of being of this or that model, and decide whether it is justified or not.

While there is a true danger that the average Athenian will be misled by perspectival genealogical seeming about vital things, such as through the tricks of rhetoric, there is a limit to sophist or the rhetor's power. As long as the discourses refer to one's line of work, where he has experience with how things really are, perspectival seeming loses its power, and the heuristic one is simply out of the question.⁵³ But how will one protect himself from heuristic genealogical seeming about more abstract things like virtue, justice and the like? The notions of what virtue or justice is has been instilled in them by the poets from a young age using, in Plato's view, untrue images. In other words, how does one come to find out that the models he uses in identifying what is virtuous and just are themselves false? By proposing an answer to this question, we can get a glimpse at how these types of seeming intertwine.

I propose that for Plato heuristic genealogical seeming about abstract notions has as its fundamental origin radical ontological seeming. Bad models, or bad descriptions of models, are created because they derive from an unreflective total reliance on particular instances of "F" as paradigms for what it means to be "F".

⁵² See Deleuze (1983) for the position that the myth usually plays this role for Plato in the dialogues.

⁵³ Socrates makes Polus concede the point that the rhetor can only convince the ignorant that he is a better medic than the actual medic, but not the knowledgeable in the art of medicine (*Gorgias*, 459b).

If we accept that heuristic genealogical seeming has as its essential source radical ontological seeming, then there could be only two ways out of it according to Plato. The first one would suppose placing trust in the images produced by the philosopher. This is the attitude expected of the auxiliaries (*Republic*, 414b). Analogously, the producers of Book X must place their trust in the advice of the users (601d-602). Yet how could one really be sure that he is following a truly wise person and not just a fraud, a sophist?⁵⁴ The second way is that of the philosopher, and it implies arriving on your own at the model. This is described by Plato as: “whenever someone tries through argument and apart from all sense perceptions to find the being itself of each thing and doesn’t give up until he grasps the good itself with understanding itself” (*Republic*, 532a-b), thereby identifying this process as dialectic.

Conclusion

As I hope to have shown, seeming is no straightforward, univocal notion for Plato. The differences I have argued for here are nowhere explicitly distinguished in the dialogues. This, of course, does not mean that they are not at play. I believe that by reading Plato with these distinctions in mind one can benefit from a ground from where to interrogate the text in a more systematic fashion whenever he comes across seeming or its cognates. Many times, the reason we feel a sense of confusion regarding a passage and are unable to tackle it directly lies in the fact that we lack the conceptual ground from where to ask questions that would, if not dispel the confusion, at least articulate it as a problem. As is often the case with philosophical research, my goal here was not primarily to provide answers for any questions or problems, but to provide a ground for asking questions. There are many passages that when read without these distinctions in mind can seem simply baffling. How can there be something “truer” than something else? What are we to make of the fact

⁵⁴ Cf. Notomi’s (1999) idea that in order to be able to identify the sophist, one must do so by philosophizing, and consequently by becoming a philosopher in the process.

that just things also appear unjust? Is Plato somehow a relativist all of the sudden? Or what sense does it make to talk about “fake pleasure”?

Keeping in mind the distinctions that I have provided here we can ask in a more systematic way what Plato has in mind whenever he talks about seeming:

- a) Is he pointing to a radical ontological seeming? Is he trying to say that the deception consists in believing that particulars constitute the ultimate reality, and that they, instead of the Forms, are invested with the function of providing a criterion for what is real, and for what is true?
- b) Is he concerned with a perspectival type of seeming, where deception arises from something that interferes with our access to phenomena? Is he referring here to optical effects and illusion, rhetorical devices that charm the soul, and other types of what he calls “magic tricks”?
- c) Is he thematizing radical genealogical seeming and deception that arises from being in possession of notions or models that, upon elenctic trial, prove to be unsound and self-contradictory? Or is he concerned with the heuristic side of genealogical seeming, the imparting of crooked models by way of improper images?

In this way, we are provided with a lot more interpretative room when trying to figure out what Plato is aiming at when he makes a statement involving deception. Let’s take the proposition “Callicles’ act only seems to be just” as a sample case. It can be interpreted along the lines of a), as stating that it is not true of the act that it is true justice, meaning that it is not the Form of Justice. This has no bearing whatsoever on whether the act is actually just or not, in the sense of it being an image of justice. What it denies is the act’s being what Justice is, but not whether it is just or not. We can also read it as b), a problem of improper access to the phenomenon. On this reading it is denied that the act is truly an image of justice. The reason we thought it was stems either from the fact that we knew too little of the situation, and “saw” it only from where it appeared just, or that we were charmed by some discourse that made it appear so. Lastly, c) we could read this line as denying that the act is an image of justice, but in this case the accent is placed not on our access to the act, but on our criteria from which we

access something as just, or our criteria of predication. In this case it is a problem of bad models, which need to be tested on their own through dialectic means.

I am in no way trying to suggest that we can find these questions as separated thematical inquiries. Rather, most of the time they are intertwined, either in the way of illustrating each other, or as constituting interconnected moments of each other, where one presupposes and anticipates the other one. As such, we cannot expect to find these senses at work as different autonomous themes of inquiry. Rather, the sole purpose of these distinctions is to highlight the different senses that Plato relies on whenever he makes a case about what is essentially a whole, unitary concern. By asking questions like "What is virtue?" Socrates in effect asks: "By reference to what do you make your genealogies of virtue?". The model is thus brought to light from its unreflective use and tested for cracks. Seeing that the model is full of cracks, though, is just half the journey. Seeing that by virtue of which you can see the cracks brings one's soul to its proper home through *anamnesis*. It is only thus that the most radical seeming is unveiled.

REFERENCES

- Allen, Reginald E. "Participation and Predication in Plato's Middle Dialogues." *Philosophical Review* 69, no. 2 (1960): 147-164.
- Benardete, Seth. *The Being of the Beautiful: Plato's Theaetetus, Sophist and Statesman*. Chicago: University of Chicago Press, 1984.
- Deleuze, Gilles. "Plato and the Simulacrum." Translated by Rosalind Krauss. *October* 27 (Winter 1983): 45-56.
- Fine, Gail. *On Ideas. Aristotle's Criticism of Plato's Theory of Forms*. New York: Oxford University Press, 2004.
- Geach, Peter T. "Plato's *Euthyphro*: An Analysis and Commentary." *Monist* 50, no. 3 (1966): 369-382.
- Heidegger, Martin. *Plato's Sophist*. Bloomington and Indianapolis: Indiana University Press, 1997.
- Heidegger, Martin. *The Essence of Truth. On Plato's Cave Allegory and Theaetetus*. London: Continuum Publishing Group, 2002.

- Lee, Edward N. "On the Metaphysics of the Image in Plato's *Timaeus*." *Monist* 50, no. 3 (1964): 341-368.
- McPherran, Mark L. "Plato's Particulars." *Southern Journal of Philosophy* 26 (1988): 527-553.
- Moss, Jessica. "Pleasure and Illusion in Plato." *Philosophy and Phenomenological Research* 72, no. 3 (2006): 503-35.
- Moss, Jessica. "What is Imitative Poetry and Why is it Bad?" *The Cambridge Companion to Plato's Republic*. Edited by G.R.F. Ferrari. Cambridge: Cambridge University Press, 2007.
- Moss, Jessica. "Appearances and calculations: Plato's division of the soul." *Oxford Studies in Ancient Philosophy* 34 (2008): 35-68.
- Moss, Jessica. "Plato's Appearance-Assent Account of Belief." *Proceedings of the Aristotelian Society* 114 (2014): 213-238.
- Nehamas, Alexander. "Plato on the Imperfection of the Sensible World." *American Philosophical Quarterly* 12, no. 2 (1975): 105-117.
- Nehamas, Alexander. "Participation and Predication in Plato's Later Thought." *Review of Metaphysics* 36, no. 2 (1982): 343-374.
- Notomi, Noburu. *The Unity of Plato's Sophist*. Cambridge: Cambridge University Press, 1999.
- Patterson, Richard. *Image and Reality in Plato's Metaphysics*. Indianapolis: Hackett Publishing Company, 1985.
- Platon. *Complete Works*. Edited by John M. Cooper and D. S. Hutchinson. Indianapolis: Hackett Publishing Company, 1997.
- Puric, Octavian. "The Platonic Receptacle: between Pure Mediality and Determining Cause." *Romanian Journal of Analytic Philosophy* 11, no. 1 (2017): 21-50.
- Silverman, Allan. "Plato on Phantasia." *Classical Antiquity* 10, no. 1 (1991): 123-47.
- Smith, Nicholas. D. "Plato on the Power of Ignorance." *Oxford Studies in Ancient Philosophy*, supplementary volume 2012: 51-73.

FREEDOM VS. ETERNALISM: SOME OBJECTIONS TO THE INSIDE OUT PERSPECTIVE

BOGDAN DUMITRESCU¹

Abstract: The free will problem has traditionally been viewed as an incompatibility between the concept of freedom and the concept of determinism. This paper is concerned with a slightly different framing of the problem: with the compatibility between free will and the metaphysics of time.

Carl Hoefer, in his 2002 article "*Freedom from the Inside Out*" has argued that the source of the free will problem is our unconscious assumption of the A-theory of time. He also argued that if we adopt a B-theory of time and imagine our actions from a static block universe perspective, then freedom would be saved. He argues that this is the case, because bidirectional determinism in the static block does not privilege past to future determination.

My aim in this paper is to present two new objections to Hoefer's view. Firstly, I argue that his description of the A-series is problematic and does not help him establish that the A-series is the source of the free will problem. Secondly, I argue that his theory is susceptible to the threat of ontological fatalism and that this is in conflict with freedom understood as the ability to do otherwise

Keywords: free will; determinism; eternalism; A- and B-theories; block universe; metaphysics of time

1. Introduction

Free will is often believed to be in conflict with the thesis of determinism: the idea that a past state of the world in conjunction with the laws of nature entails one single possible future. If freedom is the ability to act

¹ Bogdan Dumitrescu is a doctoral student in the Department of Theoretical Philosophy at the University of Bucharest. This paper is part of his research on free will and its relationship with the metaphysics of time. Contact: bogdan.dumitrescu@drd.unibuc.ro

otherwise than how in fact I've acted, then it seems that I cannot be free in a deterministic world. Traditionally, there have been two notable positions regarding the tension between the two concepts. One can be a compatibilist and argue that free will and determinism are actually compatible and can co-exist together in the same world, or one can be an incompatibilist and argue that the two concepts cannot co-exist together.

I am concerned with a different approach to the problem of free will in this paper: the relation between free will and the metaphysics of time. Carl Hoefer in his article "*Freedom from the Inside Out*" (2002) has defended a compatibilist account of freedom that I shall refer to as the Inside Out Perspective. This proposal seeks to make free will compatible with determinism by adopting a different conception of time than the one we unconsciously assume.

The aim of this paper is to present a number of objections to the Inside Out Perspective. Two of them belong to Jason Brennan (2007) and the other two belong to me. The main objection I raise is that Hoefer's theory is susceptible to a certain kind of fatalism: **ontological** fatalism. Why this matters for the Inside Out Perspective will become clear in the later sections of the paper.

