Revista Română de Filosofie Analitică

Romanian Journal of Analytic Philosophy

Volumul XVI, Nr. 2, 2022

EDITURA UNIVERSITĂȚII DIN BUCUREȘTI

BUCHAREST UNIVERSITY PRESS

2025

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)

www.srfa.ro

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

REDACȚIA / 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 Bucureşti – ROMÂNIA
Tel. +40 213053703

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

DTP / COPERTĂ
EUB-BUP

Revista Română de Filosofie Analitică
Romanian Journal of Analytic Philosophy

Vol. XVI **Iulie-Decembrie 2022** **Nr. 2**

CUPRINS
CONTENTS

Revista Română de Filosofie Analitică,
Romanian Journal of Analytic Philosophy
Vol. XVI, Iulie-Decembrie 2022, Nr. 2, p. 7-22
ISSN (ediția electronică): 1843-9969,
ISSN (ediția tipărită): 1844-2218
DOI: 10.62229/rrfaxvi-2/1

VIZIUNEA PLATONICIANĂ MODIFICATĂ ȘI O NOUĂ PERSPECTIVĂ[1] (THE MODIFIED PLATONIC VIEW AND A NEW PERSPECTIVE)

BOGDAN-ALEXANDRU APOSTOLESCU[2]

***Abstract*:** This paper examines a *modified Platonic view*, characterized by gauge symmetry and non-Euclidean geometry, as a framework for understanding the laws governing our known Universe. Building upon the work of physicist Lee Smolin, this perspective has also undergone a critical analysis that has provided the basis for a conceptual clarification of the modified Platonic view. As a result, the article proposes a more nuanced approach, termed *the approximate, modified Platonic view*, which aims to highlight the limitations but also the attractions of Platonism.

***Abstract*:** Lucrarea examinează *concepția platoniciană modificată*, caracterizată prin simetrie de etalonare și geometrie non-euclidiană, ca un cadru pentru înțelegerea legilor care guvernează Universul cunoscut de noi. În baza cercetărilor fizicianului Lee Smolin, această perspectivă a fost supusă și unei analize critice care a oferit premisele pentru o clarificare conceptuală *viziunii platoniciene modificate*. Ca urmare, articolul propune o abordare mai nuanțată, denumită *concepție platoniciană modificată și aproximativă,* care își propune să evidențieze limitele, dar şi avantajele perspectivei platonice.

***Keywords*:** gauge symmetry, non-Euclidean, modified Platonic view, approximate, Smolin.

---

[1] Acest articol redă, cu unele adăugiri și modificări, fragmente din lucrarea de doctorat a autorului: *Proiectul de unificare a forțelor fundamentale ale naturii. O abordare istorică și sistematică*. Traducerile îi aparțin.

[2] Bogdan-Alexandru Apostolescu este doctorand în filosofie al Universităţii din București. Contact: bogdan.apostolescu@gmail.com

1. Introducere

Tradiția filosofică ne transmite încercările întrăznețe de a explica Universul pe baze matematice. O astfel de tradiție este moștenită de la Platon, care a încercat să explice fundamentele Cosmosului prin relații geometrice euclidiene. A fost un demers remarcabil, care nu a reușit totuși să aibă o coresponență reală cu structura Universului cunoscut de noi. Cu toate acestea, ideea relațiilor de tip geometric s-a păstrat, iar cercetările secolului al XIX-lea și de început de secol XX au scos în evidență o nouă perspectivă, mult mai nuanțată, ce are la bază teoria grupurilor, care, de data aceasta, are o corespondență în lumea materială. Chiar și așa, modelele matematice actuale nu pot avea pretenția că înglobează toate legile universului, ci doar pe cele pe care le cunoaștem.

Concepţia actuală pune la fundamentul lumii cunoscute de noi nu formele geometrice simetrice[3] *(viziunea platoniciană)* prezentate în dialogul *Timaios*, ci, potrivit modelelor fizicii energiilor înalte, o simetrie extrem de profundă (simetrie de etalonare) care poate fi pusă, totuși, în relație cu Platon.

Teoriile fizice evoluează și ne dăm seama, așa cum susține fizicianul teoretician Chen-Ning Yang, că simetriile din fizica avansată nu urmează traiectorii diferite și că înțelegerea relațiilor dintre ele, cât și înțelegerea unificării conceptuale a lor, este o provocare importantă pentru fizicieni.

> „Observăm că, pe măsură ce domeniul fizicii se extinde, în timp ce principiile de simetrie cresc în număr, creșterea nu pare să urmeze linii independente. A înțelege relația și unificarea conceptuală între ele este o provocare majoră."[4]

---

[3] Platon transformă elementele constitutive ale Naturii din tradiția greacă de până la el în forme matematice simetrice, poliedre regulate: *Cubul pentru elementul Pământ, Tetraedrul regulat pentru Foc, Icosaedru regulat pentru Apă, Octaedrul regulat pentru Aer*, Dodecaedrul regulat pentru Cosmos. Primele patru poliedre regulate au fețe alcătuite din forme geometrice simetrice: *pătrate* (formate din două triunghiuri dreptunghice isoscele) și *triunghiuri echilaterale* (construite din două triunghiuri scalene „frumoase", după cu le numea Platon. Dodecaedrul nu e format din triunghiurile elementare amintite, însă are o relație geometrică simetrică, de exemplu, cu icosaedrul: *au același grup de simetrii și același subgrup de rotații*.

[4] Chen Ning Yang, *Selected Papers (1945-1980) with commentary*, World Scientific Publishng, vol 36, 2005, p. 279.

Prelegerile științifice legate de simetrie încep, aproape fără excepție, cu raportări la tradiție, iar dialogul *Timaios* este un reper. Este și cazul lui Yang, care insistă, ca și Platon, că simetria prevalează:

> „Din punct de vedere istoric, această dezvoltare este deosebit de interesantă, deoarece din cele mai vechi timpuri filosofii au încercat să coreleze simetria cu structura universului. În Grecia antică, Timaios și Platon au asociat cele patru substanțe naturale «fundamentale», respectiv focul, aerul, apa și pământul, cu tetraedrul obișnuit, octaedru, icosaedru și cubul. În China antică, I Ching a asociat simbolurile trigramei și hexagramelor cu fenomene naturale. *Desigur, simetria de azi nu înseamnă același lucru cu cea expusă de filosofii antici* [subl.m.]. Dar faptul că există relații conceptuale generale între aceste semnificații este greu de negat."[5]

Am depăşit, cum semnalează Dumitru (2011, p. 146 ff.) teoretizarea genurilor naturale clasice, de la apă la oxigen. Dimpotrivă, cum spune Yang (2005), simetria zilelor noastre nu este aceeași cu a filosofilor antici, Modelul Standard și Relativitatea Generală fiind un exemplu în acest sens.

2. Simetriile Modelului Standard, reprezentate ca Grupuri Lie; nu prin poliedre regulate platoniciene

Simetriile de etalonare nu sunt platoniciene, ele comportând alte tipuri de evidențiere decât cele ale poliedrelor regulate. Matematicianul și astrofizicianul Mario Livio afirmă că pentru oamenii de știință simetria este un far călăuzitor:

> „Biblia ne spune că, după ce au părăsit Egiptul, israeliții au fost călăuziți prin deșert «de un stâlp de foc noaptea, ca să le dea lumină». Simetria a fost stâlpul de foc al oamenilor de știință, conducându-i către relativitatea

[5] *Ibidem*, p. 81.

generală și modelul standard. Îi poate ea conduce și la unificarea celor două?"[6]

În continuare voi supune spre analiză această întrebare. Livio face apel la grupul de simetrii Lie (modele matematice), grup ce este compatibil cu fizica particulelor elementare. Practic – prin aceste modele – oamenii de știință au descoperit că „toate simetriile ce caracterizează Modelul Standard al particulelor elementare ... pot fi reprezentate ca produse de grupuri Lie simple".[7] Și astfel, în urma cercetărilor asupra fizicii particulelor și grupurilor de simetrie Lie, „grupul caracteristic al Modelului Standard a fost identificat cu un produs de grupuri Lie notate cu U(1), SU(2), SU(3)"[8], iar „drumul spre definitiva unificare a forțelor naturii trebuie să treacă prin descoperirea celui mai convenabil grup Lie ce conține produsul U(1) x SU(2) x SU(3)".[9]

Marius Sophus Lie a studiat grupurile de simetrie matematice și a încercat să explice conexiunea dintre aceste grupuri și geometrie, iar împreună cu Felix Klein a pus bazele programului Erlangen, potrivit căruia, subliniază matematicianul Ian Stewart, *geometria și teoria grupurilor reprezintă același lucru*. Stewart explică ce înseamnă că geometria și aceste grupuri sunt identice: există un grup de simetrie al unei geometrii date, în același timp manifestându-se reversul care ne arată că „geometria corespunzătoare unui grup este formată din toate obiectele asupra cărora acționează grupul de simetrie".[10] Cu alte cuvinte, simetriile geometriei euclidiene „sunt acele transformări ale planului care lasă invariante lungimile, unghiurile, liniile drepte și cercurile"[11] și invers, „orice obiect care este invariant la mișcările rigide intră în mod natural în sfera

[6] Mario Livio, *Ecuația care n-a putut fi rezolvată. Matematicieni de geniu descoperă limbajul simetriilor*, traducere din engleză de Mihnea Moroianu, ediția digitală, Editura Humanitas, București, 2014, p. 262.
[7] *Ibidem*, p. 261.
[8] *Ibidem*, p. 261.
[9] *Ibidem*, p. 261.
[10] Ian Stewart, *De ce frumusețea este adevărul. O istorie a simetriei*, traducere din engleză de Irinel Caprini, Editura Humanitas, 2010 p. 189.
[11] *Ibidem*, p. 190.

geometriei euclidiene",[12] iar Stewart ne spune, analizând acest subiect, că „geometriile neuclidiene folosesc pur și simplu alte grupuri de transformări".[13]

În această categorie a unei altfel de geometrii intră și grupurile Lie care sunt folosite în fizica energiilor înalte, aceste grupuri (fiind o descriere matematică a forțelor) prezintă modul cum particulele se cuplează, interacționează între ele, interacțiune care nu eludează simetria de la nivel cuantic. Astfel, grupul de simetrie asociat electromagnetismului este cel simbolizat cu U(1), forța slabă nu se abate de la simetria SU(2), iar forța tare nu deviază de la simetria SU(3). Pentru că forța electromagnetică și slabă au fost unificate într-o singură forța electroslabă, ea a fost asimilată grupului de simetrie SU(2) x U(1). Toate aceste trei simetrii sunt supuse studiului de către fizicieni pentru a fi unificate într-un grup de simetrie Lie și mai mare, una dintre soluțiile cele mai cunoscute fiind SU(5), care ar trebui să dea o formă matematică conformă cu experimentul Modelului Standard.