To Hoefer, the free will problem arises because of a tension between determinism and our common sense view of time. Using McTaggart's distinction between **A-series time** and **B-series time**, Hoefer argues that we unconsciously assume an A-series view of time, meaning we assume that the past is fixed, the present is a flowing instant that moves through time and that the future is indeterminate. Since this view of time privileges a past→future determination, we are led to believe that a past time slice in conjunction with the laws of nature can entail only one single possible future (Hoefer 2002, 206).

Therefore, in order to make free will and determinism compatible we need to change our conception of time. Hoefer considers that the A-series' rival, the B-series would be the suitable alternative. He invites us to adopt a static block universe perspective, a model of spacetime inspired by Minkowski that consists of three spatial dimensions and one temporal dimension. This temporal dimension is the B-series, an ordering of events in time that retains the temporal direction present in the A-series, but loses the ontological distinction between past, present

and future. In a B-theory of time, we deal only with static and permanent relations between events. These relations between events are *earlier than*, *simultaneous with* and *later than*. All events are thus ontologically the same in the static block. There is no difference in “realness” between events that are “future” with respect to me and events that are “past” to me.

Hoefer proceeds to argue that within this block universe, we are able to view our actions as **not** being determined by events in the past in conjunction with the laws, but by our own volitions. This is possible because the B-theoretic block universe allows for bidirectional determinism. This means that past→future determination is no longer privileged. Future→past determination is also possible and, more importantly, determination from the *inside-out* is also acceptable. My actions within the block can thus be viewed as partially determining both earlier time slices and later time slices relative to me. For Hoefer, it seems, the fact that one direction of determination is no longer privileged in a block universe seems to be sufficient to make freedom possible. The proposal seems to make free will compatible with bidirectional determinism. However, as I shall argue, the Inside Out Perspective is an eternalist theory and does not seem to escape the threat of ontological fatalism.

The plan of this paper is as follows: In the second section I will present the temporal assumptions of two incompatibilist arguments (the Forking Road Argument and the Consequence Argument) in order to demonstrate the relevance of the metaphysics of time to the free will debate.

In the third section I will present Carl Hoefer’s compatibilist proposal. In the fourth section I will present two lines of criticism that have been put forward by Jason Brennan. The first involves the asymmetry of causation and its threat to free will that is left unaddressed in Hoefer’s article, and the second involves the uncertain ontological status of actions within the block.

Finally, in the fifth section, I will present two of my own objections to Hoefer’s view. Firstly, I will argue that his characterization of the A-series is too specific and does not help establish his conclusion that the A-series is responsible for the tension between free will and determinism. Secondly, I will show why ontological fatalism is a threat to the Inside Out Perspective defended by Hoefer. This kind of fatalism (coined by Joseph Diekemper) is the type that is entailed by eternalist ontology.

2. Free will and its relation to time: Two incompatibilist arguments and their temporal assumptions

There is a common argument for incompatibilism that is typically known as the Forking Road Argument. This is the idea that whenever we make a choice we are like a traveler that chooses one of the alternate routes from a forking road. The alternate roads stand for possible futures and the singular road behind us is analogous to the past. This picture shows us that if determinism is true (the thesis that the actual past in conjunction with the laws of nature entails one single possible future), then there cannot be multiple alternate routes to choose from. The past plus the laws entails only one road and that's it. If indeterminism holds true, then there is such a thing as a set of alternatives to choose from when we act. If I think of baking a cake, then I could choose the road containing my baking of the cake or the road which does not contain my baking of the cake.

What is this argument's conclusion? The conclusion is that determinism and freedom are incompatible concepts, because freedom presupposes the existence of alternatives. A common definition of free will views it as the ability to do otherwise (or dual ability). According to Peter van Inwagen, I freely act if and only if I have the ability to act otherwise than I actually do. A denial of free will would be to say that what I *can* do and what I *do* coincide (van Inwagen 1975, 188).

However, Kadri Vihvelin criticizes this common argument for incompatibilism. She says:

But several assumptions have been smuggled into this analogy: assumptions about time and causation and assumptions about possibility. The assumptions about time and causation needed to make the argument work include the following: that we "move" through time in something like the way we move down a road: that our movement is necessarily in one direction only, from past to future; that the past is *necessarily* "fixed" or beyond our control in some way that the future is not. These assumptions are all controversial; on some theories of time and causation (the four-dimensionalist or eternalist theory of time, a theory of causation

that doesn't deny the possibility of time travel and backward causation), they are all false. (Vihvelin 2013, 126-127)

Thus, this incompatibilist argument carries with it some metaphysical assumptions. What is of concern in this article are its assumptions on time. As Vihvelin rightly remarks, the Forking Road Argument presupposes that we as agents move through time like we move on a road, that the direction of time is past to future and that the past is fixed. These assumptions are typically the theses of the *A-theory* of time. Although not all A-theories of time share all three assumptions, it is clear that they do share the assumption that time is flowing. There is an objective privileged present that moves through time and is the explanation for why it passes.

The most popular A-theories are the following:

Presentism: The view that only present entities exist and no non-present entities exist. The past and the future are, thus, unreal according to presentists.

Growing Block Theory: The view that the past and the present are real, but the future is unreal. The passage of time here is simply the addition of new slices of existence onto the block as the present moves further and further.

The Branching Theory: The view that the past and the present are real and the future consists of multiple branching courses of events. The possibility of each of these non-actual branching futures is entailed by the actual past and the laws of nature. One must note that the branching theory is clearly an indeterminist view and actually entails an open future. Presentism and the growing block are compatible with an open future, but don't necessarily entail one (Miller 2005, 198).

The Moving Spotlight Theory: The view that time is a four-dimensional block universe of events onto which a privileged present is added. On this view, all events exist eternally, however

the distinction between past and future is kept due to the add-on of an objective present that “shines” on the line of temporal events.²

If one were to compare these A-theories, she or he would see that the Forking Road Argument works best on the branching theory or on the growing block theory. As Miller (2005) argued, the growing block theory is compatible with an open future and is also compatible with a closed future. It does not predispose us to accept an open or a closed future. However, the branching model explicitly commits us to an open future with multiple alternatives. Therefore, it seems that the validity of the Forking Road Argument rests on the assumption that either the branching theory or the growing block theory is true. This is an obvious blow to the strength of the argument itself. An eternalist theorist (which I happen to be) can reject this argument right from the start.

What, then, is eternalism?³ The A-theory is contrasted with its rival, the *B-theory* of time. This is a static theory of time in which there is no objective becoming of events, no passage of time and no objective distinction between past, present and future. The present is no longer objective, but merely perspectival. The flowing “now” is reduced to a simple indexical. On an A-theory, events have different properties such as pastness, presentness or futurity. On a B-theory, no such properties exist. All events are equally real and they are distinguished with respect to their relations with each other. These static relations are *earlier than*, *simultaneous with* and *later than*. In contrast, the properties of events on an A-theory are ever-changing. As John Ellis McTaggart put it (the philosopher who coined the distinction between A-series time and B-series time):

Positions in time, as time appears to us *prima facie*, are distinguished in two ways. Each position is Earlier than some, and Later than some,

² For defenses of presentism, see Bigelow (1996) and Markosian (2004). For defenses of the growing block, see Broad (1923) and Tooley (1997). For defenses of the branching theory, see Belnap (1992) and Belnap and Green (1994). Finally, for a defense of the moving spotlight theory, see Skow (2009).

³ For presentations and overviews of the issues with eternalism, see Wasserman (2018), Miller (2013) and Le Poidevin (2013).

of the other positions. And each position is either Past, Present, or Future. The distinctions of the former class are permanent, while those of the latter are not. If M is ever earlier than N, it is always earlier. But an event, which is now present, was future and will be past. (McTaggart 1908, 456)

Thus, A-theoretic change (or *genuine* change, as McTaggart calls it) is the change of an event from being future to being present and then to being past. B-theoretic change is supposed to be the change in an object's properties at a time. An object can have a set of properties in an earlier state and then have a different set of properties in a later state.⁴

Perhaps what the A-theory and B-theory have in common is the thesis that time has a direction and, of course, an ordering of events. As remarked by Matt Farr, McTaggart's A-series has a classification of events into past, present and future, a direction of time and an order. The B-series series eliminates this classification and retains the directionality (represented by the *earlier-later* relation) and the order of time. Of course, there is the less discussed C-series which retains only an ordering of events, but this series does not concern us in this article (Farr 2012, 87-88).

What can we gather from all this? We can observe that the Forking Road Argument for incompatibilism fails to be convincing for philosophers that are B-theorists, because they reject the argument's assumptions on time. However, the relation between the metaphysics of time and the problem of free will reaches deeper than this. Let's consider the most popular and, arguably, the most powerful argument for incompatibilism: the Consequence Argument. Although this particular argument has been expressed in different forms, its clearest and most notable formulation is that given by Peter van Inwagen. In his famous book, *An Essay on Free Will* (1983) he states it as follows:

If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us

⁴ For an introduction into the debates between A-theorists and B-theorists and their differing views on tense and change, see Le Poidevin, Robin (1998) *Questions of Time and Tense*, Oxford University Press, Chapter 1.

what went on before we were born, and neither is it up to us what the laws of nature are. Therefore the consequences of these things (including our present acts) are not up to us. (van Inwagen 1983, 16)

If the consequences of the past in conjunction with the laws of nature are our present actions, then our present actions are not up to us. If our present actions are not up to us, then we do not have free will. Therefore, if determinism is true, then we do not have free will. If I have the ability to act otherwise than how in fact I do, then I must be able to make the conjunction of the laws with the actual past to be false. However, if I can make the conjunction (let's call it P & L) false, then I can make either P false or L false or both. Typically, we do not think that our present actions can render P false, because the past is remote, beyond our control and *fixed*. Furthermore, the idea of backward causation (present actions having a causal influence on past events) is met with a lot of resistance by our intuitions.

As for rendering L false, it's quite doubtful that we as mere human beings are capable of rendering false a law of nature. However, David Lewis had a lot to say on this point in response to van Inwagen's argument. It's ridiculous to claim that we are able to break the laws by simply acting freely, but it may not be ridiculous to claim that I have the ability to do something which, if I had done, would have broken a law of nature (Lewis 1981, 123). The tactic used here is to attack the LAW premise of the Consequence Argument: It is not up to us what the laws of nature are.⁵

What of the PAST premise? The idea that the past is not up to us seems to belong to common sense. However, one must be more precise here. What does it mean to say that the past is not up to me? It may mean that I cannot change the past, but one could say that it wouldn't make sense to claim that I could change the future either. A simple

⁵ Another way to reject the LAW premise is to argue for a Humean conception of laws in which the laws of nature are simply contingent generalizations of how the fundamental events or particles in the world move. The laws here do not have the power to necessitate or dictate the way in which objects or events behave, but are simply patterns of those behaviors. For more on this view, one could read Beebe and Mele (2002), Esfeld (2021), Loewer (2012).

tautology is that the future will be what it will be. This does not imply that my actions are causally powerless towards the future. There is a perfectly reasonable sense in which my actions causally affect the future. So this would mean that there is a distinction to be made between *affecting* and *changing* and an observation to be made: that affecting does not imply changing (Le Poidevin 2013, 537).