Astfel, metoda de lucru a acestui proiect încearcă unificarea forțelor fundamentale ale Modelului Standard folosind grupurile Lie, algebra Lie și reprezentările lor,[14] adică un alt fel de model matematic simetric ce poate fi pus în analogie cu cel al lui Platon. În sensul acesta, Mario Livio ne oferă conceptul de legătură dintre abordarea lui Platon și modul de lucru de astăzi, *concepția platoniciană modificată:*

> „Concepția platoniciană modificată susține că legile fizicii se exprimă ca ecuații matematice, structura universului este fractală, galaxiile se așează de la sine în spirale logaritmice, ș.a.m.d, deoarece matematica este limbajul universului."[15]

Așadar, modelele matematice avansate sunt încadrate de Mario Livio în noua cale platoniciană.

---

[12] *Ibidem*, p. 190.

[13] *Ibidem*, p. 190.

[14] John Baez; John Huerta, *The Algebra of Grand Unified Theories*, Bulletin (new series) of the American Mathematical Society, Volume 47, Number 3, July 2010, pp. 483–552, p. 483.

[15] Mario Livio, *Secțiunea de aur. Povestea lui phi, cel mai uimitor număr*, traducere din engleză de Mihnea Moroianu, Editura Humanistas, București, 2012, p. 275

3. Geometrie non-euclidiană pentru spațiu, nu platoniciană

Geometria non-euclidiană, alături de simetria de etalonare, poate fi încadrată, fără a greși, în *concepția platoniciană modificată*. Atât noua geometrie, cât și matematica particulelor, păstrează ideea simetriei, dar cu mari îmbunătățiri față de propunerile predecesorilor.

Albert Einstein a studiat ecuațiile lui James Clerk Maxwell (1831-1879) și a produs o nouă revoluție în fizică. Maxwell a folosit calculul vectorial și și-a seama că unda pe care o cerceta era de fapt lumina, pentru că această undă, din calcule, avea aproximativ aceeași viteză cu cea a luminii; această undă sinusoidală fiind compusă din două câmpuri, electric și magnetic, care se transformă unul în celălalt. Deci lumina este o undă electromagnetică. Există o dualitate caracterizată de simetrie și care i-a permis lui Maxwell să unifice electricitatea cu magnetismul și să o arate publicului printr-un limbaj matematic exact. Iată ce ne transmitea Maxwell în a doua parte a secolului al XIX-lea:

> „Deoarece expresia legii forței între cantități date de magnetism are exact aceeași formă matematică ca legea forței între cantități de energie electrică de valoare numerică egală, o mare parte din tratamentul matematic al magnetismului trebuie să fie similar cu cel al electricității.”[16]

Einstein și-a dat seama că ecuațiile lui Maxwell aveau o nouă simetrie, care până atunci nu fusese descoperită. El a modificat ecuațiile transpunându-le în patru dimensiuni și a schimbat între ele coordonatele, dându-și seama că rămân la fel. Așadar, Einstein a arătat că energia, materia, spațiul și timpul sunt toate componente ale unei simetrii 4-dimensionale, o simetrie fundamentată pe rotațiile în cvadridimensionalitate.[17] Ne aflăm, prin urmare, pe terenul teoriei relativității restrânse, care a adus într-un tot unitar particulele de lumină cu spațiu și timpul.

---

[16] James Clerk Maxwell, *A Treatise on electricity and magnetism, vol II*, Clarendon Press, Oxford, 1873, p. 4

[17] Michio Kaku, *Ecuația lui Dumnezeu. În cautarea unei teorii a tuturor lucrurilor*, traducere din limba engleză de Constantin Dumitru Palcus, editura Trei, 2022, p. 45

Mai mult, Einstein și-a continuat cercetările, dând o nouă definire Gravitației, în raport cu Newton, și definit-o prin curbarea spațiu-timpului, folosindu-se de geometrie, de o nouă geometrie.

Negarea postulatul V a lui Euclid a dus la dezvăluirea de noi tipuri de geometrie (*geometrie lobacevskiană – geometrie hiperbolică și geometrie riemanniană – geometrie eliptică*). Unde aceste noi tipuri de geometrii, numite neeuclidiene, au reușit să arate că „se pot construi teorii geometrice absolut valabile, perfect coerente, ... deşi în aceste geometrii postulatul lui Euclid asupra dreptelor paralele nu mai este considerat valabil”[18]. Aceste geometrii prezentate de Nikolai Lobacevski (1792–1856) și Georg Riemann (1826–1866) au atras nu numai interesul matematicienilor, fizicienilor, ci și al filosofilor.

Postulatul V poate fi exprimat în felul următor: *printr-un punct exterior unei drepte se poate duce o singură paralelă şi numai una la dreapta dată.* Însă, explică Anton Dumitriu, printr-o concepţie revoluţionară, s-a acceptat atunci, ca nou postulat, propoziţia contradictorie cu aceasta: *printr-un punct exterior unei drepte se poate duce o infinitate de paralele la această dreaptă*; sau încă, *propoziția printr-un punct exterior unei drepte nu se poate duce nici o paralelă la acea dreaptă* [19] (adică soluția Lobacevski și, respectiv, soluția Riemann). Astfel, „înlocuindu-se postulatul lui Euclid cu una din aceste propoziţii, s-a putut construi o geometrie perfect coerentă, necontradictorie, care nu e mai puţin valabilă decât geometria lui Euclid”.[20] Noutatea era că într-o asemenea geometrie „apar teoreme care nu mai au acelaşi enunţ ca în geometria lui Euclid”,[21] un exemplu fiind acela că suma unghiurilor unui triunghi nu mai este egală cu două unghiuri drepte, suma putând fi mai mică sau mai mare de 180 de grade. Această realitate a tras după sine o nouă viziune asupra spațiului, el nu mai era unul rigid.

De exemplu, pe Einstein geometria riemanniană l-a ajutat în formularea teoriei generale a relativității. Acest tip de geometrie i-a

[18] Anton Dumitriu, *Istoria logicii,* ediția a II-a revăzută și adăugită, Editura Didactică și Pedagogică, 1975, p. 987.
[19] *Ibidem*, p. 544.
[20] *Ibidem*, p. 544.
[21] *Ibidem*, p. 544.

înlesnit lui Einstein să înțeleagă modul în care materia determină gravitația și să înțeleagă faptul că aceasta nu este forță, ci curbura pe care o fac obiectele în dimensiunea spațiu-timp.

Einstein a acordat atenție sporită acestei geometrii, pentru că în lipsa ei i-ar fi fost imposibil să construiască teoria relativității. Astfel, în urma cercetărilor sale, el și-a dat seama că geometria euclidiană nu este compatibilă cu teoria relativității generale și a mers spre un altfel de geometrie, cea riemanniană, pentru a da sensul adevărat continuului spațiu-timp și a stabilit că acesta se supune unei geometrii neeuclidiene, unde metrica unei astfel de geometrii este curbată.

Ulterior lui Riemann, fizicianul-matematician Hermann Minkovski (1864-1909) își expune un punct de vedere, subliniat de Marco Andreatta, hotărâtor pentru teoria relativității generale a lui Einstein și important pentru schimbarea de optică asupra spațiului și timpului: ele nu mai au un caracter ideal, ca în mecanica clasică, newtoniană:

> "Conceptele de spațiu și de timp pe care aș vrea să vi le expun provin de pe tărâmul fizicii experimentale și în asta constă forța lor. Sunt radicale. De aici încolo, spațiul înțeles de sine stătător și timpul înțeles de sine stătător sunt sortite să dispară printre umbre și doar un fel de reunire a lor va mai putea avea o realitate independent."[22]

Bazându-se pe cercetările lui Riemann și Minkovski, Einstein formulează o teorie geometrică, teoria relativității generale, înglobând gravitația, unde propune un model geometric neeuclidian și unde realitatea fizică trebuie să fie interpretată ca fiind dependentă de patru coordonate: timpul și cele trei dimensiuni ale spațiului, acolo unde metrica spațiului este determinată de forțe care acționează asupra lui.

Einstein a găsit în acest tip de geometrie instrumentul potrivit pentru introducerea noilor sale idei cu privire la gravitație. Geometria lui Riemann este aplicabilă spațiilor largi, ea fiind o aproximație foarte bună, dovadă fiind validitatea relativității generale.

[22]Marco Andretta, *op. cit.,* p. 160

Așadar, lumea macroscopică este definită de o geometrie rimenniană, iar mecanica cuantică (microcosmosul) schimbă această perspectivă, lumea particulelor elementare funcționând după alte norme:

> „la scară ultramicroscopică trăsătura centrală a mecanicii cuantice – principiul de incertitudine – este în conflict direct cu principiul de bază al teoriei generale a relativității- modelul neted al spațiului geometric (și al spațiu-timpului)".[23]

În spații foarte restrânse, la nivel cuantic, formalismul matematic al lui Riemann nu mai poate fi aplicat. În teoria cuantică, cea mai simplă matematic,

> „există ceva mai fundamental decât spațiul nostru tridimensional și particulele din el: funcția de undă și locul cu dimensiuni infinite, numit spațiul Hilbert, unde acesta se află … [iar] funcția de undă și spațiul Hilbert, care constituie, se pare, realitatea fizică cea mai fundamentală, sunt obiecte pur matematice".[24]

Altfel spus, mecanica cuantică înlocuieşte spațiul perceptibil cu spațiul Hilbert, punând „funcția de undă"[25] în locul „traiectoriei"[26], *acest tip de spațiu fiind în strânsă legătură cu grupurile Lie*. Spațiul Hilbert este baza pentru formularea matematică a teoriei cuantice.

Cele două modele matematice care dau seamă de modul de funcționare al macrocosmosului și microcosmosului sunt folosite astăzi de fizicieni pentru a crea un model și mai complex (Teoria Totului) care să definească unitar Universul. Sintagma de „Teorie a Totului" rămâne,

---

[23] Brian Green, *Universul elegant. Supercorzi, dimensiuni ascunse și căutarea teoriei ultime*, traducere din engleză de Dragoș Anghel și Anamirela-Paula Anghel, Editura Humanitas, 2015, p. 148

[24] Max Tegmark, *Universul nostru matematic. În căutarea naturii ultime a realității*, ediția a doua revăzută și adăugită, traducere din limba engleză de Dumitru Dorian, Editura ASCR, 2016, p. 267

[25] Funcția de stare a unei particule, descrie starea cuantică a particulei.

[26] Locul geometric prin care trece un corp.

totuși, improprie, chiar dacă fizicienii vor reuși, la un moment, să unească Relativitatea Generală cu Modelul Standard al particulelor elementare.