There might then be a reasonable sense in which the past is affected by my actions or even a sense in which it is determined by my actions. Let's consider the concept of determinism. One could distinguish two kinds of formulations of the thesis of determinism. Consider the following two:

The world is deterministic if and only if, given a specified way things are at a time T, together with the laws of nature, jointly logico-mathematically determine a single possible future of the world. (Hofer 2004, 101)

Now consider:

(...) the propositions stating the laws of nature and the propositions describing the state of the world at an arbitrary time t (i.e. the propositions describing the initial conditions) entail the propositions describing the state of the world at any other time. (Esfeld 2019, 78)

The first kind of formulation I shall call the Tensed Definition of Determinism and the second I shall call the Tenseless Definition of Determinism. The first is usually stated as follows:

Tensed Definition Of Determinism: (Actual Past & Actual Laws of Nature) → One possible Present (or Future) course of events and only one.

This is typically the most common statement of the thesis. We are often told that the past in conjunction with the laws of nature entail our present actions or our future. However, this formulation contains temporal terms and assumes that past to future determination is privileged. A tenseless version of determinism does no such thing. We can state it as follows:

Tenseless Definition of Determinism: (State of the world A & Actual Laws of Nature) → State of the world B.

Or in other words, for every X, if X is a state of the world, then in conjunction with the actual laws, X logically entails any state of the world. By “state of the world” I may mean a conjunctive statement in which I enumerate the simultaneous events that occur at an instant in time (or on a time slice). Peter van Inwagen uses this kind of definition and claims that deterministic relations are actually entailments between propositions (van Inwagen 1975, 186-190). Michael Esfeld also holds a similar view as it is evident from this passage: “Thus formulated, it is clear that determinism in science is—only—about entailment relations among propositions.” (Esfeld 2019, 78)

This tenseless version of determinism does not, of course, assume that a certain direction of determination is privileged and, as a result, it favors a bidirectional relation between states of the world. My actions in the present both logically entail later states and also logically entail earlier states. If this is so and the relation of determination goes both ways rather than one way only, then the Consequence Argument’s PAST Premise is false. This is the move that Carl Hoefer makes in his interesting article *Freedom from the Inside Out*. In the following section, I will present Hoefer’s Inside Out Perspective, in which he attempts to make free will compatible with the deterministic physics inside a four-dimensional static block universe.

3. The Inside Out Perspective

Carl Hoefer’s Inside Out Perspective is a compatibilist account of free will, but this is not the usual compatibilism that is generally being talked about in the literature. Hoefer attempts to make free will compatible with deterministic physics by appealing to a certain theory of time: the static block eternalist theory. He writes:

The challenge to free will from determinism has not come from the physics, but rather from the unholy marriage of deterministic physics with our A-series view of time. (Hoefer 2002, 206)

According to him, we unconsciously assume a common sense view of time and that time is the A-series. Typically, he says, we think of the past as being fixed, the present as a moving instant and the future as being open to possibilities. I argue that this characterization is a bit problematic and I will explain why I believe so in a later section. However, Hofer states:

The worry we have is that a *past* slice (...) determines our actions *now*. We never think of a *now*-slice (including the voluntary actions we perform now) determining what happened in the past. Why not? (Hofer 2002, 206)

This is why the problem of free will arises. The culprit isn't actually determinism *per se*, but the assumption that past to future determination is privileged. This assumption is present, because we have this A-theoretic intuition of time. Then how could someone try to make free will compatible with determinism again? Naturally, we could try harmonizing free will with a different conception of time, and this is exactly what Hofer attempts to do.

Given a proper understanding of time, we will see that freedom and determinism are compatible in a much more robust sense than has ever been thought possible. (Hofer 2002, 202-203)

The saviour theory would be the B-theory of time. Hofer invites us to view time as a four-dimensional static block universe inspired by Einstein-Minkowski spacetime that contains all temporal events of the world. The block has three spatial dimensions and one temporal dimension which is thought to be the B-series time. As on any eternalist theory, there is no single event or time slice within the block that can be identified as the "flowing now" (Hofer 2002, 203-205). Past, present and future are A-theoretic terms and have no place in a B-theory. Events in the block are all equally real. The events of today are not ontologically different from the events of one billion years ago. As stated in the previous section, any B-theory must preserve a direction of time. This direction is given by the earlier-later relation between events within the

block and can also work as a later-earlier relation since tenseless determinism is bidirectional.

As McTaggart stated, these relations are permanent and static. One might also note that statements about these B-relations don't change their truth values at different times. For example, it will always be **true** that World War II is later than World War I and it will always be **false** that the Romanian Revolution of 1989 is earlier than World War II.⁶

Given such a view of time, how could one argue that free will is possible under it? If there is no ontological distinction between events and all exist indiscriminately, then the later time slices seem to be fixed and beyond my control. I lack control over the future just as I lack control over the past. Eternalism seems to entail fatalism: the thesis that no matter what we do, the way the future will be is unavoidable. Hofer doesn't share this thought. He writes:

The very "timelessness" of the 4-D block (in an A-series sense) leaves us free to reject the customary view that past events determine present choices. From the B series perspective, there is no reason to think of past→future determination as more important or real than future→past determination. And, even more to the point, one can equally view a set of events in the *middle* as determiners of both past and future events. (Hofer 2002, 205)

Furthermore, he states:

The idea of freedom from the inside out is this: we are perfectly justified in viewing our own actions not as determined by the past, nor as determined by the future, but rather as simply determined (to the extent that this word sensibly applies) by ourselves, by our own wills. (...) Instead, we can view our own actions, qua physical events, as primary explainers, determining – in a very partial way – physical events outside ourselves to the past and future of our actions, in the block. (Hofer 2002, 207)

⁶ For an introduction to the debate between tense and tenseless semantics and for an overview of the issues in the philosophy of time, see Fischer (2016).

Thus, if I am an agent within the block, my actions partially determine both later time slices and earlier time slices of the universe. The reason we are reluctant to accept this kind of bidirectional determinism is because we think it implies absurd backward causation and wish to avoid it. The asymmetry of causation shows us that causes precede their effects and effects follow their causes. The direction of causation is only one way and, thus, our actions can causally influence only the future. However, this need not be so in the case of determinism. If we accept bidirectional determinism, then our actions within the block can influence both earlier and later time slices.

Hoefer argues that we need not worry about backward causation. He makes it clear that deterministic relations are logical entailments and not causal relations. Our actions constrain how the earlier (or later) states can be, but they do not have causal influence over the earlier states (Hoefer 2002, 209-210). But what do these logical constraints amount to?

My typing on my keyboard at t_2 determines the later state t_3 which contains, let's say, the finishing of this current sentence. My typing at t_2 also constrains the earlier state t_1 to be in a certain way, but these are not *macro*-level constraints (events like my previous typing of sentences in this document, objects such as the keyboard on which the typing is being done or my functional computer). Hoefer's proposal is that our actions impose logical constraints on how the past⁷ is at the *micro*-level, not at the macro-level. He writes:

(...) let's assume that a human action (including the perceived surroundings of the agent's context) is a physical event type that has innumerable instantiations at the microphysical level. We assume, in other words, that there is some ill-defined and probably infinite set of microphysical-state types that are 'good enough' to count as a supervenience base for my typing 't' in the assumed context. (...) If I freely choose to type this letter, 't', the choice in its context entails that some one of this enormous micro-state types shall be, and that is all. The constraints this places on how the past

⁷ "Past" is here used as meaning "earlier time slices than the ones which contain my current actions".

should be, even (say) the past of only one minute earlier, are probably either trivial or non-existent. (Hofer 2002, 210)

And later he writes:

At the microphysical level the constraint is just that earlier microphysical states have to be logically consistent with a microstate of the correct type (i.e., one corresponding to my typing a 't') obtaining, at the time and place that it does. (Hofer 2002, 211)

My current typing logically constrains earlier states of the universe to be in a certain way microscopically, not macroscopically. Hofer does not mean to say that if I had chosen to type a different sequence of letters now,⁸ then the *macroscopic* past would have been different in perceptible ways, but that the *microscopic* past would have been different in certain ways. He notes:

I think I have freedom of the following kind: even given that the past history of the world is, macroscopically, as I (and indeed every other agent) knows it to be, I can either type the 's' or the 'z' (depending on which I choose). (Hofer 2002, 215)

To sum up, Hofer believes that the free will problem arises because of the clash between our A-series intuitions on time with deterministic physics and that the problem can be avoided if we adopt the Inside Out Perspective which assumes B-series time and the block universe model with a bidirectional determinism. This, he maintains, allows us to conceive our actions as fundamental partial determiners of both earlier and later time slices within the block. This bidirectionality of determinism, obviously, does not privilege past to future determination over future to past determination and removes the worry that a past time slice plus the laws could determine our current actions.

Worries of backward causation are also put to rest, because deterministic relations in the Inside Out Perspective are held to be logical entailments

⁸ This is an indexical use of "now", of course.

or logical constraints on how earlier or later time slices are at the microphysical level. They are not causal relations. Hoefer explains causality's unidirectionality (and the fact that we have causal control over the future and not towards the past) by appealing to other asymmetries such as the asymmetry of entropy, as B-theorists usually do (Hoefer 2002, 212).

I believe that Hoefer's arguments can be formulated as follows:

A-series Argument:

- (i) If A-series time privileges a past→future determination, then A-series is the source of the tension between determinism and free will.
- (ii) If the A-series is the source of the tension, then the A-series must be discarded in favor of a better metaphysical theory of time that supports compatibilism.

Therefore, (iii) If A-series time privileges a past→future determination, then the A-series must be discarded in favor of a better metaphysical theory of time that supports compatibilism.

B-series Argument:

- (iv) If we adopt a static block universe perspective (with B-series time), then we can also assume bidirectional determinism.
- (v) If we assume bidirectional determinism, then there is no privileged past→future determination.
- (vi) If there is no privileged past→future determination, then the conflict between determinism and free will is avoided.

Therefore, (vii) If we adopt a static block universe perspective (with B-series time), then the conflict between determinism and free will is avoided.

In the following section I shall present Jason Brennan's criticism of Hoefer's proposal and then, in the final section of this paper, I will present my own objections to the Inside Out Perspective.

4. Two criticisms of the view

4.1 *The asymmetry of causation*

Two lines of criticism have been proposed by Jason Brennan (2007) to Carl Hoefer's Inside Out Perspective. One objection concerns the very relevance of Hoefer's proposal to the problem of free will. There is the assumption that the past to future determination is the reason why determinism seems to conflict with free will. The Inside Out perspective seems to avoid this privileged direction of determination by adopting bidirectional determinism within an eternalist static block universe. My actions in the present are determined both by earlier time slices and also by later time slices. However, nothing keeps me from conceiving my own actions as being part of a time slice that determines earlier states and later states. The time slice I inhabit is as much a determiner as any other time slice.

Brennan isn't convinced by this argument. Hoefer, he argues, may succeed in showing that determinism wasn't the threat to free will, but it seems that the asymmetry of causation could very well be one. In the block universe, our actions have logical consequences towards the past and the future, but this symmetry does not hold with respect to causation though.⁹ Causation remains asymmetrical and unidirectional. Our actions have causal effects towards later time slices and not towards earlier time slices. I can finish writing this paper at t_1 and make it the case that it will be ready for submission to a journal at t_2 and this would perhaps lead to the paper being reviewed at t_3 . No such effects can occur towards earlier states. Hoefer explains this asymmetry by appealing to asymmetries in physics such as the direction of entropy given by the second law of thermodynamics, but Brennan remarks that Hoefer has not given an argument for why unidirectional causation isn't a threat to free will (Brennan 2007, 211-212).

⁹ Carl Hoefer views determinism and causation as distinct concepts. He gives arguments for this separation in his article "*Causality and Determinism: Tension or outright conflict?*" in which he states that deterministic relations are relations of entailment and causal relations are metaphysical and non-logical relations. (Hoefer 2004, 101)

One could still argue that given the fact that the direction of causation is past to future, then all my actions are causally necessitated by antecedent events. If all my actions are necessitated by antecedent events, then, one could say, I have no free will. This past to future necessitation seems to be exactly the kind of privileged determination that Hoefer thought caused the tension between free will and determinism. In other words, Brennan's question to Hoefer would be: If a privileged past→future determination is responsible for the tension between free will and determinism, then why isn't past→future **causation** a threat to free will just as much?¹⁰

Of course, given that the Inside Out Perspective is a B-theory, Hoefer could run into another problem regarding causation: the issue of whether or not causation is actually compatible with eternalism. Causation is thought of as having a past to future direction, but in B-theoretic terms, this would be the earlier-later relation. This concept, *prima facie*, doesn't seem to be incompatible with eternalism. Later states are simply causally dependent on earlier states. Time has thus a direction on the B-theory, because of the direction of causation. There does not seem to be a problem.

However, we generally think that the effect does not exist prior to its cause. Causes bring their effects into existence. When I light a match on fire, I bring forth an event into existence. Causation seems thus to imply bringing events into existence. I strike the match at t1 and at t2 it lights on fire. Prior to t2 there could have been various events in the world or conditions unknown to me that could have interfered with the bringing about of said effect. This seems to be in tension with a basic tenet of the B-theory: the idea that it is static. If a theory of time is static rather than dynamic, then existence is not time-relative. Causation understood as the bringing of events into existence seems to be incompatible with the B-theory thus. Robin Le Poidevin writes:

¹⁰ It has been pointed out to me by a reviewer that the threat to free will by universal causation could arise only if our actions have prior sufficient causes, not simply because our actions have causes. Many compatibilists do not see the fact that our actions have causes as a threat to free will at all. Brennan does not seem to provide reasons to think that all our actions have prior sufficient causes.

Consider the various connotations that causation has: we think of a cause as bringing about its effect, as bringing that effect into being, to make real what before was unreal. And if the direction from cause to effect is from earlier to later, then that, of course, implies the unreality of later times. At the time of the cause, the effect is still unreal. And that runs entirely counter to the view of (what we call) the future advocated by the B-theory. (Le Poidevin 2013, 540)

However, this problem is not mentioned by Brennan, because it is not an issue that is specific to Hofer's Inside Out Perspective, but to any B-theory in general. Brennan suggests that Hofer, in order to escape the causal asymmetry objection, should accept the symmetry of causal efficacy. This should imply that our present actions have a causal influence over earlier time slices and towards later time slices (Brennan 2007, 214). Hofer would object here, of course, because symmetric efficacy, being a causal notion, implies that we bring about effects towards the past. This is unacceptable backward causation that needs to be avoided.

Brennan argues that Hofer is faced with a **dilemma**:

- (viii) Either he accepts that unidirectional causation is a threat to free will or he accepts that it is not.
- (ix) If unidirectional causation is a threat to free will, then the Inside Out Perspective is irrelevant to the free will problem.
- (x) If unidirectional causation is not a threat to free will, then Hofer needs to argue why it is not.

Therefore, (xi) Either the Inside Out Perspective is irrelevant to the free will problem or Hofer needs to argue why unidirectional causation isn't a threat to free will (cf. Brennan 2007, 213-214).

This dilemma is perhaps a false dichotomy. Hofer may simply revise his theory and be an eliminativist with respect to causation. This position was also endorsed by Bertrand Russell, who argued that causation is a "*relic of a bygone age*", a concept that physics does not use or need (Russell 1912, 1). Since Hofer argued that A-series time concepts such as past, present or future have no place in contemporary physics (from

Einstein's theory of relativity onwards) and are unnecessary, then he could argue the same about causation, much like Russell did (Hoeyer 2002, 203). However, of course, the price of eliminating causation from a theory of time may be too high.

4.2 *An ontology of free actions*

The second line of criticism that Jason Brennan brings concerns the ontology of our free actions. What kind of things are free actions in the block universe? Of course, they are events, but what explains their occurrence? They may be fully explainable by later states or earlier states, but we must remember that Hoeyer proposes something different:

Instead, we can view our own actions, *qua* physical events, as primary explainers, determining – in a very partial way – physical events outside ourselves to the past and future of our actions, in the block. (Hoeyer 2002, 207)

Our actions are best explained by our beliefs and intentions and not by physical events outside our time slice, he argues. This seems like a very common compatibilist view: the idea that a necessary condition of a free action is being caused by its agent. This is very much like agent causation, the view that free actions must be caused by the agent's will and not be necessitated by antecedent events. However, Brennan sees a problem with this approach.

He argued that if our choices in the block universe are to be viewed as basic, fundamental determiners, then they are either brute facts or random occurrences. If they are brute facts, then they lack any explanation for their occurrence. If they are random, then, of course, our actions are not within our control. We would have as much control over our actions as we have over the outcome of a coin toss. This is particularly devastating for the Inside Out Perspective, because it leaves it open to the problems of indeterminism (Brennan 2007, 215).

Furthermore, if Hoeyer wishes to adopt agent causation into the Inside Out Perspective, then, Brennan argues, agent causation seems to

be sufficient to solve the free will problem and it would then be unclear what role the Inside Out Perspective is supposed to play (Brennan 2007, 215). Adopting agent causation seems to make Hofer's B-theory irrelevant again. However, one could argue that even if agent causation could be successfully integrated in Hofer's static block without making the Inside Out Perspective useless, then one could still not get rid of the threats of indeterminism.¹¹

This concludes the presentation of Jason Brennan's objections to Carl Hofer's Inside Out Perspective. If one were to sum up the criticisms brought forth, it would be through the following question: How exactly is the Inside Out Perspective relevant to the problem of free will? In the following and final section of this paper, I will present my own objections to this compatibilist proposal and argue for the changes that could be made in order to save some of it.

5. Further objections to the view

5.1 *The problem with the A-series argument*

In the article "Freedom from the Inside Out", we are presented with a characterization of the A-series that serves to show how our common sense view of time is responsible for the conflict between free will and determinism. Hofer writes:

First, we unconsciously assume a metaphysical picture that is A-series based and incompatible with the block universe: we think of the past as 'real', fixed or determinate, the present as 'real' (or becoming so), but the future as indeterminate or 'open'. (Hofer 2002, 206)

This characterization of the A-series, I argue, is problematic. McTaggart referred to the A-series and the B-series as possible orderings of events or positions in time, not as fully fledged metaphysical theories. The traits

¹¹ For an argument against agent causation as a viable solution to the free will problem, one could see van Inwagen (2000, 15-16).

that Hoefer enumerates are not present in all A-theories. The reality of the past and the unreality of the future, for example, are not necessary assumptions. In fact, presentism views the past as unreal and the moving spotlight theory views the future as real. Then what A-theory fits Hoefer's description best? One possible candidate would be the growing block theory.¹²

I have described the growing block theory as a metaphysical theory of time that assumes a real past, a moving present and an unreal future. The movement of the present is simply the adding of time slices to the block. C.D Broad, a proponent of the theory, writes:

It will be observed that such a theory as this accepts the reality of the present and the past, but holds that the future is simply nothing at all. Nothing has happened to the present by becoming past except that fresh slices of existence have been added to the total history of the world. The past is thus as real as the present. (Broad 1923, 66)

On this theory of time, the past→future determination is clearly favored. The trunk of the past together with the laws of nature could imply a closed fixed future. As Kristie Miller argued, the growing block is compatible with an open future, but it does not entail it (Miller 2005, 198). Thus, the growing block theory is the best candidate for Hoefer's A-series characterization for two reasons:

1. One could think of it as having a closed and fixed past, a moving present and an open future.
2. One could equally worry that, because the past→future determination is the direction of determination, the real past in conjunction with the laws of nature entails a closed and fixed future, making our free will seem illusory.

And from here on, one could argue like Hoefer, that our freedom is not the thing that is illusory, but our common sense A-series conception of time.

The problem here is that Hoefer's characterization of the A-series is too specific. It appears to only adequately apply to the growing block

¹² One could say that the branching theory fits Hoefer's description better, because it entails an open future. However, the branching theory is an indeterminist theory of time and would not help Hoefer in proving that the A-series favors past to future determination.

theory. The branching theory, on the other hand, necessarily entails an open future and is an indeterminist theory and thus cannot lead to Hoefer's A-series conclusion. Presentism seems to be compatible both with an open past and with a closed future and thus does not predispose us to privilege past to future determination.¹³ The moving spotlight theory rejects a nonexistent future by default, because it assumes the thesis of eternalism as true and simply adds a moving objective present. Therefore, Hoefer's description of the A-series seems to actually be the description of the growing block theory.

This makes his case (that A-series time is responsible for the free will problem) a bit weak. He did not manage to show that A-series is the source of the free will problem, but that a very specific theory of time is: the growing block. One natural response available to Hoefer would be to claim that the growing block is actually our common sense view of time and that is why the threat of free will from determinism is such a powerful intuition.¹⁴ However, this doesn't establish the conclusion that A-series time conflicts with free will. The branching theory with its indeterminate future clearly seems to not support it.

Hoefer could make a slightly different claim: He could say that free will is actually made more plausible within a B-theory of time than within any A-theory of time. This proposal is, obviously, not without its problems.

5.2 *The threat of ontological fatalism*

While in the previous section I raise an objection to Hoefer's perspective regarding his characterization of the A-theory, in this section I point out

¹³ Also, presentism seems to have a problem of its own regarding causality. The presentist needs to explain how causality is possible given the fact that causation is generally a relation between two non-contemporaneous events. The presentist seems to allow only for the possibility of simultaneous causation. Thus, if I talk of the past determining the future, under presentism, I would seem to be saying that a nonexistent determines another nonexistent. See Bigelow (1996) and Markosian (2004).

¹⁴ Hoefer's characterization is very similar to what Miller calls "the intuitive view of time". According to Miller, presentism, the growing block and the branching theory fit our intuitive view of time. (Miller 2008, 173) As I argued, only the growing block helps establish Hoefer's conclusion about the A-series though.

a threat to free will made possible by the adoption of the B-theory of time. In order to comment further on the Inside Out Perspective, I must introduce the distinction between **logical** fatalism and **ontological** fatalism. Fatalism is a thesis concerning human actions. It is the idea that our actions are causally inefficacious towards the future much like they are causally inefficacious towards the past. The fatalist, thus, believes that our actions are ineffective towards the future. No matter what we do in the present, we will not change what will become future.