4. Contra universalizării modelelor simetrice. Perspective: Smolin și Hossenfelder

Atât Modelul Standard, cât și Relativitatea Generală constituie, conform fizicianului Lee Smolin, o fragmentare a naturii:

> "Toate teoriile cu care operăm, inclusiv Modelul Standard al particulelor elementare și Relativitatea Generală sunt teorii aproximative, care se aplică unor trunchieri ale naturii care includ numai o submulțime a gradelor de libertate din univers."[27]

Smolin apreciază că acest model nu ia în considerare, din cauza limitării sale, alte fenomene, încă necunoscute, ce ar putea fi vizibile dacă fizicienii ar testa distanțe și mai scurte. Ca urmare, subliniază Smolin,

> „fenomenele lipsă ar putea include nu doar noi tipuri de particule elementare, ci și forțe anterior necunoscute (*n.m.* – deja cercetătorii analizează existența unei noi forțe) … s-ar putea dovedi că principiile fundamentale ale mecanicii cuantice sunt greșite și au nevoie de modificări pentru a descrie corect fenomenele care se ascund la distanțe mai mici și energii mai mari".[28]

Prin urmare, în conformitate cu punctul de vedere smolian, din cauza acestor neajunsuri „spunem că Modelul Standard e o teorie efectivă, compatibilă cu experimentul, dar demnă de încredere numai într-un anumit domeniu".[29]

---

[27] Lee Smolin, *Timpul renăscut. De la criza fizicii la viitorul universului*, traducere din engleză de Walter Fotescu, Humanitas, 2022, p. 153.
[28] *Ibidem*, p. 157.
[29] *Ibidem*, p. 157.

Smolin accentuează că Modelul Standard exclude relativitatea, iar relativitatea nu înglobează modelul cuantic. Ca și Modelul Standard, „relativitatea generală este, în cel mai bun caz, o aproximație a unei teorii cuantice unificate a naturii”,[30]fiind o trunchiere a unei teorii mai fundamentale.

Sprijinindu-se pe faptul că simetria se găsește doar la nivelul teoriilor aproximative pentru că acestea sunt caracterizate de o abordare restrânsă, el lansează ipoteza că la nivelul legilor fundamentale, nu locale, nu există simetrie.

Lee Smolin îl aduce în sprijinul său pe Gotfried Wilhelm Leibniz și susține că *principiul rațiunii suficiente* constrânge o teorie cosmologică la anumite tipare: *în univers nu trebuie să existe ceva care să acționeze asupra lucrurilor fără ca acel ceva să nu sufere, la rândul lui, nici o acțiune*. El introduce, prin acest principiu, perspectiva relațională dezvoltată la începutul secolului al XVIII-lea de Leibniz și afirmă că „dacă insistăm asupra acțiunii reciproce și excludem structurile de fundal fixe, susținem că de fapt orice entitate din univers evoluează dinamic, în interacțiune cu tot restul”.[31] Acest raționament atrage după sine ideea că *toate proprietățile reflectă relații aflate în evoluție*.

Acest raționament relațional, potrivit lui Smolin, are consecințe interesante, însemnând „că nu pot exista două corpuri care să aibă același set de de relații cu restul universului”,[32] el introducând aici *principiul indiscernabilelor* care nu se aplică, după Leibniz,[33] lumii materiale: dacă nu există două corpuri asemenea în univers, înseamnă că nici simetrie nu

[30] *Ibidem*, p. 159.
[31] *Ibidem*, p. 164.
[32] *Ibidem*, p. 164.
[33] Pentru Leibniz în lumea sensibilă, materială nu există indiscernabilitate, ele există doar în plan abstract: […] similitudinea perfectă nu are loc decât în noțiunile abstracte, în care lucrurile nu sunt luate în seamă în toate privințele, ci numai după o manieră deterministă de a le considera: de exemplu, atunci când luăm în seamă în mod exclusiv figurile, neglijăm cu totul materia figurată; de aceea două triunghiuri pot cu toată îndreptățirea să fie considerate asemenea în geometrie, deși nu se află nicăieri două triunghiuri material perfect asemănătoare. (Vezi Gotfried Wilhelm Leibniz, *Primae Varitates și alte scrieri de logică și metafizică. Leibniz prin el însuș*i, ediția a II-a, traducere, note și postfață de Alexandru Boboc, Editura Paideia, București, 2021, p. 21.)

există. Cu toate acestea, Smolin recunoaște că simetriile se întâlnesc în toate teoriile fizice cunoscute,[34] însă „dacă principiile lui Leibniz sunt corecte, aceste simetrii nu trebuie să fie fundamentale".[35]

Simetria este prezentă, crede Smolin, „pentru că un subsistem al universului e tratat ca și cum ar fi singurul lucru care există".[36] Practic, considerăm acel subsistem ceva fix și separat și neglijăm restul universului sau perspectiva relațională. Punctul de vedere al lui Smolin transmite faptul că, dacă privim universul dintr-un unghi de vedere relațional, rotația și translația nu comportă nici o însemnătate: „simetriile, ca de pildă translațiile și rotațiile, nu sunt fundamentale; ele apar din diviziunea lumii în două părți",[37] omul de știință fixând newtonian anumite cadre stabile care dau practic simetria. Ca urmare a acestui raționament, Smolin pregătește pasul către schimarea paradigmei de lucru:

> "dacă aceste simetrii sunt aproximative, atunci la fel sunt și legile de conservare a energiei, impulsului și momentului cinetic. *Aceste legi de conservare fundamentale depind de presupunerea că spațiul și timpul sunt simetrice în raport cu translațiile în timp, translațiile și rotațiile în spațiu"* [*subl. m.*][38]

Smolin, în contextul acesta, aduce în discuție cercetarea matematicienei Emmy Noether, care *stabilește o legătură între legile de conservare și simetrie*. Smolin spune despre Noether că, din unghiul ei de abordare, aceasta stabilește doar legi aproximative. Interogând observațiile lui Noether și aducând în discuție perspectiva relațională, Smolin afirmă că o viitoare teorie cosmologică nu trebuie să excludă actualele teorii, însă ele trebuie abordate ca fiind efective. Fiind aproximații, simetriile acestor teorii existente vor fi limitate la subsisteme ale universului. Prin urmare, noua teorie a Totului „nu va postula simetrii și nici legi de conservare"[39] la nivel fundamental. Mai mult, aceasta

[34] Smolin, Lee, *op. cit.,* p. 165.
[35] *Ibidem*, p. 165.
[36] *Ibidem*, p. 165.
[37] *Ibidem*, p. 165.
[38] *Ibidem*, p. 165.
[39] *Ibidem*, p. 170.

„trebuie să satisfacă principiul *rațiunii suficiente*, *principiul inexistenței acțiunilor fără reacțiune* și *principiul identității indiscernabilelor*"[40] și nu trebuie să aibă „structuri de fundal fixe, inclusiv legi fixe ale naturii".[41]

Smolin insistă asupra paradigmei newtoniene la care încă nu s-a renunțat, unde obiectul matematic care descrie legile naturii este unul atemporal, deci fix, și care intră în contradicție cu viziunea sa asupra *realității timpului*. Matematica, din perspectiva lui, înregistrează doar procesele fizice ce s-au încheiat, ea neavând posibilitatea de a cuprinde tot universul. Pentru Smolin, paradigma newtoniană nu este potrivită pentru explicarea întregului univers pentru că ea nu poate fi extinsă pentru tot universul, acesta fiind într-o continuă evoluție:

> "Lumea rămâne însă mereu un mănunchi de procese care evoluează în timp și doar mici părți ale ei sunt reprezentabile prin obiecte matematice atemporale. Întrucât paradigma newtoniană nu poate fi extinsă pentru a include tot universul, *nu este necesar să existe un obiect matematic care să corespundă istoriei exacte a întregului univers"* [*subl. m.*] [42]

El afirmă că permițând existența *realitatății timpului*, vom crește puterea științei și că soluția newtoniană o diminuează:

> *"Permițând legilor să evolueze în timp* [subl.m.], ne sporim șansele de a le explica prin ipoteze care au consecințe testabile. A accepta că legile evoluează în timp pare să le diminueze puterea, dar în realitate crește astfel puterea globală a științei, pe când extinderea ideilor care funcționează în paradigma newtoniană asupra domeniului cosmologiei diminuează puterea științei. *Dacă în perspectiva noastră asupra lumii includem evoluția și timpul la nivelurile cele mai profunde, avem mai multe șanse să înțelegem universul misterios în care ne aflăm."* [*subl. m.*] [43]

Și fizicianul Sabine Hossenfelder are un punct de vedere asemănător cu al lui Smolin despre legile naturii cunoscute de noi, numindu-le *incomplete*:

---

[40] *Ibidem*, p. 170.
[41] *Ibidem*, p. 170.
[42] *Ibidem*, p. 310.
[43] *Ibidem*, p. 317.

> "Știm că legile naturii pe care le avem în prezent sunt incomplete. Pentru a le completa, trebuie să înțelegem comportamentul cuantic al spațiului și timpului, revizuind fie gravitația, fie fizica cuantică, fie ambele. Iar răspunsul va conduce, fără îndoială, la noi întrebări."[44]

Hossenfelder are, ca și Smolin, o atitudinte rezervată față de frumusețe, estetică (simetrie, simplitate, eleganță) în fizică. Hossenfelder recunoaște succesele predecesorilor care au făcut descoperi remarcabile sprijnindu-se pe conceptele de simetrie, unificare, eleganță, însă spune că nu trebuie să ne cramponăm de ele dacă nu mai produc rezultatele de până acum.[45] Deși Hossenfelder ridică semne de întrebare asupra abordării estetice în fizică, ea nu oferă și criterii clare pentru ca fizica să avanseze fără ideea de estetică.

Reveninid la Smolin, el caută o altă abordare și consideră mai oportună o reciprocă la teorema lui Emmy Noether pentru că teorema deja clasică[46] „sugerează că spațiul este fundamental"[47] și că energia și impulsul sunt calități emergente care-i oglindesc simetriile:

> „Ceea ce vrem este o reciprocă a teoremei lui Neother, care să pornească de la presupunerea că energia, impulsul și conservarea lor sunt fundamentale și să ne spună în ce condiții spațiul poate emerge ca o descriere aproximativă a subsistemelor întregului."[48]

Reciproca vine ca urmare a postulării sale că o teorie fundamentală este independentă de fundal, ceea ce presupune că nu există simetrii și, prin urmare, nu putem considera că energia, impulsul și conservarea lor

---

[44] Sabine Hossenfelder, *Rătăciți printre formule. Cum îi deruteză frumusețea pe fizicieni*, traducere din engleză de Radu Sobodeanu, Humanitas, 2020, p. 271.

[45] *Ibidem*, p. 41.

[46] Smolin arată că teorema lui Noether implică noțiunea de simetrie și afirmă că ea nu se poate aplica unei teorii fundamentale, deoarece teoria fundamentală trebuie să satisfacă principiul identității indiscerabilelor, iar din acest principiu rezultă că nu există simetrii în natură.