Using Joseph Diekemper's taxonomy, **logical** fatalism is the kind of fatalism that has its source in the unrestricted application of the principle of bivalence to all declarative statements. If all statements are true or false, then future contingent statements like "Tomorrow there will be a sea-battle" (Aristotle's famous example used in *De Interpretatione*) are either true or false too. If they are already true or false at the moment of their utterance, then it's doubtful whether we could affect the future in a way in which we would change their truth values.

Eternalism seems to go hand in hand with fatalism. If eternalism is true, then for every statement about future contingent events there seems to be a fact in a later state of the world which acts as a truthmaker for that statement. If "Tomorrow there will be a sea-battle" is true, then the time slice of tomorrow must contain the event of the sea-battle already and would thus make my statement already true at the moment of utterance.

However, logical fatalism is not the type of fatalism I wish to address here. I wish to address **ontological** fatalism, the kind that follows directly from eternalism.¹⁵ Diekemper states that:

The thought here is that ontological fatalism is meant to follow directly from the nature of future events, and that this implication is independent of any implications arising from the status of propositions about those events. (Diekemper 2007, 434)

¹⁵ It could, perhaps, be argued that logical fatalism follows directly from eternalism too, but that would mean that bivalence would also follow directly from eternalism and that thesis would need a separate defense. The crucial point of distinction between logical and ontological fatalism (on my understanding of Diekemper's work) is that logical fatalism has as its source the application of logical bivalence to all declarative statements and ontological fatalism has as its source the eternal existence of future events.

If eternalism is the thesis that all events exist eternally; that there is no ontological distinction between events in terms of reality, then it seems that all events are fixed. If all events are fixed, then ontological fatalism seems to follow. Diekemper states:

If, however, we are considering the variety of ontological fatalism that is meant to follow from a temporal eternalism (whereby all events exist eternally), *and* fixity is grounded in ontology, then it is plausible that the fixity of the future both implies, *and* is a consequence of, (this variety of) fatalism. (Diekemper 2007, 436)

Thus, the Inside Out Perspective (being an eternalist theory) needs to address the issue of ontological fatalism. Put simply, the Perspective does not seem to make eternalism and free will compatible if ontological fatalism is not avoided somehow. Hofer mentioned in a footnote that Paul Horwich in his book *The Asymmetries of Time* has argued for the “correct refutation of the argument for fatalism (‘logical’ fatalism) based on the block universe” (Hofer 2002, 205). However, as noted, Horwich’s argument applies to logical fatalism and not to the fixity of eternally existing events that gives rise to ontological fatalism.

What I must add is that this objection from ontological fatalism is not an objection specific to Hofer’s proposal, but to compatibilists that hold eternalist views in general. A handful of other authors have also held similar eternalist compatibilist views regarding free will, but none seem to emphasize free will’s relationship with time more than Hofer. I will return to this issue at the end of the section.

Because of ontological fatalism, eternalism may be incompatible with a certain understanding of free will, that of freedom to do otherwise. How might this be so? This is the point at which the Forking Road Argument creeps in uninvited. Under the assumption of eternalism, there seems to be only one road in front of an agent when acting. Freedom to do otherwise seems to entail that there must be at least one other road I could take.

Libertarians often tie the concept of freedom to that of the existence of open alternatives. The existence of alternatives might be taken to imply an open future. Robert Kane, the famous libertarian, considered that “Such

a picture of an open future with forking paths – a garden of forking paths, it has been called – is essential to our understanding of free will.” (Kane 2007, 6)

More recently, Marius Backmann has also tied freedom to do otherwise with the concept of an open future. He stated:

In libertarianism, one standard criterion for freedom is the power to do otherwise: in order for a decision to be free, it must be possible to decide between at least two actually open alternatives: If I want to freely choose whether to drink red or white wine, it must be possible that the decision goes either way. In the standard reading, this implies that there are, at the instant of a decision, at least two real alternative future courses of events available and the agent can bring one of them about by his decision. (Backmann 2016, 259)

Assuming this understanding of libertarian freedom, a genuinely open future seems to be a necessary condition for freedom to do otherwise. Since an open future is incompatible with eternalist ontology, it would seem that, as a consequence, freedom to do otherwise is incompatible with eternalism. If one would accept that (i) libertarian free will implies an open future; and (ii) eternalism implies a non-open future, then one can have a strong argument for the incompatibility between eternalism and freedom to do otherwise.

This might not seem convincing at first, but I believe that upon closer inspection, this argument does carry some weight. The argument is not so different from the Forking Road Analogy that I mentioned earlier. Let’s assume eternalism, and let’s say I wish to bake a cake today. The decision to bake it occurs at 14:00 PM. At 15:00, after shopping for ingredients, I start the process of baking the cake. Then at 16:00 the cake is ready. Now, from a God’s-eye view (the view which the block theorist invites us to take), there is this causal sequence of events that is extended in spacetime. However, all these events exist *simpliciter*. They do not exist simultaneously, of course, but all of them equally exist at different locations in spacetime.

We must remember that we have assumed eternalism, which means that the forking road analogy cannot apply. The only road I have in front of me at 14:00 is that which leads to the cake at 16:00. I seem to

be perfectly able at 14:00 to refrain from baking the cake and thus preventing it from bringing it about at 16:00. However, if I could choose not to bake it, then, assuming the forking road, I should have a route before the moment of my decision that sprouts from 14:00 to 15:00 where I begin doing something else instead of baking the cake (writing this paper, for example) and to 16:00 where no cake baked by me exists. In other words, I would need to have another branch, another alternate route that is open to me before I decide at 14:00 to bake the cake or not.

But this does not seem to be the case under eternalism. We do not have the advantage of the growing block theory of time or of the branching theory here. Both these theories are compatible with an open future that contains genuine alternatives, but eternalism cannot admit of the openness of the future since there is only one actually¹⁶ existing future¹⁷ in the static block.

Another point that the eternalist compatibilist needs to address, that is closely tied to the issue of causation in a B-theory, is the fact that bringing something into existence does not seem to make sense in eternalist ontology. Because on an eternalist B-theory we have an ontological symmetry regarding events (no difference in “realness” between earlier, simultaneous or later times), nothing can be said to be brought into existence (Le Poidevin 2013, 540-541). I cannot claim that at 16:00 the cake is brought into existence unless I take objective temporal becoming seriously (or an A-series account of change).

This point was also made by Niall Shanks. If the B-theory of time is true, then I lack any existential control over the cake. The cake does not get brought into existence if it exists eternally located at a later time slice relative to me (Shanks 1994, 57). Under the assumption of eternalism, we already assume that for every X, if X is an event, then X exists eternally. This is not to say that the cake will necessarily sprout into existence by some other causal chain, if I choose not to bake it, but that my baking it is already entailed by earlier and later events.

¹⁶ I take it as a given that all events in an eternalist ontology are actual since if there is no distinction between events in terms of “realness” (the event of me typing “now” is just as real as the start of the Romanian Revolution), then all existing events must be actual.

¹⁷ Here I, of course, use a perspectival meaning of “future” which is “the sequence of events that exist later relative to me”.

The following reasoning is valid: "If I had chosen at 14:00 to not bake the cake, then no cake would exist at 16:00. But a cake at 16:00 does exist! Therefore, I did decide to bake the cake at 14:00." However, the compatibilist could rightly add that this does not show that the existence of the cake at 16:00 determines or forces me to choose to bake it at 14:00. The existence of the cake at 16:00 merely implies that I in fact chose to bake it.

The compatibilist is, of course, correct here. This does not change the issue of the compatibility between eternalism and freedom to do otherwise though. Even if ontological fatalism does not actually follow from eternalism or doesn't actually pose a threat to free will, then there would still remain this tension between the concept of freedom to do otherwise (which may require an open future) and the ontology of eternalism (which does not admit of an open future).

One could argue that I have merely shown that there may be an incompatibility between eternalism and libertarian free will and that I have not shown that ontological fatalism is a threat to free will at all. But one would need to be a bit more careful here. Eternalism is not necessarily in conflict with freedom to do otherwise just because it presupposes that all events exist, but because it necessarily presupposes that future events exist. Then ontological fatalism can follow not just from eternalist theories of time, but from any theory of time that supposes that the future is ontologically real. For example, the very unpopular (but logically possible) shrinking block theory of time (in which the present and the future are real, but *not* the past) also entails ontological fatalism in virtue of the fact that the future is real and thus fixed.

Thus, it would seem that eternalism is incompatible with freedom to do otherwise, because it entails ontological fatalism. One could still remain unconvinced of the idea that ontological fatalism is a threat at all, of course. This point may perhaps be made stronger if one considers Joseph Diekemper's (2007) usage of the distinction between logical and ontological fatalism when considering time travel scenarios. Suppose that at 13:00 PM today I am visited by my future self who is a time traveler. My future self tells me that I will decide to bake the cake at 14:00, bake it at 15:00 and have it ready at 16:00. He even gives me a detailed description of how I will do so. Now, of course, my actions will

make the proposition "I will bake the cake today" true and the backward counterfactual "Had I not decided to bake a cake, the cake would not exist at 16:00 PM today" would also be true. It is true that I will bake it *not* because I am constrained by the truth of the future propositions told by the time traveler, but because I will do so by my own will.

However, this sort of anti-fatalistic response is effective against logical fatalism, not ontological fatalism. The truth of future propositions depends on the occurrence of those future events in question; the occurrence of those future events depends on my current actions. But in ontological fatalism we are dealing with events that already occur (tenselessly) and thus seem fixed as a matter of what I will do. As Diekemper states: "we have not only abstract future truth, but concrete future existence" (Diekemper 2007, 448). We must remember that the time traveler has already experienced and done the deed of baking the cake and had it ready at 16:00. Thus, he continues:

So, the response to the ontological fatalist cannot rely upon counterfactual claims, but must rely upon counter-existence claims: something along the lines of, 'If I don't pass through Village C, then it doesn't lie along my route.' (Diekemper 2007, 448)

Maybe not even this will be convincing for the eternalist compatibilist, but consider what has to be the case if I had freedom to do otherwise assuming the road analogy and eternalism. If I would be able to choose not to bake the cake, thus diverting from the road in front of me which contains the cake at 16:00 PM, then I would have to have an alternate route to go through. This kind of branching world metaphysics that makes an open future is at odds with the ontology of eternalism, but it would seem that this is the kind of ontology that freedom to do otherwise requires.

It would thus seem that any eternalist compatibilist that claims that libertarian free will is compatible with eternalism must first show how there is no tension between the two. One obvious route would be to show somehow that freedom to do otherwise does not entail an open future. Nonetheless, I believe that the tension is real and that it must be addressed by eternalist compatibilists.

Perhaps Hoefler could think of free will not in libertarian terms, but in traditional compatibilist terms. If the freedom to do otherwise is incompatible with eternalism, then perhaps a weaker sense of freedom could suffice. One could think of freedom as the absence of constraints or coercion. This negative sense of freedom has been advocated by many compatibilists in the past and could indeed be used by Hoefler or any eternalist that assumes the block universe perspective. However, Jason Brennan could of course reply that the shift towards compatibilism or to agent causation theory would make the block universe perspective useless to the debate on free will. Further modifications would then be needed to Hoefler's theory in order to make it relevant again to the free will problem.