[47] Lee Smolin, *Revoluția neterminată a lui Einstein. Căutarea a ceea ce se află dincolo de cuante*, traducere din engleză de Walter Fotescu, *Humantias*, București, 2023, p. 286

[48] *Ibidem*, p. 287.

sunt emergente din proprietățile spațiului. Rocada propusă de Smolin presupune ca timpul (reprezentat de energie și impuls) să preceadă spațiul și „prin urmare, dacă vrem ca energia și impulsul să joace un rol în fizică, se pare că singura posibilitate e să le introducem dintru început".[49]

Punctul de vedere al lui Smolin este că, în urma schimbării de paradigmă, rămânem cu o imagine în care sunt fundamentale relațiile cauzale, energia și impulsul. În acest sens, Smolin afirmă că principiile și ipotezele relaționismului temporal sunt exprimate de modelele de mulțimi cauzale[50], iar „în aceste principii, timpul, în sensul devenirii continue a momentului prezent, e fundamental pentru natură"[51]. Consecința este aceea că legile naturii nu sunt atemporale și evoluează în timp, iar „aceasta răstoarnă convingerea, răspândită printre fizicieni, că timpul nu e prezent în «legile cele mai fundamentale», ci emerge din aceste legi".[52]

Așadar, *timpul*, în sensul momentului prezent și al trecerii sale, e fundamental, în vreme ce legile sunt emergente și supuse schimbării – acesta este punctul central al studiului smolian.

Căutarea unei noi metode de lucru în spirit smolian este în momentul de față într-o fază incipientă, fiind foare greu să se înlocuiască soluțiile matematice prezente pentru că ele au dat rezultate totuși valide. Cu toate acestea, unghiul de abordare smolian, potrivit căruia noi în momentul de față studiem legi aproximative, a deschis calea acestui articol către *o nouă abordare platoniciană mai nuanțată.*

5. Propunere supusă studiului filosofic: concepția platoniciană modificată și aproximativă

Viziunii platoniciene modificate, mediatoare în momentul de față a unei Mari Unificări și unei Teorii a Totului, i se pot aduce corecții. Din punctul meu de

[49] *Ibidem*, p. 287.

[50] Mulțimile cauzale energetice sunt modele ale unor universuri cuantice care explorează conjecturile noastre legate de timp și spațiu. Aceste mulțimi iau ca fundamentale noțiuni active, ireversibile, de timp și cauzalitate, precum energia și impulsul.

[51] *Ibidem*, p. 287-288.

[52] *Ibidem*, p. 288.

vedere, o clarificare conceptuală este necesară, pentru că – dincolo de ceea ce noi cunoaștem în momentul de față – pot exista alte principii mult mai fundamentale, acest fapt neputând fi exclus. Poate că simetria naturii, pe care noi o postulăm prin diferite modele matematice, este doar un strat intermediar. Dacă Universul este în plină expansiune, atunci există posibilitatea ca în regiunile îndepărtate ale Universului legile să nu mai aibă aceeași formă cunoscută acum de oamenii de știință. Poate că viitorul fizicii va defini definitiv simetria ca având o arie de aplicabilitate limitată.

Pentru că, foarte probabil, ceea ce cunoaștem noi în momentul de față are cadre limitate, este indicat ca *viziunea platoniciană modificată* prezentată în această lucrare să fie mai bine încadrată şi delimitată prin adăugarea noțiunii evidențiate de Smolin, conceptul de *aproximativ*: *viziune (concepție) platoniciană modificată și aproximativă*.

BIBLIOGRAFIE:

Andreatta, M. (2021). *Forma Lucrurilor. Alfabetul Geometriei*, traducere din italiană de Liviu Ornea, Humanitas.

Baez, J. şi Huerta, J. (2010). The Algebra of Grand Unified Theories, *Bulletin (new series) of the American Mathematical Society*, Volume 47, Number 3, July 2010. https://doi.org/10.48550/arXiv.0904.1556

Dumitriu, A. (1975). *Istoria logicii*, ediția a II-a revăzută și adăugită, Editura Didactică și Pedagogică.

Dumitru, M. (2004). *Explorări logico-filozofice*, Humanitas.

Green, B. (2015). *Universul elegant. Supercorzi, dimensiuni ascunse și căutarea teoriei ultime*, traducere din engleză de Dragoș Anghel și Anamirela-Paula Anghel, Editura Humanitas.

Hossenfelder, S. (2020). *Rătăciți printre formule. Cum îi deruteză frumusețea pe fizicieni*, traducere din engleză de Radu Sobodeanu, Humanitas.

Leibniz, G. W. (2021). *Primae Varitates și alte scrieri de logică și metafizică. Leibniz prin el însuși*, ediția a II-a, traducere, note și postfață de Alexandru Boboc, Editura Paideia.

Kaku, M. (2022). *Ecuația lui Dumnezeu. În cautarea unei teorii a tuturor lucrurilor*, traducere din limba engleză de Constantin Dumitru Palcus, editura Trei.

Livio, M. (2014). *Ecuația care n-a putut fi rezolvată. Matematicieni de geniu descoperă limbajul simetriilor*, traducere din engleză de Mihnea Moroianu, ediția digitală, Editura Humanitas, București.

Livio, M. (2012). *Secțiunea de aur. Povestea lui phi, cel mai uimitor număr*, traducere din engleză de Mihnea Moroianu, Editura Humanistas, București.

Maxwell, J. C. (1873). *A Treatise on electricity and magnetism*, vol. II, Clarendon Press, Oxford.

Stewart, I. (2010). *De ce frumusețea este adevărul. O istorie a simetriei*, traducere din engleză de Irinel Caprini, Editura Humanitas.

Smolin, L. (2022). *Timpul renăscut. De la criza fizicii la viitorul universului*, traducere din engleză de Walter Fotescu, Humanitas.

Smolin, L. (2023). *Revoluția neterminată a lui Einstein. Căutarea a ceea ce se află dincolo de cuante*, traducere din engleză de Walter Fotescu, Humantias, București.

Tegmark. M. (2016). *Universul nostru matematic. În căutarea naturii ultime a realității*, ediția a doua revăzută și adăugită, traducere din limba engleză de Dumitru Dorian, Editura ASCR.

Yang, C. N. (Ed.). (2005). *Selected Papers (1945-1980) Of Chen Ning Yang (With Commentary)* (Vol. 36). World Scientific. https://doi.org/10.1142/5853

Revista Română de Filosofie Analitică,
Romanian Journal of Analytic Philosophy
Vol. XVI, Iulie-Decembrie 2022, Nr. 2, p. 25-44
ISSN (ediția electronică): 1843-9969,
ISSN (ediția tipărită): 1844-2218
DOI: 10.62229/rrfaxvi-2/2

PHENOMENAL HOLISM AND QUALIA CATEGORIES

MIHAI ALEXANDRU BÎCLEA[1]

***Abstract*:** Scientists have attempted to find consciousness and, more specifically, qualia in the physical world ever since philosophers such as Thomas Nagel (1974) or Frank Jackson (1986) have commented upon the elusive experiential properties of such mental states that are characterized by a specific "what's it like". One of the proposals to minimize the metaphysical and epistemological tension that arises once the existence of such phenomena is acknowledged originates in the influential paper "What is it like to be a bat?" (Nagel, 1974): the development of an objective phenomenology. Current research programmes follow this idea and aim to understand consciousness using mathematical-empirical models. However, these endeavors seem to be missing the point when studying consciousness because they do not provide any evidence about how qualia correspond to neural states. I argue against a proposal to account for the missing link between physical structures and qualia, namely the use of category theory (Tsuchiya et al., 2016). Instead, I conjecture that the endeavor is futile because it relies on the assumption that qualia can be described structurally from an epistemological point of view. I support my conjecture arguing that phenomenal holism has not been ruled out.

1. Introduction

Scientists have attempted to identify consciousness and, more specifically, qualia in the physical world ever since philosophers such as Thomas Nagel (Nagel, 1974) or Frank Jackson (Jackson, 1986) have commented

[1] Mihai Alexandru Bîclea is a graduate student in the „Research in Theoretical Sociology" master's programme at the University of Bucharest.

upon the elusive experiential properties of such mental states that are characterized by a specific "what's it like". One of the proposals to minimize the metaphysical and epistemological tension that arises once the existence of such phenomena is acknowledged originates in Nagel's influential paper "What is it like to be a bat?" (Nagel, 1974): the development of an objective phenomenology. While the author did not expand on what he meant by such a project, it seems that recently his idea reemerged in the context of scientists who want to use mathematical concepts in order to explain consciousness. (Fekete & Edelman 2011; Oizumi et al., 2014; Tononi et al., 2016; Kleiner, 2020). Maybe the most influential research programme among these is The Integrated Information Theory.

The Integrated Information Theory (IIT), first proposed by Giulio Tononi (Tononi, 2004), tries to account for the phenomenological cross-experiential properties of consciousness in order to find physical correlates for consciousness. It aims to minimize the range of possible mechanisms that could implement consciousness by looking at certain axioms related to the general characteristics of qualia and by excluding every physical thing that cannot be conceived so as to account for all of them. The last version of the theory identifies five axioms that can be briefly summarized as follows:

1. **The axiom of intrinsic existence**: consciousness has an intrinsically real and actual existence.
2. **The axiom of composition**: consciousness is structured; one can discriminate between different qualia at a given time.
3. **The axiom of information**: consciousness has a specific existence that is different than any other of its iterations.
4. **The axiom of integration**: consciousness is unified and it cannot be reduced to any non-independent parts.
5. **The axiom of exclusion**: consciousness is definite in terms of content and speed of being perceived. (Tononi, 2015)

From these, theorists derive five postulates that, taken together, describe the properties that a physical system should meet in order to be conscious. I will not mention them here, as they are not necessarily

relevant for our discussion. I will go further and describe a similar approach, the attempt to find the neural correlates of consciousness.

Searching for the neural correlates of consciousness (NCC) is an endeavor that aims to look at the minimum conditions that should be met at a physical-neuronal level so that consciousness could accompany a given mental state (Northoff, 2014; Crick & Koch, 2003). NCC typically looks at perceptions or intransitive conscious states like wakefulness or sleep as it is more difficult to account for the contents of conscious states, namely for qualia, especially if you cannot define, isolate and differentiate between them properly in an experimental setting. It is similar to IIT as its main goal is also trying to connect the physical level with the phenomenal level,[2] our qualia states. The only difference is that NCC starts from the physical level of analysis, namely the neuronal events, while IIT starts from the phenomenological level of analysis, namely the way it is for us to undergo certain experiences.

These approaches have arguably made significant contributions to our understanding of either one of the levels, or the other (for example Haun & Tononi, 2019). Nevertheless, it seems that neither have made any definitive progress in accounting for the connection between neuronal events and qualia. This inter-level relationship is still not clear because both IIT and NCC lack an a priori account of causality in their models between the two levels and about how each level, especially the phenomenal one, can be best described in more systematic terms. Following a distinction made by Tsuchiya et al. (Tsuchiya et al., 2016), it seems that the main limit of these approaches is precisely this one: they lack the theoretical foundations that would allow for more research in the areas of either reducing one level to another, or of finding interactions between the two levels in question. Tsuchiya et al., more specifically,

[2] I use the word "level" and the distinction between different levels corresponding to the implementational part of a system, namely the physical substrate, and the representational part, which I take to be in this particular case the phenomenal component, similarly with the terminology introduced by Marr (Marr, 1982). I find it a helpful tool to make sure no confusion arises between the analysis of different dimensions.

reason that the theories do not include structural characterizations of phenomenal consciousness. Without this sort of characterization, it would be, according to them, impossible to describe and operationalize empirically how a certain quale is instantiated in a physical structure, as opposed to other qualia. So, is "quale" a genuinely non-referring term?[3]

The solution Tsuchiya et al. provide is to apply the structure typical to the physical level at the phenomenological level of analysis. They argue that this can be done by identifying quale instances and their relationships with one another in terms of degrees of similarities, and conversely dissimilarities, so that these relationships can be, in turn, understood as part of a structure out of which we can define certain qualia based on the relationships that are postulated between them and all the other qualia. The proposal once again makes use of mathematics, and, more specifically, category theory (Tsuchiya et al., 2016, 2021; Tsuchiya & Saigo, 2021). The resulting mappings between either intransitive levels of consciousness, or transitive phenomenal contents, would help in bringing us closer to mapping conscious mental states structurally to neural states.

In this text, I argue against this proposal. It seeks to account for the missing link between physical structures and qualia by imposing a structure upon these kinds of phenomenal contents, by appeal to category theory. Firstly, I argue that this endeavor is not achievable, nor that it is completely compatible with IIT axioms. Secondly, I argue that the project seems futile as it relies on the assumption that qualia can be pragmatically described structurally from an epistemological point of view. In doing so, I would briefly mention how proponents would apply category theory to the study of consciousness, I would discuss their assumptions and argue against one of the premises of their model by introducing a counterargument based on a phenomenal holism thesis. The problem of compositionality that arises, I would contend, renders the whole endeavor moot.

---

[3] For a discussion on negative existential claims and empty names, see Dumitru, M., and Kroon, F. (2008). What to say when there is nothing to talk about. *Crítica (México, DF), 40*(120), 97-109.

2. Structural Qualia

Tsuchiya et al. (Tsuchiya et al, 2016) are empirically inclined to argue that a quale is a representational property of a given experience. This is easily explainable since, by establishing an identity theory between content that is represented mentally and an external experience that is highly influenced by numerous factors, one essentially can hold true that the nature of the apparent metaphysical content, namely that of the properties associated with the experience, supervenes on the actual physical properties of the entities which make the experience possible in the first place. This kind of externalism, it seems, is the first premise that guarantees the development of a phenomenology objective enough as to account for qualia with the tools that science has gathered up to this point in time.

The second premise, which I will discuss more, seems to be at least partially derived from the first one. By saying that qualia can be characterized intentionally by making use of the state of affairs in the external, physical world, and by acknowledging the fact that this physical world is essentially a structured one, one can argue that qualia, in a similar manner, can be characterized structurally. The reasoning is valid, and the premises seem to be true as far as our intuitions hold. In fact, as we have previously seen, one of the axioms of Information Integration Theory is the one related to the compositionality of qualia (Tononi, 2015). This, however, is more related to the fact that we can identify different qualia instances at a given time and derive from them the idea that our overall conscious state is made up of these individual instances subsumed. It seems that there might be another way to look at the idea of a structure which would face certain epistemological limits. But, first, let us understand the way in which category theory would actually account for qualia or, better said, to analyze what kind of a structure is imposed on qualia.

3. Consciousness and Category Theory

Category theory is a mathematical framework that allows for formalizing and comparing the relationships between objects originating in the same

categories, but also the relationships between objects that do not share any categorial origin. These are calculated based on the degree of analogousness in terms of the relationships that each object has in their original category. I would not go deeper into the mathematical proofs and formalities; I would just mention the relevant aspects related to the framework, intuitively. If a relationship is found in terms of similarity, or analogousness, then there is a functor that can preserve the structure of any of the categories based on the other category. This brief description of the theory already suggests that if we were to apply the model to the study of consciousness, we would be able to settle if a quale instance that one person has is similar to the one of another person, solely based on the relationships that each has in terms of similarities with their own similar qualia, which we take generally to be, at least for this case, accessible to introspection. Similarly, it seems that these qualia objects, if we take them as such, could be also mapped with the objects of other categories, for example the ones corresponding to the neural underpinnings based solely on the configurations each of these categories has between their objects. This is essentially the goal, to aid current theories of consciousness such as IIT or NCC to account for the connection between phenomenal states, on one hand, and physical states, on the other hand. But, before going on to explain how finding a functor across these categories would actually work, one has to check whether the domains in question can be thought as categories in the first place.

There are two ways in which we can apply the category framework to the study of consciousness. On one hand, we can consider the states that fall under the scope of intransitive consciousness, namely those found in wakefulness, sleep, or coma, as objects, under the category of, say, degrees of consciousness. On the other hand, we can apply it to the quale instances as part of the category of transitively conscious states. The latter approach is the one that I find has more explanatory power for the development of the mapping between different domains as previously discussed, which is why I will focus mainly on it while only briefly describing the idea of levels of consciousness as categories.

4. Conditions for a Category of Qualia

According to Tsuchiya et al. (2021), the conditions that a collection of objects has to meet in order for it to be called a category are the following:

> 'Definition: For a collection of objects to be considered as a category, they must satisfy the following five axioms.
> 1. An arrow has its "source" object called domain and "target" object called codomain.
> 2. For every object, there is a self-referential arrow called identity.
> 3. A pair of arrows is composable if the domain of one arrow equals the codomain of another.
> 4. Identities do not change other arrows by composition.
> 5. Composition is associative.' (Tsuchiya et al., 2021)

The authors illustrate how we can conceptualize an arrow *f* between two objects representing two different degrees of consciousness so that every condition is met if the meaning of the arrow is understood as "higher or equal". Essentially, what they do is associate numbers with the said degrees of intransitive consciousness, and to prove that the function *f* meets the conditions for composition, associativity, and unit, for the numbers assigned to the wakefulness levels in the domain and the codomain. There is not much that can be argued against here, as there is no difficulty in imagining that these kinds of states are part of a continuum ranging from 0, when the individual might be dead or in a vegetative state, to an upper limit that would be ultimately a fully wakeful state. What could be mentioned, however, is that the assignment of values would be difficult without referring to the physical conditions in which the agent finds himself. This would go against the initial idea to determine the relationships solely inside a category and then apply them to the category of neural states by using a natural transformation. Additionally, the upper limit of this continuum might not be easy to settle, especially if we consider the states of different species about which we do not have generally the intuitive anthropocentric assumption that their states are similar to ours, more specifically that they are as wakeful as we are. This

would be influenced as well by the cases in which the experience of being awake would feel different from an individual to the other not only in terms of the intensity. This brings us to the second interpretation, that of the category of transitive consciousness consisting of qualia objects.

Tsuchiya et al. (Tsuchiya et al., 2016) talk of a different function in this case, that can be understood as "similar" and that can be mapped between three objects representing three instances in which an individual sees the color red in three different objects. They go on to argue that these representational contents of qualia can be understood to be part of a category as well, mainly relying on the function's nature, that is understood isomorphically to correspond to the one of equivalence in mathematics. It seems that for the sorts of things that its objects are, namely quale instances, the function does not work as clear-cut or as objectively as they picture once we think how it would be applied in real life, in an environment that is not controlled. For this, we can also remember the fourth Information Integration Theory axiom that states that consciousness is unified and irreducible (Tononi, 2015). If we think of the contextual qualia that might influence the particular perceived quale content which we compare with others in terms of similarities, then it would seem that compositionality would be a problem that would not allow us to compare quale instances independently of context. In order to explain this, I would develop the example given by the authors with the three red objects: the sunset, the crayon, and the wine.

5. Contextual Qualia

We do not have a single qualia instance at a given time. What might trick us into believing this, is that we seem to be able to change the introspectable access and discriminate in certain cases between such quale that compose an overall conscious state at a given time. For example, when we focus on the redness of a sunset, we also perceive what might be in the background: the shape of the sun, the light, the warmth, the other colors, maybe the blueness of the sky that contrasts the other chromatic properties of the landscape. The same applies to the redness of

the crayon: we might perceive its size, its shape, its texture, and what is near it, for example if it is in a pencil case along with other pencils and crayons that have different or similar colors. In the case of the red wine, we might feel additional quale based on our previous experiences related to how its consumption made us feel, the taste it had, but also how the brightness of the room was.

This is compatible with both representationalism and externalism. It is also what the axiom of integration and combination, taken together, argue: consciousness is structured as long as we can identify and discriminate between different qualia, but, at the same time, we cannot explain a conscious state purely as the sum of all discriminated qualia that we have access to: integrated into one, the nature of the content changes. We do not have a single quale associated with a single experience, but we have a couple of them, all integrated, unified into one representation, and all influencing each other up to some extent, by virtue of being part of our conscious global state, at a given time. There is no doubt that we can pinpoint the redness of each of these objects, but it seems that this cannot be done without subtracting the influence of the other chromatic properties that we perceive in the vicinity of the objects or even the other cross-modal quale that might affect the way in which we perceive the redness. This brings us to the phenomenal holism thesis.

6. Phenomenal Holism

There are two approaches that can be broadly taken when we are talking about the idea of unity of consciousness, along with the idea of a structure. On one side, we can consider that the global conscious state, that we have at a certain moment in time, is made up of independent units. On the other side, we can argue that it is made up of interdependent units. If we take the axioms of composition and integration as granted, then one of these views should logically follow. This has to do with the type of structure that is imposed on qualia. The first point of view can be considered as a thesis of atomism, while the second one can be

understood as a thesis of holism about conscious states. Tim Baynes summarizes this as follows:

> "Theorists who adopt an atomistic orientation assume that the phenomenal field is composed of 'atoms of consciousness' — states that are independently conscious. Holists, by contrast, hold that the components of the phenomenal field are conscious only as the components of that field. Holists deny that there are any independent conscious states that need to be bound together to form a phenomenal field. Holists can allow that the phenomenal field can be formally decomposed into discrete experiences, but they will deny that these elements are independent atoms or units of consciousness." (Bayne, 2010)

The proponents of *category theory* for mapping qualia structurally seem to endorse the atomistic view, because they do not talk about any context or any variation across experiences perceived in terms of quale instances. They implicitly assume that seeing redness when looking at a sunset and seeing redness when looking at a pencil, both can be compared in terms of degrees of similarities, without acknowledging the other possible factors which might influence the particular experience of looking at these particular objects in separate contexts, or at separate times. By taking into account the possibility of having some other factors as part of the global conscious state, which may affect the way in which we see redness, one adopts a holist or a context-dependence view.