At the beginning of the section, I noted that the objection presented is not specific to Hoefler's theory, but that it is quite general and may apply to any eternalist that defends freedom to do otherwise within an eternalist ontology. In the last decade, a number of philosophers including Michael Esfeld (2021), Barry Loewer (forthcoming) and Jenann Ismael (2016) have offered similar defenses of free will within eternalistic frameworks. The similarity that these proposals share with Hoefler's is that all seem to reject the PAST premise of the Consequence Argument by defending the idea that our actions also influence the past states of the world.

It is worth noting that Esfeld has explicitly stated in his "Super-Humeanism and free will" article that his own proposal does not appeal to a block universe perspective and his rejection of the PAST premise of the Consequence Argument is distinct from the way in which block theorists like Hoefler reject it (Esfeld 2021, 10-11). His argument does not rely on assumptions about the nature of time, but on a slightly different conception of Humeanism about laws – Super-Humeanism – which he uses to reject both PAST and LAWS premises from van Inwagen's Argument.

The reason why I have chosen Carl Hoefler's Inside Out Perspective as the target of the ontological fatalism objection of this section is because his proposal seems to be the first robust defense of free will within a block universe eternalist perspective.¹⁸

¹⁸ Hoefler acknowledges that there is a similarity between his proposal and Peter Forrest's (1985) backward causation defense of free will, but Forrest's account does not make any particular assumptions on the metaphysics of time.

6. Conclusion

In this paper, I aimed to show old and new possible objections towards Carl Hoefer's Inside Out Perspective. Two of them belong to Jason Brennan (2007) and the other two belong to me.

After introducing the problem, I showed in the second section how the free will problem can be linked to the metaphysics of time by stating the temporal assumptions of two well-known incompatibilist arguments: the Forking Road Argument and the Consequence Argument.

In the third section, I presented the Inside Out Perspective that was proposed by Carl Hoefer in his 2002 article "Freedom from the Inside Out." The main idea there is that free will's conflict with determinism is not because of determinism itself, but because of our common sense view of time, the A-series, which privileges past→future determination. Hoefer also argues that if we adopt a B-theory of time, the block universe, then we can make freedom compatible with determinism.

In the fourth section I presented Jason Brennan's two criticisms against Hoefer's proposal: one involving the threat to free will from the asymmetry of causation that is left unaddressed by the block universe perspective and the other involving the unexplained ontological status of our actions within the block.

In the fifth and final section I presented my own objections to the view. I argued, firstly, that Hoefer's characterization of the A-series is too specific to sustain his general conclusion about the A-series. The conclusion that A-series privileges past→future determinism can perhaps be true only of the growing block theory of time and not necessarily true of the other A-theories. Secondly, I argued that if his proposal is successful in making determinism compatible with free will, then it must also address the threat of ontological fatalism. Ontological fatalism has been framed as the idea that the nature of eternally existing future events threatens freedom. I have suggested that the freedom under threat is the ability to do otherwise and that because such an ability, according to prominent libertarians, implies an open future with genuine alternatives, then this kind of freedom is in direct conflict with eternalism, which is incompatible with an open future. Thus, if Hoefer's perspective does not avoid ontological fatalism or explain how

eternalism can be reconciled with an open future, then his theory fails to secure freedom to do otherwise within the block universe.

The threat of fatalism was the objection that led people to believe that the block universe is incompatible with free will in the first place. And it is this initial objection that I believe Hofer's account (and any other eternalist compatibilist account) does not avoid. One could still say: If there is no ontological distinction between past, present and future events, then the future is just as real as the past. If so, can I make those future events *not* occur? If I cannot, does this not threaten my ability to do otherwise?

The defender of the Inside Out Perspective might be able to respond to many or most of the objections presented in this paper. In order to avoid Brennan's first criticism, one might be an eliminativist towards causation in the spirit of Bertrand Russell, and claim that causation is a concept much like the A-series: they both have no place in contemporary physics. Although, of course, the price of renouncing causation might be too high even for the eternalist.

The brute fact/random occurrence problem raised by Brennan is a bit trickier. If one would bring in agent causation into the block, then one must then show how the block universe isn't beside the point and how agent causation isn't sufficient by itself to defend free will. One might try to counter this objection by arguing that it was never assumed that our actions are not entailed also by earlier and later time slices and argue that agents partially determine other time slices only from the inside-out perspective and not from the outside-in perspective. Our actions are indeed fully explainable at a certain level of analysis (the microphysical, for example) by earlier or later states of the world, but only if we view determination in the block from the outside-in and not if we view it from the inside-out, as Hofer suggests.

As for my personal objections, the defender of the Inside Out Perspective can guard against the first one by shifting gears. One can claim that while there certainly are A-theories in which free will is possible, freedom is more compatible with determinism in a B-theory of time. As for my main objection towards the perspective (the threat of ontological fatalism), the defender could probably escape fatalism if the freedom under discussion would not be defined in terms of the ability to

do otherwise, but in terms of the absence of coercion and other influences (substances that alter brain states, hypnosis and so on). This “traditional compatibilist freedom” would probably raise the question of the Inside Out Perspective’s relevance to free will and this is perhaps the main point on which the defender of the static block must insist on. Nonetheless, it must be specified what kind of freedom we can have within the block.

As a final remark, I believe Hofer’s article is a very interesting contribution to the free will debate and pushes us to think about topics less discussed, such as the relation between free will and the various metaphysical theories of time. The compatibilist account of freedom within the block universe may have its issues, but the conversation started is one that warrants more pages to be written.

Acknowledgments

I am very grateful to an anonymous reviewer whose extensive comments and criticisms have helped me with substantially improving this paper. The present work has benefitted immensely from his helpful advice.

REFERENCES

- Backmann, Marius. “I Tensed the Laws and the Laws Won: Non-Eternalist Humeanism.” *Manuscrito* 39, no. 4 (2016): 255-277.
- Beebe, Helen, & Mele, Alfred. “Humean Compatibilism.” *Mind* 111, no. 442 (2002): 201-223.
- Belnap, Nuel, & Green, Mitchell. “Indeterminism and the Thin Red Line.” *Philosophical Perspectives* 8 (1994): 365-388.
- Belnap, Nuel. “Branching space-time.” *Synthese* 92, no. 3 (1992): 385-434.
- Bigelow, John. “Presentism and Properties.” *Philosophical Perspectives* 10 (1996): 35-52.
- Brennan, Jason. “Free will in the Block universe.” *Philosophia* 35, no. 2 (2007): 207-217.
- Broad, Charlie Dunbar. *Scientific Thought*. London: Kegan Paul, 1923.

- Diekemper, Joseph. "B-Theory, Fixity, and Fatalism." *Noûs*, 41, no. 3 (2007): 429-452.
- Esfeld, Michael. "Why Determinism in Physics Has No Implications for Free Will." *Organon F* 26, no. 1 (2019): 62-85.
- Esfeld, Michael. "Super-Humeanism and free will." *Synthese* 198, no. 7 (2021): 6245-6258.
- Farr, Matt. "On A- and B-theoretic elements of branching spacetimes." *Synthese* 188, no. 1 (2012): 85-116.
- Fischer, John Martin; Kane, Robert; Pereboom, Derk & Vargas, Manuel. *Four Views on Free Will*. New York: Wiley-Blackwell, 2007.
- Fischer, Florian. "Philosophy of time: A slightly opinionated introduction." *Kriterion – Journal of Philosophy* 30, no. 2 (2016): 3-28.
- Forrest, Peter. "Backwards causation in defense of free will." *Mind* 94 (1985): 210-217.
- Hofer, Carl. "Freedom from the Inside Out." *Royal Institute of Philosophy*, 50 (2002): 201-222.
- Hofer, Carl. "Causality and Determinism: Tension or Outright Conflict?" *Revista de Filosofia* 29, no. 2 (2004): 99-115.
- Horwich, Paul. *Asymmetries in Time: Problems in the Philosophy of Science*. Cambridge, Mass.: MIT Press.
- Ismael, Jenann. *How Physics Makes Us Free*. New York: Oxford University Press, 2016.
- Le Poidevin, Robin (ed.) *Questions of Time and Tense*. Oxford: Oxford University Press, 1998.
- Le Poidevin, Robin. "Time and Freedom." In: *A Companion to the Philosophy of Time*. Edited by Heather Dyke and Adrian Bardon. New York: Wiley-Blackwell, 2013. 535-549.
- Lewis, David K. "Are we free to break the laws?" *Theoria* 47, no. 3 (1981): 113-121.
- Loewer, Barry. "Two accounts of laws and time." *Philosophical Studies* 160, no. 1 (2012): 115-137.
- Loewer, Barry. "The consequence argument meets the Mentaculus." In *Time's Arrow and the Probability Structure of the World*. Edited by Barry Loewer, Brad Weslake and Eric Winsberg. Cambridge, Mass.: Harvard University Press, forthcoming 2023.

- Markosian, Ned. "A Defense of Presentism." *Oxford Studies in Metaphysics* 1 (2004): 47-82.
- McTaggart, J. Ellis. "The Unreality of Time." *Mind* 17 (1908): 456-473.
- Miller, Kristie. "Time Travel and the Open Future." *Disputatio* 1, no. 19 (2005): 223-232.
- Miller, Kristie. "Backwards Causation, Time, and the Open Future." *Metaphysica* 9, no. 2 (2008): 173-191.
- Miller, Kristie. "Presentism, eternalism, and the growing block." In Heather Dyke & Adrian Bardon (eds.), *A Companion to the Philosophy of Time*. New York: Wiley-Blackwell, 2013. 345-364.
- Russell, Bertrand. "On the Notion of Cause." *Proceedings of the Aristotelian Society* 13 (1912): 1-26.
- Shanks, Niall. "Time, Physics and Freedom." *Metaphilosophy* 25, no. 1 (1994): 45-59.
- Skow, Bradford. "Relativity and the Moving Spotlight." *Journal of Philosophy* 106, no. 12 (2009): 666-678.
- Tooley, Michael. *Time, Tense, and Causation*. Oxford: Oxford University Press, 1997.
- van Inwagen, Peter. "The Incompatibility of Free Will and Determinism." *Philosophical Studies* 27 (1975): 185-199.
- van Inwagen, Peter. *An Essay on Free Will*. Oxford: Oxford University Press, 1983.
- van Inwagen, Peter. "Free Will Remains a Mystery." *Philosophical Perspectives* 14 (2000): 1-19.
- Vihvelin, Kadri. *Causes, Laws, and Free Will: Why Determinism Doesn't Matter*. Oxford: Oxford University Press, 2013.
- Wasserman, Ryan. *Paradoxes of Time Travel*. Oxford: Oxford University Press, 2018.

THE LAST NAIL IN THE COFFIN OF SCIENTIFIC THEORIES¹

DAIAN BICA²

Review of Steven French, *There Are No Such Things as Theories*, Oxford: Oxford University Press, 2019, 288 pp.³

Steven French's last book *There Are No Such Things as Theories* brings about a new and provocative way of rethinking and reshaping the debates in philosophy of science by jettisoning the concept of scientific theory and replacing it instead with a rich ontology of scientific practices. The focal point of this approach seems to be that we still lack a good set of criteria to make sense of theories — which French takes to mean no less than that *there are no such things as theories* out there in the world ready to be discovered (223). This rather revolutionary framework encourages the reader to reassess the scope of scientific theory in the light of theory eliminativism — more precisely to free herself “from this illusory ontology” (239). As I will argue in what follows, French's main argument is a *reductio ad absurdum* that operates throughout the book: given the fact that approaches to theories fail to specify what a theory *is*, philosophers should discard the very idea of such a thing (180-182).