Visual illusions are a good example for illustrating how one could argue for the holistic approach, as opposed to the atomistic one. We are familiar with how certain visual configurations of colors and shapes can trick us into having certain global representations, unreachable by division into smaller parts, like pieces of a puzzle, and by experiencing each on its own. The exact sum of all the micro-representations that we could derive from a macro-representation, would not be equivalent to the macro-representation itself. This can be explained in virtue of how each piece of puzzle gains a novel information once associated with other pieces, namely the way in which all relate to each other, so as to give birth to the bigger image. Going back to our example, it is conceivable to say the redness of the crayon is similar to the redness of the sunset, not only

because of their individual contents, but also based on what other quale we have or have had while perceiving them, in the context of a global conscious state. Based on the color configuration of the background, we are typically tricked to say that the same color instantiated in two places is different, as a result of the other perceptions that accompany it.

We deal with a different level of complexity if we take into account cross-modal perceptual interferences. One famous example is the "McGurk" effect (McGurk & MacDonald, 1976) which proves an interdependency between visual and auditory stimuli. Another example can be the "parchment skin illusion" (Jousmäki & Hari, 1998), which has found an interdependency between auditory and tactile stimuli. This kind of illusions can arise based on the idiosyncratic ways in which our species integrates different perceptions, but this does not prove our argument wrong: that the qualia arising from these perceptions, can be intuitively thought to be dependent on the global conscious state that they are part of.

This view creates a problem for the approach of category theory, because it seems that the identity of a given quale instance is not stable or context-independent. In fact, using the mathematical terminology, it seems that the identity changes once composed with another representational content. To take a more familiar example, it seems that certain moods affect the way in which we perceive certain external things. With the example of the red color it might not be as easy to notice if and how our quale of redness changes when we are sad or in a negative mood, when we are happy or in a positive mood, but if we think of listening to Beethoven's Moonlight Sonata, then it might be more intuitive to say that the representation of our experience would be substantially different, depending on the mood we were in when the experience happened.

Based on all previous considerations, I would formulate the phenomenal holism thesis in a similar way to how it was previously articulated (Dainton, 2010):

> (**PHT**) Two phenomenal contents perceived in a single state of consciousness are impacted in a significant way as a result of being perceived in a single state of consciousness.

It can be understood in a narrow form, if we look solely at the how a local quale and a global conscious state are influenced bidirectionally. If we look further, we can understand the impact in a broader sense: essentially every local quale is affected by all the other local quale because these are all affecting the perceived global state. Thus, it seems that if we accept this thesis as opposed to the atomistic one, the fourth condition, the one related to how identities should not change other arrows by composition, would not be met. My argument can be summarized as follows: supposing qualia can only be understood structurally from a holistic point of view, namely one that does not minimize the dependence effect across different perceived instances and contexts, then it would not make sense to think of qualia as classical objects in category theory because they would not have a stable identity once they are composed in different configurations. Moreover, even if we were to classify them as such by adopting a special enriched category as it has been proposed (Tsuchiya et al., 2021), the similarities between quale instances taken outside of their original contexts, would not suffice to create an objective framework, since the postulated relationships would entirely correspond to similarities perceived only for the quale instances in question. They would also incorporate the contextual influences, so the model would need an additional component, subtracted from every similarity degree reported, a component related to the interferences caused by our different conscious global states.

These do not seem to be something that can be accounted for, considering the fact that we only perceive conscious unities made up of some components that are always discriminated in relation to the global state of affairs. In other words, it seems that if we take the IIT axioms for granted, and in addition accept the phenomenal holism thesis, then we would not be epistemologically equipped to argue that the redness of a sunset and the redness of a pencil are similar, purely based on their local or atomical properties, without taking into account the interdependency between them and the other perceived qualia "atoms" that each could be said to shape each other, up to some extent. Any attempt of finding equivalence between quale instances, which is at the core of using category theory to map the qualia relationship, and to correlate them with

the relationships between physical events, would include a degree of error - not because of the subjectivity of the one who perceives and reports the qualia similarities, but because of their inability to recognize whether a quale is in itself similar with another one, or if it is similar by association with the global conscious state that one finds oneself in.

7. Epistemological Constraints

Another way of stating the phenomenal holism thesis, following the distinction made by Pitt (Pitt, 2018) between ante and post-hoc compositionality, is to say that qualia are not ante-hoc compositional: their phenomenology in context cannot be composed of the phenomenologies its representations have out of context, which is why their identities in context, or in composition, cannot be composed by summing the identities that create the context, as they would appear outside or in no context. Everything needs to be put into a context, every quale depends on the other ones that are perceived at a given time. The authors' proposal for a functor that aims to make a correspondence between objects based on perceived similarities, has to account, as I previously mentioned, for why the quale are said to be similar, but this task is difficult. We do not know how much the redness of a sunset is similar to the redness of a crayon, or a wine, without taking into account the environmental factors that might generate a setting capable of influencing the way in which we perceive the same redness instance, differently, in two different spatial-temporal contexts. This inability is similar to the one we have when we are exposed to an illusion, when we are informed about the nature of the illusion and about the fact that we have been tricked, and when we cannot escape the way in which we perceive the particular illusion, even if we do not know that it is not in the same way that it appears to us.

The only escape, it seems, if we want to apply category theory to the study of consciousness, would be to either think of global conscious states as categories composed of local quale objects, or to choose global conscious states themselves as objects part of a wider category of

intransitive conscious states. The first option would not have any utility in providing the missing link between the phenomenal level and the physical level because of our current state of methodology that needs a universal structure similar to the one of the neural mechanisms, rather than a fragmented, context-dependent structural characterization. The second one, it seems, would be hardly feasible. Instead of taking the redness of a sunset as an object, one could take it, along with the other qualia influencing the perception of the redness of the sunset, as an object that is not decomposable, or easy to individuate. By trying to account for all the factors, however, we would be getting at the global conscious state, since every perceived factor would be caused or influenced by another factor. It would be difficult to weigh in all the representational influences that have shaped a given quale, without regressing to the totality of the representations that we have access to at that given time. This would be another epistemological limit, especially if we accept the fact that we do not have access to all our qualia.

Surely, one could argue that as long as we can find similarities between different qualia, seeing the redness of a sunset and seeing the redness of a pencil or a wine, then there should be something common between them, which is context independent. Nobody denies that qualia cannot be stripped out of their interdependencies so that a pure content could remain. However, this pure content could not be taken as the basis for applying category theory, as this would not capture the reality of how qualia are presented to us. It would not run counter to the problem exposed, as we can identify certain elements as similar even though they are partly influenced by different factors and contexts, but we still could not point out how much the difference is related to the way the quale is perceived in itself, and how much the quale perceived is dependent on or influenced by other factors. We would not have the capacity to approximate this common ground between, for instance, seeing redness in different spatial-temporal settings.

Other than arguing against the phenomenal holism thesis, which as I have presented seems to be consistent with certain intuitions and perceptions that we typically have, the proponents of category theory could reason that category theory helps us in precisely identifying the

common element of a content across contexts. In order to explain this, let us suppose that the problem of compositionality does not exist, and accept the fact that qualia can be understood as objects of a category. We would go on to define the objects, purely on relational terms, but in order to understood how it would pragmatically happen, we seem to have to continue the explanation of the tools provided by the category theory.

A natural transformation, in category theory, is a relationship between functors of different categories that enable each one of them to be translated by making use of the relationships of the objects of the other one. This concept brings us to the Yoneda Lemma, from which we can derive in an intuitive form the following sentence: an object A of a category X is equivalent with an object B of a category Y, if the relationships that A has inside category X are equivalent with the relationships that B has in category Y. In other words, if we apply this idea derived from the Yoneda Lemma to our discussion, my sunset redness can be equivalent with another person's sunset redness if my sunset redness can be described *in such way*[4] that would form a certain relational configuration with the other qualia I have, that is similar to the configuration that the redness of the other person creates, once it is compared with the other qualia the other person has. Provided that the context in which the sunset redness was perceived by the two individuals was the same, then it would be fair to assume that their contents are equivalent.

It does not matter what the representation is, it matters only how the representation relates to the other representations or local qualia and, in turn, to the global conscious state. If one sees the three objects as having different shades of green, then it seems that the similarity-based relationships between these qualia instances would be equivalent with the ones that another individual might have, even though they would perceive the different shades of the objects as red. In other words, all the relationship configurations of a quale with the other qualia, being the

[4] As Quine points out: "what makes a sentence an observation sentence is not what sort of event or situation it describes, but how it describes it" (Quine and Ullian, 2007:39).

same interpersonally, does not entail that the quale itself is the same. In fact, it seems that this principle of Similarity-Congruence is not logically strong enough as to help one deduce whether the type of the quale in question is similar or not (Pautz, 2019). This issue is inescapable even if we describe relationships between qualia by adopting an interval or a ratio level variable assignment. This has been attempted by "enriching" the category theory as to account for more flexible "relationships" in the qualia space that can be mapped in a more nuanced manner, on a continuum. The motivation is that when one maps qualia in a metric space, one fails to account for the phenomenal properties that go beyond the represented points. Tsuchiya et al. (Tsuchiya et al., 2021) introduce a monoidal category called "dissimilarity", to complete the initial framework, but this endeavor fails to address both the problem of compositionality, and the epistemological limits previously mentioned. By assigning numbers to the similarities perceived, not only the congruence would have to be approximated, as it does not seem to be an *inter-individually objective similarity* between any two qualia instances, but also the degree of variation introduced. This does not manage to render the endeavor more objective, because it would depend, maybe in a more significant manner, on the individual subjective ratings and everything else which might influence them, from the degrees of access to introspection, to the range of possible perceived limits between which an experience can be represented as similar or not with another one. Such model might, however, help us find the "noise" s coming from our global conscious experience, because it would allow for comparisons between the same agent's local qualia in different contexts, so that the variations in terms of interdependencies could be closer to being controlled. However, such an approach would be ultimately an atomistic one, because qualia would be perceived as the phenomenal building blocks that would suffer from a certain degree of dependency capable of altering their contents, a degree that can be mathematically subtracted from the parts, so that the content remains in its pure form.

8. Conclusion

I started the paper by describing the most prominent approaches for investigating the elusive phenomenon of consciousness, namely IIT and the NCC. The latter starts at the physical level and tries to infer from the neural events the phenomenal states thought to intervene upon them, while the former follows the exact opposite move. Both have to account for, as Tsuchiya et al. (Tsuchiya et al., 2016) argued, the way in which the two levels relate to each other and how one can possibly implement or cause the other, if they want to offer a more complete framework for research. Applying a structure to the phenomenal level, as the authors go on to suggest, is an idea that could aid in such an endeavor, by delimitating qualia in the same way we delimitate brain regions. It would make it possible for our current scientific tools to map the connection between the two levels.

I, then, presented one of their assumptions, namely atomism, and argued against it by defending the alternative position, namely the context-dependence or phenomenal holism thesis. This can be understood as a view which is entailed by two axioms of IIT – the axiom of integration and the axiom of unity-, which states that two representational contents perceived at the same time, by virtue of being perceived at the same time in an integrated whole, are different than what they would have been if they were perceived each on their own, and then subsumed. If we accept this view, we cannot think of qualia as categories, because they do not maintain their identity while they are being composed.

The endeavor inspired by category theory to map qualia structurally creates a problem because developing such an objective reporting of our qualia does not seem to be entirely pragmatically achievable, especially if we take into account the epistemological limits that would constrain us from identifying how much the similarity perceived between two qualia instances in two different contexts would rely on the local content itself, and how much it would be influenced by the global conscious state, the amalgamation of the phenomenal contents all being interconnected.

I want to end by saying that even though category theory does not help us in laying the foundations for a more concrete study of the nature of the supervenience arising between the two levels mentioned, this does not mean that we should all end up supporting either mysterianism, or dualism. In fact, quite the contrary: we should continue to think of ways in which we can bring consciousness closer to our scientific tools, or, better said, we should bring our scientific tools closer to consciousness[5]. We should not assume that, by default, the characteristics of the physical level could be isomorphically applied to the phenomenal level. We are most certainly slowed down by certain epistemological limits, especially in interpreting the interactions that arise between different levels of analysis that we perceive, but we are also constrained by certain views we have about science. These are the ones that we are more in power to change at present. Applying mathematics to our object of study is most certainly an asset. However, a positivist approach does not always benefit science, especially if we are talking about the study of such a mysterious and elusive, yet utterly familiar and widespread phenomenon, as consciousness.

REFERENCES:

Bayne, T. (2010) *The Unity of Consciousness*, Oxford: Oxford University Press, p. 225.

Crick, F., and Koch, C. (2003). A framework for consciousness. *Nat. Neuroscience*, 6, 119–126. https://doi.org/10.1038/nn0203-119

Dainton, B. (2010). Phenomenal Holism. *Royal Institute of Philosophy Supplement*, 67, 113–139. doi:10.1017/s135824611000007x

Dumitru, M., and Kroon, F. (2008). What to say when there is nothing to talk about. *Crítica* (México, DF), 40(120), 97-109.

[5] It is not entirely clear which the best way to move forward is, when it comes to the study of consciousness – for a discussion on the way this problem weighs on experimentalists and arm chair philosophy see Chapter 8 – "Explicarea conştiinţei fenomenale. Conceptibilitate epistemică şi posibilitate metafizică", in Dumitru, M. (2019). *Lumi ale gândirii: zece eseuri logico-metafizice*. Iași: Polirom. Edited by Andrei Pleșu.

Dumitru, M. (2019). *Lumi ale gândirii: zece eseuri logico-metafizice*. Iași: Polirom. Edited by Andrei Pleșu.

Fekete, T. and Edelman S. (2011) Towards a computational theory of experience. *Conscious Cogn*; 20:807–27. doi:10.1016/j.concog.2011.02.010

Haun, A. and Tononi, G. (2019). Why Does Space Feel the Way it Does? Towards a Principled Account of Spatial Experience. *Entropy*. 21 (12): 1160 doi:10.3390/e21121160. PMC 7514505.

Jackson, F. (1986). What Mary Didn't Know. *In Journal of Philosophy*, 83, pp. 291–295. https://doi.org/10.2307/2026143

Jousmäki, V., and Hari, R. (1998). Parchment-skin illusion: Sound-biased touch. *Current Biology*, 12(8), 190. DOI: 10.1016/s0960-9822(98)70120-4

Kleiner, J. (2020). Mathematical models of consciousness. *Entropy*, 22:609. https://doi.org/10.3390/e22060609

Koksvik, O. (2014). Three Models of Phenomenal Unity. *Journal of Consciousness Studies*, 21 (7-8):105-131.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Henry Holt. ISBN 978-0716715672.

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746–748. https://doi.org/10.1038/264746a0

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, Vol. 83, No. 4, p.436, 449. https://doi.org/10.2307/2183914

Northoff, G. (2014). *Unlocking the Brain: Volume II: Consciousness*; Oxford University Press: Oxford, UK.

Oizumi, M., Albantakis, L. and Tononi, G. (2014*) From the phenomenology to the mechanisms of consciousness: integrated information theory* 3.0. *PLoS Comp*;10: e1003588.

Pautz, A. (2019) What is the integrated information theory of consciousness? A catalogue of questions. *J Conscious Stud*; 26:188–215

Pitt, D. (2018) Phenomenal compositionality and context effects. *Inquiry*; 61:494–8

Quine, W. V., and J. S. Ullian. 2007. *Ţesătura Opiniilor. (*Romanian edition of *Web of Belief)*. Translated by Mircea Dumitru. Pitești, Romania: Paralela 45.

Tononi. G., Boly, M., Massimini, M. et al. (2016). Integrated information theory: from consciousness to its physical substrate. *Nat Rev Neuro*; 17: 450–61. https://doi.org/10.1038/nrn.2016.44

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*. 5 (1): 42. doi:10.1186/1471-2202-5-42. ISSN 1471-2202. PMC 543470. PMID 15522121.

Tononi, G. (2015). Integrated information theory. *Scholarpedia*. 10 (1): 4164. Bibcode:2015SchpJ..10.4164T. doi:10.4249/scholarpedia.4164

Tsuchiya, N., Taguchi, S., and Saigo, H. (2016). Using category theory to assess the relationship between consciousness and integrated information theory. *Neurosci Res* ;107:1–7. https://doi.org/10.1016/j.neures.2015.12.007

Tsuchiya, N., Saigo, H. (2021). A relational approach to consciousness: categories of level and contents of consciousness, *Neuroscience of Consciousness*, Volume 2021, Issue 2, niab034, https://doi.org/10.1093/nc/niab034

Tsuchiya, N., Saigo, H., and Phillips, S. (2021). *Enriched category as a model of qualia structure based on similarity judgements*. https://doi.org/10.31219/osf.io/ucjmz

Revista Română de Filosofie Analitică,
Romanian Journal of Analytic Philosophy
Vol. XVI, Iulie-Decembrie 2022, Nr. 2, p. 45-65
ISSN (ediția electronică): 1843-9969,
ISSN (ediția tipărită): 1844-2218
DOI: 10.62229/rrfaxvi-2/3

EXPLAINING ACTIONS: THE MODEL OF TELEOLOGICAL EXPLANATION AND ITS DIFFICULTIES

NORA GRIGORE[1]

Abstract: The classical model for explaining an action (i.e. for answering the question why someone has done something) usually puts an agent's *desires* at the center of the explanation: it is because the agent desired something that the action took place. Collins argues that an appeal to *purposes* could be a more appropriate explanation, that is, he offers a teleological account of action. In his view, actions could be described as a "compensation" for a perceived lack, for an unaccomplished purpose of the agent, in the way a thermostat or a helmsman brings corrections to a perceived state of affairs. The purpose of my essay is to discuss the difficulties encountered by his proposal. I argue that one needs a clearer account for what may count as "compensatory" in order for the theory to be able to distinguish between a random event, a causal effect and a compensatory action.

Keywords: action, purpose, teleological, explanation, compensatory

1. Introduction

People have desires. These desires are mental events of some sort that cause humans to act. In order to explain their action, i.e. to explain why

[1] Nora Grigore is a researcher at the "Constantin Rădulescu-Motru" Institute of Philosophy and Psychology of the Romanian Academy.

people acted the way they did, we only need to refer to the desire (or desires) that made them act, namely we need to provide an account of the cause of their actions. Once we have the cause, we also have the explanation of why they acted or the reason explaining their action. This is, in its simplified version, what Arthur Collins calls "the standard view of reason-giving", a view that he wants to criticize in his article "Action, Causality, and Teleological Explanation".[2] His own proposal is a teleological kind of explanation for action, one in which the effect rather than the cause is taken into consideration, but not *qua* effect; the intended effect of one's actions explains that person's action not by its being an effect but by its being the purpose and the outcome of the action.[3]

"Teleological" is to be distinguished here from "finalistic"; in a merely teleological explanation an appeal is made to purposes and goals but these are not considered to be efficient causes of the action, as in the "finalistic" explanation. For example, it is a teleological explanation to say that a living organism behaves in various ways because it has its own survival as a purpose, but it is finalistic to say that the purpose of surviving is what causes the organism to behave in that way.

(Naturally, if there were no genuine teleology but only pseudo-teolology (von Wright 1972), then teleological talk would be fictional. For a possible treatment, see Dumitru and Kroon (2006). It is unclear, however, whether this would change anything with respect to teolology. Whether nothing would change were genuine teleology not to exist might itself be methodologically remarkable.)

For Collins, the main proponent of the standard view is Donald Davidson with his famous "Actions, Reasons, and Causes".[4] Whether this is a correct interpretation of Davidson's text will not be the object of the present discussion. Even though Collins' attribution of the standard view

---

[2] Collins, A.W., "Action, Causality and Teleological Explanation", *Midwest Studies in Philosophy*, 9 (1), 1984, pp. 345–369.

[3] Collins has the implicit but quite clear position that he names the same thing with "purpose" and " intended outcome of an action", as a purpose is not supposed to be a mental entity or state. One of his declared purposes in the paper is to get rid of mysterious mental entities.

[4] Davidson, D., "Actions, Reasons and Causes", *The Journal of Philosophy, 60* (23), 1963, pp. 685–700.

to Davidson does not seem especially eccentric, one might become cautious about this step when reading Geoff Schueler's book, *Reasons and Purposes:*[5] the analysis of several hidden assumptions and possible interpretations of Davidson's theses might shake the conviction that one knows exactly what those theses say.

Fortunately, for the purposes of the present research, the matter can be left aside. The purpose of this research is not to examine the dispute between Collins' teleological proposal and the standard view or to take sides in this dispute. The aim is rather to analyze in greater detail the structure of what Collins calls a "compensational" kind of explanations for actions, to see the exact steps presupposed by such an explanation and to address several possible concerns about it. These concerns or possible objections will not be made from a Davidsonian point of view or on behalf of the standard view, even though they might be used by such an opponent, as I will show below. Rather, they will be concerns that can be raised from the 'inside' of Collins' position, i.e. difficulties that one may reach even though one started by accepting Collins' arguments.

I bracket the question of what it might mean to be fully rigorous or precise in providing a teleological explanation. Perhaps, inter alia, natural language has no exact logic: this would raise thorny issues concerning what it might be to genuinely verbally explain phenomena (Dumitru 2019, pp. 187–200).