¹ This review originally appeared on the “Let's talk about books” academic blog. It was accessed here: <https://letstalkaboutbooks.blog/2020/12/05/the-last-nail-in-the-coffin-of-scientific-theories/> The editors thank both the author and the coordinator of the blog, professor Dana Jalobeanu from the University of Bucharest, for agreeing to reprint the review.

² Daian Bica is a doctoral student at the University of Düsseldorf. Contact: <dabic100@uni-duesseldorf.de>

³ Unless otherwise noted, the page references below are to the book being reviewed.

In addition, French's book presents the reader with arguments coming from fields as different as philosophy of music and philosophy of art regarding the ontological status of artworks, fictions, or music pieces. Questions such as *Does Picasso's Guernica exist as an abstract object? Is Beethoven's Fifth Symphony a real Platonist entity inhabiting a realm rather different than the physical one?*, are imported directly in philosophy of science (French 2020). Subject to reflection, the problem refers to the mere possibility of justifying analogies between art and science, and if that is the case, why those analogies hold and where the analogies lead to. An example would be that theory eliminativism stems from discussions regarding whether statues exist or not (184). Such an option in the philosophy of art, taken by Cameron, is to assert that "There are statues" is false since at a fundamental level there are only statue-shaped atoms (Cameron 2008, 301). To a certain extent, the same move is done in theory eliminativism ("theory-shaped bits of practice"), exhibiting a relation of a certain kind with artwork eliminativism (192, 239).

Steven French is a well-known British author, much appreciated for many contributions of great value in the English-speaking philosophy of science – in debates, to name a few, regarding philosophical problems in quantum mechanics, scientific realism, metaphysics of science, the interplay between science and art, or the role of models in scientific activity. In this biographical respect, *There Are No Such Things as Theories* also relies on previous approaches that the author has been elaborating elsewhere in his work. Resurfaced here, for instance, is the problem of the so-called "Viking" or "toolbox" (meta-philosophical) approach to philosophy of science from his 2014 book *The Structure of the World* – roughly speaking, a concept that is imported from, say, metaphysics or philosophy of art *should* be domain-specific to scientific practice (French 2014, 49-50). To speculate a bit, an example of such a conceptual import is, in fact, the much-disputed concept of scientific theory. Given that in the plurality of practices there is no place for theories and, consequently, *no* metaphysical commitment thereafter, theory is not a topic-specific tool for understanding modern science.

Nevertheless, another example of what the Viking Approach amounts to is to again consider the interplay between artwork eliminativism and theory eliminativism — an idea that is tailored for topics and debates in

philosophy of science. Nonetheless, theory eliminativism is backed up with a belief (another domain-specific tool in the Viking Approach's sense) in a fundamental ontology that is informed by contemporary practices in quantum mechanics — that is, a metaphysics of structure replacing one of self-sustaining objects (French 2014, 205). To finish this biographical *detour* with a concluding remark, French's previous Viking Approach already had the philosophical ammo to fuel such a stance as theory eliminativism.

There Are No Such Things As Theories should be integrated in the *status quo* of contemporary philosophy of science in order to understand the transition from theories to scientific practices as units of philosophical analysis. The concept of theory was, from a historical point of view, the bastion of philosophy of science from its early days, arguably, the 19th century, until very recent times, roughly, the last decades of the previous century, when it underwent as an academic field a turn to the role of scientific practices. In other words, (before its turn to practice) philosophy of science is centred around the concept of scientific theory. The meta-philosophical orientation towards theories sets down an (explanatory) agenda for what philosophers of science *should* do – to show scientists make sense of *theories* from the Scientific Revolution to the days of the Large Hadron Collider and of the Standard Model of subatomic particles.

Let's consider three textbook examples of the most relevant theory-oriented philosophies of science. Take a look at Pierre Duhem's definition of theory from his *Aim and Structure of Physical Theory* that grounds 19th century debates in philosophy of science: "a physical theory *is* an abstract system whose aim is to summarize and classify logically a group of experimental laws" (Duhem 1991, 7). Consequently, in Duhem's view, the aim of science revolves around searching for such abstract systems (Duhem 1991, 7-9). Or let's go some decades later and analyse Nagel's Syntactic View of theories from *The Structure of Science* where a theory *is* "an abstract calculus" and "a set of rules" that relate the calculus to "empirical content" (Nagel 1979, 90). Elsewhere in the book, Nagel firmly says that "the distinctive aim of the scientific enterprise is to provide systematic and responsibly supported explanations" – the process of systematization is achieved by way of scientific theorizing (Nagel 1979, 15). Perhaps we should make a step even further and take

as an example van Fraassen's classical defence of the Semantic Approach to theories from *The Scientific Image*: "to present a theory is to specify a family of structures; its *models*" (van Fraassen 1980, 64).

It is almost obvious that these *loci classici* of philosophy of science, be it Duhem's mathematical representational approach, or logical positivism, or constructive empiricism, were using the very concept of theory as a "sortal term" that *is* an entity of a certain kind (181). The concept of theory was identified, in turn, either with fictional set-theoretic structures (van Fraassen advocating the Semantic Approach), or with highly abstract mathematical representation (Duhem defending the Syntactic View), or with linguistic propositions (Nagel, also providing a version of the Syntactic View). Consequently, French brings into critical consideration each of those alternatives. I will comment on these approaches below.

As a critical reaction to the theory-based approaches, various philosophers of science challenged the basic assumptions of theory-centered projects by emphasizing the role of scientific practices, and giving birth henceforth to an array of trends tied together under the umbrella concept "the practice turn". One influential alternative was to raise the problem of practices under the form of the genuine knowledge furnished by *techne*, crafts, technologies, or experiments – on this view, theories are only tools relative to these modelling practices (Cartwright 2019, 4). Or, taking another practices-based conceptual route, other philosophers hold scientific practices are culturally and historically-situated perspectives or points of view (Giere 2004). In the perspectivist understanding, theories are highly theoretical principles that define "a quite abstract object" that is in turned used in building up representational models (Giere 2004, 69).

Bearing in mind the switch from theories to practices, *There Are No Such Things As Theories* is perhaps the last nail in the coffin of the concept of theory. I should stress that French also departs from the usual practice-based approaches in the sense that Cartwright *identifies* theories with tools and Giere *identifies* theories with perspectives (191-192). If we follow the eliminativist stance, it is not possible, to begin with, to ask how the identification should take place since philosophers do not have what theories to identify with. French compares his eliminativism with Cartwright's instrumentalist view: "this is the crucial difference: there

are no theories in my view" (192). One can say that only by now, with the publishing of *There are no such things as theories*, the transition or turn from theories to practices in philosophy of science is finally achieved!

Steven French accomplishes the ultimate turn to practices, so to speak, in the 7th and 8th chapters of the book under the form of eliminativism, that is an ontological framework, such that at stake it is the problem of what exists (there are no theories) and what is not (there are practices). Within the ontological framework, eliminativism endorses two distinct core-theses. The first core-thesis includes a theory about truth-makers according to which true statements are 'made' true by certain features of realities (182). Secondly, theory eliminativism has also a proper fundamental ontology – that characterises "how the world is at its most fundamental level" (183). Consequently, elements of this ontology will serve as the truth-makers for sentences that mention both these fundamental elements and other non-fundamental elements (183). The truth-makers of propositions concerning theories are not theories *tout court* but "the complex of practices of the scientific community" that are "all that *really* exists in this context". Steven French's concept of practice is rather broad, it ranges from "the writing and dissemination of articles, the performance of experiments, (...) heuristic moves", to journals, papers, PhD thesis, "an arrays of human activity" or (arguably) concept formation (191).

How does theory eliminativism work out after all? The readers get a clue of how eliminativism is looking at work in the 7th chapter. To begin with, both core-theses rest on a later distinction drawn between English (as a non-fundamental language) and Ontologuese (as a fundamental language) (187). For instance, one may say "Quantum mechanics is an elegant theory" (189). This latter sentence is formulated in English. When a speaker utters this proposition (in a non-fundamental language), she is not metaphysically committed to the existence of quantum mechanics as a theory (as an entity of a certain kind). Instead, the metaphysical commitment of the speaker is to the plurality of scientific practices concerning quantum mechanics – that works in turn as the truth-maker of the proper proposition. Concerning the real reference of the proposition, it is formulated consequently in Ontologuese. Perhaps the sentence "Quantum mechanics is an elegant

theory” refers to a corresponding practice, one that involves writing strings of equations on a whiteboard in a physics course, or writing down the set of equations in a quantum mechanics textbook. Elegance as an aesthetic predicate is a property of a certain associate practice, but not “of the theory in any metaphysical serious sense” (197).

The author returns upon this example in the 8th chapter to highlight that quantum mechanics is not a theoretical monolith, “a unitary and well-defined entity, with define identity conditions” (208). A scrupulous analysis of the history of quantum mechanics shows that the very idea “of a parade of putative theories” is precisely a construction, whether done by historians or by scientists themselves (as historians of their own field) (203). How does French ground this ambitious claim regarding modern science? The author cuts the Gordian knot by showing that the historiographical claim of a Quantum Revolution which takes place somewhere between 1927-1928 is not that obvious an historical fact. To put it briefly, French claims in this regard: “the quantum revolutionaries differed with regard to what they took ‘the’ theory to be and what principles they felt at the heart of it” (205).

Considering the principles that are supposed to lay the foundations of quantum mechanics, one can become aware of the fact that it is not entirely what those principles are. Whether one examines von Neumann’s formulation, or Weyl’s group-theoretic approach, or Schrodinger’s wave mechanics, or Dirac’s wave mechanics, each and every approach is different in regards of the foundational principles (e.g. distinct mathematical formalism) (203-207). Those principles seem to be embodied in famous textbooks on quantum mechanics (scientific practices in other words). The same situation arises again in considering what interpretation of quantum mechanics is ‘the’ theory (207-208). Both situations concerning quantum physics show that philosophers should cast doubt on that there are theories and, on the other hand, “come up with an ontology of theories that reflect these practices” (223).

There Are No Such Things As Theories is elegantly structured as it follows. The chapters (1)-(6) provide a general survey of the literature on what theories *could be*, about which Steven French offers a cost-benefit analysis. As a consequence of the overall discussion, which shows the failure of all those theory approaches to specify the conditions of

identity for what theories *are*, the author proposes instead theory eliminativism (chapters 7 and 8). Generally speaking, chapters (1) — (6) should be read by the read as a step-by-step elimination of alternatives. Let's consider the alternatives one by one.

Chapter (1) revisits the Syntactic View on theories, on the face of which theories are collections of *logico-linguistic propositions*. The Syntactic View is, historically, one of the core features of logical positivism. One can distinguish between weaker and stronger versions of the Syntactic View (10). According to the stronger version, the variant defended in fact by positivists, theories are abstract logical calculi, logically closed under first-order logic and that are further on subject to interpretation (3-4). In its turn, the interpretation is determined by "correspondence rules" that bind together or "bridge the gap" between the observable and theoretical languages one with another (12-13). We as philosophers of science reach an "understanding of what is a theory of (...) once the correspondence rules are laid down" (13). Within the strong approach, correspondence rules deploy a certain role in individuating theories (13). On the other hand, the weak version retains the very idea of theories as propositions, but rejects the framework of correspondence rules – according to the weaker version, the proper rules do not pick out what a theory *is* (10).

The next alternative (approached in chapter 2) is the Semantic View of theories, wherein theories are taken to be *collections of models*, that are nonetheless *extra-linguistic entities*. Not having a linguistic nature (*contra* the Syntactic View), theories as models can get a number of different linguistic formulations (33-34). Contrary to the Syntactic View, theories as models possess "linguistic independence" (36). What are models, truly? As mathematical objects and represented in a formal framework, we define models as *structures*: $\mathcal{M} = \langle A, R_i, f_j, a_k \rangle$ $i \in I, j \in J, k \in K$, where A stands for a non-empty set, R is a family of relations, f is a family of functions and a refers to a family of individuals of set A (36). More precisely, the approach of theories as collection of models understands the latter as mathematical structures (36). This approach enables a certain understanding of models as the vehicle of scientific representation in order "to describe the relations theories have to each other and to phenomena" (37). For a model (theoretical model) to represent its system target (data model) means that the former is totally

or partially isomorphic with the letter. In this sense, isomorphism is the relation of sharing the same structure between two models, a relation that obtains when a model is said to represent a system target. One may ask: “given that scientific models are, primarily, representations, in what sense may they also be mathematical structure?” (46). But is a model more than a formal skeleton, namely, as something that has a relation of representation with a physical system? (45-46). If it is the case, how can one define the representational relationship?

This string of questions opens up the third chapter, where Steven French consequently discusses models and theories as *scientific representations*, whereas models are understood to be *the vehicle of representation* (51). In asking whether theories are representations, French is again taking the relation between art and science as a source of inspiration (52-53). Here the author examines mainly two kinds of accounts. Either one that construes representation in terms of similarity relationships between what is represented and what represents, or one that defines the relevant representation relationships in terms of isomorphism between the former and the latter (51). According to the first account, representation as similarity works as an asymmetric relation – for instance, *a* represents *b*, but *b* doesn’t represent *a* (52). An example from art: Freud’s *Benefits Supervisor* can be said to represent Sue Tilley, the subject of the painting, although Sue Tilley cannot be said to represent the painting itself (52). In addition to that, similarity is a relation of material resemblance that holds between the represented and the representation. The model of billiard balls represents (is similar to!) the behavior of the gaseous particles in terms of motion, collision, or momentum (material features that the billiard balls are said to share with the gaseous particles). However, the real particles do not have the same size as the billiard balls.

On this view of the isomorphism account, a relation of representation stands for sharing the same (formal) structure between the model and the physical system that is represented. In other words, “certain relations which hold in the real system will be represented by corresponding relations holding between elements of the sets, but others will not” (61). Taking the example of a pendulum model: scientists describe it in terms of a point-like bob that lacks friction and of a massless string (61). In the structuralist view, a relation of scientific representation holds between

the pendulum model and a real actual pendulum – “what the material and the ideal pendulums have in common are *aspects of the relevant structure*” (61). However, both views are not however mutually exclusive – the isomorphism account can be understood as a more formalized version of the similarity approach (95-96).

The next three chapters (4), (5), and (6) ask what theories and models are as abstract entities: Are they fictional entities similar to fictional characters? Or do theories behave like artifacts (paints or music pieces)? Or perhaps do theories exist out there in a Platonist world? In chapter (4), French explores a debate in the philosophy of art regarding whether artworks are abstract objects, and if this is the case, how are they brought about since there is a tension between the very idea of an abstract object and the process of (concrete) creation (100-101). One move is to follow the Vikings approach and to import this problem in the ontology of theories (112-113). This is, precisely, the main topic of chapters (5) and (6). In chapter (5), the author explores two distinct approaches. According to the first, advocated by Karl Popper, one should distinguish between a First World (the physical world – one of physical entities, processes), a Second World (the realm of mental – the world of mental states) and the Third World (the world of theories, models, artworks) (116-118). In this view, the process of theory construction involves the discovery of *theories that are abstract entities out there in the world*. Popper argues in favor of the Third World along this line: being given that scientists manage “to discussing the same *thing*”, theories exist as abstract entities (119). Under this view, theories are not created, solving thus the above tension from the fourth chapter.

An alternative option is to take into consideration Thomasson’s account, by characterizing certain theories as *abstract artifacts*, that lack of spatio-temporal location – they can also be regarded as abstract although “they are still created, come into existence, change, and may cease to exist” (123-124). Within the fictionalist approach, models and theories, quantum mechanics and billiard ball models are compared with fictional characters and books, such as *Lord of The Rings* and *Frodo*. In which sense are those on the same par? A fictional character and a model are *abstract artifacts that come into existence in a particular set of practices*, work of fiction (in the first case) and experimental setup or modeling practices

(in the second case) alike (124). To put it the other way around, they are similar in the respect of the process of being brought about or created. Both Frodo and the billiard ball model lack spatio-temporal location and can be traced back to the relevant practices where they are embodied (125).

Recall the above example with quantum mechanics. If it is a theory, when was it discovered, according to Popper's view? With Bohr, or with Dirac, or with Heisenberg? Or, buying into Thomasson's view, when and where it did come into being? The general problem is that "we begin to think about how this sort of account might mesh with 'scientific discovery' in general and well-known heuristic moves in particular" (151). At this juncture, heuristic moves mainly mark the methodological and experimental procedures to which the development of theories is supposed to be subordinated. This kind of objection calls into question the plausibility of both Popper's and Thomasson's views.

Chapter (6) deals with a fictional account of theories (152). On this view, theories and fictions are on the same ontological par (*contra* Thomasson, for whom theories are not fictions, what do they actually share, is that both are relative to particular practices). Accordingly, propositions concerning the theory of general relativity, for instance, are not *literally true*, but true relative to or within that theoretical framework. This could mean that when scientists talk in their practice about the theory of relativity, they are engaging in a game of make-believe, prop, or pretense (20, 152-154). Briefly, those scientists are pretending that their propositions are *about* a kind of entity called "the theory of relativity" – the entire game of make-believe is "delineated by a kind of convention or principle of agreement", in this case, among scientists (21). One may ask: if *theories are fictions*, what *are* fictions? They could be *possibilia*, non-actual possible worlds (156). Describing the idealized model of the pendulum, scientists are referring to a non-actual possible world that lacks friction forces (156-157). The other option left is to explain fictions as "objects of our imagination" that draw scientists in the game of make-believe. (159). Regardless of which option one may choose, fictionalism can't account for some practice-related problems. What about models that contain "very general properties of infinite populations" from population biology (173)? In order to answer this problem, the fictionalist defender has to accept that those models are fictions by which they are

entertaining in a make-believe game (174). Or, it is not the case that scientists are aware in the scientific practice of the fictional nature of those models (174).

I think that the main problem around which chapters (5), and (6) revolve, is that although those approaches have their own theoretical merits, they are ontologically costly in the sense that they do not resolve the problems they address. The key lesson is perhaps that philosophers of science should just embrace theory eliminativism:

“We could chop through this knotty (Gordian) bundle of issues by simply denying the initial assumption, namely that theories are things or entities, abstract or otherwise, to begin with” (151)

One may balk at theory eliminativism by critically asking: if we accept that there are not conditions of identity for theories, what role will the philosopher of science perform instead? Would she be forced just to describe the doings and happenings in the scientific practices without talking about representations, models, or theories? (233). Perhaps “we should simply focus our collective attention on the practices” (234). In the last (ninth) chapter, French delivers an ingenious response to this concern. As philosophers of science, we should take the representation relationship between theories and models as a philosophical meta-construction (235). Meta-construction means, in this case, a philosophical discussion done either by professional philosophers or by scientists thinking philosophically about “theories or models representing some target system” (235). By this (meta)-philosophical assumption, we may make sense of scientific practice and its implications for how we should understand the world (235). The meta-level is distinguished from the object level, that is the level of scientific practice itself (20). The adoption of the assumptions meshes very well with a Syntactic View, or with a Semantic View, or with an isomorphism or similarity-based approach on scientific representation – that are “constructions that we philosophers of science introduce and use to do our work” (236).

Are those philosophical and historiographical constructions of any use given the eliminativism framework? Indeed, philosophers of science are not representing something at the object level. Rather, they are focusing more on constructions that enable themselves “to make sense

of certain features of scientific practice" (236). In the spirit of this object- and meta-level distinction, the role of philosophy of science is to make explicit the principles already at play in scientific practices "both current and as presented through the history of science" (238). On this basis, philosophers should assess what those principles commit scientists to and "how we can best make a consistent" theory that incorporates them (238).

Always engaging with the philosophical literature and even growing naturally out of it, *There Is No Such Things As Theories* is a critical diagnosis of the ongoing debates on scientific theories and, optimistically speaking, on how those debates should be directed from now on. Steven French manages successfully to provide a new philosophy of science that is tailored for the already established practice turn in the field. More than a new dismissive account of scientific theories, here we have the announcement of a novel way of dealing with philosophical problems related to scientific practices. Here is my guess: *There Is No Such Things As Theories* is our contemporary *Against Method*.

REFERENCES

- Cameron, Ross P. "There are No Things That are Musical Works." *British Journal of Aesthetics* 48, no. 3 (2008): 295-314.
- Cartwright, Nancy. *Nature, the Artful Modeler*. Chicago: Open Court, 2019.
- Duhem, Pierre. *The Aim and Structure of Physical Theory*. Translated by Philip F. Wiener. Princeton: Princeton University Press, 1991.
- French, Steven. *The Structure of the World: Metaphysics and Representation*. Oxford: Oxford University Press, 2014.
- Giere, Ronald. *Scientific Perspectivism*. Chicago: The University of Chicago Press, 2004.
- Nagel, Ernest. *The Structure of Science*. Indianapolis: Hackett Publishing Company, 1979.
- van Fraassen, Bas C. *The Scientific Image*. Oxford: Oxford Clarendon Press, 1980.

Reproducerea integrală sau parțială, multiplicarea prin orice mijloace și sub orice formă, cum ar fi xeroxarea, scanarea, transpunerea în format electronic sau audio, punerea la dispoziția publică, inclusiv prin internet sau prin rețele de calculatoare, stocarea permanentă sau temporară pe dispozitive sau sisteme cu posibilitatea recuperării informațiilor, cu scop comercial sau gratuit, precum și alte fapte similare săvârșite fără permisiunea scrisă a deținătorului copyright-ului reprezintă o încălcare a legislației cu privire la protecția proprietății intelectuale și se pedepsesc penal și/sau civil în conformitate cu legile în vigoare.

tipografia.unibuc@unibuc.ro

Bd. Iuliu Maniu 1-3, Complex LEU
tel: 0799 210 566

Tiparul s-a executat la Tipografia
Editurii Universității din București – Bucharest University Press (EUB – BUP)