Such worries aside, a brief way of describing a more concrete picture of the debate may be the following: Collins may be said to take into account here three main types of events; first, a peculiar sort of event outside the area of human action (i.e. what he calls "homeostasis"), second, a peculiar type of actions that strikingly resemble homeostasis and third, actions in general. For all three types he will claim, accompanied by suitable qualifications, that the "compensatory" type of explanation may apply. The characteristic of "compensation" seen in this manner is that the outcome of an activity (belonging or not to a human being) is pursued *in spite of* or *against* various obstacles (actual obstacles

[5] Schueler, G.F., *Reasons and Purposes. Human Rationality and the Teleological Explanation of Action*. Oxford, Clarendon, 2003.

or merely possible ones, as the use of counterfactuals shows) to be found in the environment. It is clear what is compensatory about an activity like homeostasis where an organism or a mechanism is set to 'come back' to a certain state, for example in order to maintain a certain temperature; it is also intelligible and intuitively plausible to call a certain activity like the one of a helmsman[6] "compensatory" (permanently 'correcting' the otherwise erratic course of the boat). But it requires a bit of imagination to see what can be called "compensatory" in an ordinary activity like turning on the lights. I will come back to this third case and its plausibility in the third section below.

For now, if we accept that all three types of activity can be seen in this manner, of compensating for the 'threats' or 'lacks' that the environment poses or might pose to a desired outcome, then a certain pattern can be found for all of them: to a multitude of events in the environment always corresponds a single same event which is the wanted outcome. For example, to various exterior temperatures – the same temperature (or range) maintained by a body, to various waves and wind blows – the same course maintained by the helmsman, and to various possible obstacles in the way of this action – the same turning on the lights victoriously.

In such cases the problem is: *how does one know that this is not a coincidental correspondence* or a conventional one like in the case of a function which yields the same values no matter which various arguments it takes? This problem brings other problems with it: if compensatory mechanisms include such cases of coincidental correspondence, then can we speak of teleological explanation for these cases?

My main task in this essay will be to provide an answer to the above question in the sense of identifying the condition that would exclude coincidence from compensatory mechanisms. Nevertheless, the answer will prove to be difficult to apply to all the three above mentioned types of activities.

The first section of this essay will present the outlines of Collins' text and his main arguments. The second section will consist in a detailed

[6] This is Collins' example.

presentation of the problem and the third will consist in discussing the problem and its meaning for the supporters of the teleological explanation of actions.

2. Collins' text

Arthur Collins argues for the viability of teleological explanation for human actions as opposed to what he calls "the standard view", namely the view that provides explanation of action by appeal to beliefs and desires seen as mental events causing the action. "Teleological explanation" is defined as "any explanation that derives its explanatory force from appeal to outcome, goal or objective of what is explained".[7] Collins agrees that mental events and causes and even mental events *as* causes may be discerned with respect to human action, but these do not play an essential role in the explanation of action (even though they might play an important role in the production of the action):

> "In the interpretation of reason-giving put forward here, I press for the elimination of any role for the fact (where it is a fact) that the agent wanted to attain the objective reference to which explains his action. Of course, I do not deny that agents commonly do want to reach the objectives that their actions do reach. The teleological interpretation removes reference to this antecedent desire in favor of reference to the outcome itself. The thesis that I called the standard view of reason-giving (...) regards antecedent desires and beliefs about their possible satisfaction as the very crux of reason-giving."[8]

According to Collins, the opinion that causes should be involved in reason-giving for actions comes from conflating two questions, namely the question "How is it that men are able to give reasons for their actions?" with the question "What does a man say about his action when he gives a

[7] Collins, A.W., "Action, Causality and Teleological Explanation", *Midwest Studies in Philosophy*, 9 (1), 1984, p. 347.
[8] Collins, A.W., "Action, Causality and Teleological Explanation", p. 364.

reason for having performed it?". Consequently, reason-giving would not usually involve mentioning causes but outcomes, even if this is not the whole story. The outcomes should be accompanied by the presence of a certain disposition to remove obstacles in attaining the outcome, namely by the "compensatory aspect":

> "I do not assert that the mere fact that an action has a certain outcome will validate an explanation averring to that outcome. We must believe that the agent was disposed to compensate for some obstacles, at least, had the outcome not occurred. Action has a compensatory aspect that is entailed by reason-giving explanations but is not legible from the outcome alone."[9]

Why the compensatory aspect should be present, and how it is connected with teleological explanations, will be detailed below. For now let us note Collins' conclusion:

> "Reasons explain actions by referring them to their effects and to the compensatory character of behavior vis a vis those effects. In light of the availability of this interpretation, there is no foundation at all for the expectation that reason-giving explanations may *also* refer to the causes of what they explain."[10]

The strategy employed to support the above conclusion rests on a paradigmatic example of teleological explanation, i.e. homeostasis. The characteristics of this example, (chiefly among them: the compensatory character) are said to be found in two other types of events: first, in special kind of actions and then, enlarging the sphere, simply at large in ordinary action.[11]

---

[9] Collins, A.W., "Action, Causality and Teleological Explanation", p. 364.

[10] Collins, A.W., "Action, Causality and Teleological Explanation", p. 363.

[11] It must be said that Collins does not make a detailed or systematized analysis of what/which these shared characteristics are or are supposed to be, even though their presence in all three types of activities is supposed to count as a proof of the teleologically explicable character of actions, i.e. the point of the matter. Rather than giving the characteristics, Collins seems to count on the intuitiveness of the examples he gives. An attempt at deciphering the structure of his paradigmatic example and the characteristics

Collins asserts that even though teleological explanations may have different kinds of objects (e.g. regularities, particular events, possession of organs) he chooses homeostasis (from the group of particular events) because it helps underline the importance of outcomes in reason-giving. Therefore, the first kind of events under discussion are instances of compensatory activity, like maintaining a stable body temperature.[12]

The same phenomenon (of compensation or pursuit of the outcome), Collins claims, may be easily noticed in the case of certain human enterprises and therefore a parallel can be drawn between teleologically organized systems and at least some of our ordinary actions. The parallel works with the exception of one point: the relation between the triggering event and the compensating event does not need to be causal in the case of human activity:

> "Whenever currents, swells, or wind would move the boat from the given heading, the helmsman acts so as to maintain the constant outcome-state. Here we find exactly the relationship that obtains in a physiological system with homeostatic compensatory activities, except for one point. In the exposition of physiologically based homeostasis, we required that the environmental event needing to be offset be causally related to the compensating event, which is then explained teleologically as occurring in order that homeostasis be maintained.
> In the context of action-based homeostasis, this relationship between the outcome-threat and the compensating occurrence is uncertain. ... Insofar as we are in doubt about the relationship of action and causality, we cannot simply claim a causal relationship here [between the shift in the wind and helmsman's action]."[13]

Collins will claim that the possible lack of a causal chain between the triggering and the compensating event is a bigger problem for the Davidsonian point of view than for him:

---

it shares with the cases of human action will be undertaken in section three of the present essay. This section is dedicated to following Collins' own steps.

[12] He accepts that his account of homeostasis might be oversimplified because it is never clear within what range an outcome counts as "the same". But still, for an important range of cases this kind of example makes obvious the fact of compensation, i.e. the pursuit of a certain same outcome.

[13] Collins, A.W., "Action, Causality and Teleological Explanation", p. 359.

> "This uncertainty is a significant issue for the teleological interpretation of reason-giving. But uncertainty here does not tell in favor of accounts like that of Davidson."[14]

But the exact dialectic of the dispute is only tangentially relevant to the issue at hand. The focus of the essay is the structure that is supposed to be similar to all the three cases presented by Collins as able to receive a teleological explanation because they can be seen more or less as compensatory activities. Up to this point, the first two kinds were presented. Collins' strategy, presumably, is to try to show that the characteristics of the first type of event (i.e. homeostasis and its obvious orientation towards the outcome) can be met in the other two types of events, namely in certain human activities, first, and then in human activities in general.

> "A satisfactory parallel between events in a teleologically organized system and actions depends upon finding something like compensation in ordinary action."[15]

The aim of this strategy is to show that ordinary action may be explained teleologically, without appeal to causes; according to Collins, if something can be seen as compensatory then it supports a teleological explanation.

For actions belonging to the paradigm of the helmsman this did not seem to be a problem, but actions in general are, indeed, a more problematic case because the 'compensation' does not seem al all to be omnipresent and obvious:

> "But most human actions do not offer such good analogies to physiological compensatory activities. ... The difficulty, however, in seeing actions as similar to compensation stems largely from the fact that there is nothing in particular to pick out as environmental menace to a given object and, therefore, nothing for which the action could be viewed as compensating."[16]

[14] Collins, A.W., "Action, Causality and Teleological Explanation", p. 359.
[15] Collins, A.W., "Action, Causality and Teleological Explanation", p. 359.
[16] Collins, A.W., "Action, Causality and Teleological Explanation", p. 361.

Collins' solution to this difficulty is to make a distinction between homeostatic activities and teleologically explicable ones:

> "It is not the case that all organic and machine activity that is teleologically explicable is homeostatic. Enzymes are released in the saliva to bring about the secretion of hydrochloric acid, but the release of enzymes does not keep the value of some organic parameter in a constant normal range on the analogy of temperature control."[17]

Presumably, the form of Collins' argument at this point is the following: all homeostatic events are compensatory and all compensatory events are teleologically explicable; but not all teleologically explicable events are compensatory and not all compensatory events are homeostatic. Ordinary action does not show signs of being homeostatic. But if we can see it as being compensatory, then surely it may be regarded as teleologically explicable. The problem therefore becomes: can we see ordinary action as compensatory in some way? Collins seems to say that we can:

> "One might say that any action that is done to bring about something or to reach some objective compensates for the fact that the ordinary course of events does not bring about that something without help. ... Somewhat more naturally, a kind of compensatory character is detectable in the fact that circumstances sometimes do block the success of the undertaken and ordinarily effective action."[18]

There are two ideas here: first that any action could be seen as a compensation for the fact that the outcome is not present yet (i.e. comparing the actual state of affairs with the desired one, we 'compensate' for the lack of the actual state of affairs by taking action); and, second, that this hidden compensatory character is better revealed when there is an actual obstacle in the way of attaining the outcome. That is due to the fact

[17] Collins, A.W., "Action, Causality and Teleological Explanation", p. 361.
[18] Collins, A.W., "Action, Causality and Teleological Explanation", p. 361.

that the characteristic shared by the other two kinds of teleologically explicable events discussed until now, namely the pursuit of the outcome, becomes in this way observable – overcoming or cancelling various obstacles might be the analogue of the homeostatic mechanism trying to cope with different destabilizing outer events.

Therefore, we would expect someone who wants to turn the light on to do everything that involves removing obstacles. But Collins thinks that more than this is needed to establish that compensation is a characteristic of ordinary action:

> "To establish that action is essentially compensatory, however, we have to go beyond appeal to customary expectations. We have to show that reason-giving actually carries the implication that compensatory actions would have been undertaken had the explained action failed."[19]

To show that reason-giving implies compensatory actions in case of failure, Collins turns to the notion of "pro-attitude" he borrows from Davidson. The pro-attitude is the disposition to perform various actions that would bring about the desired outcome: "Any one of an indefinitely large number of actions would satisfy the want and can be considered equally eligible as its object".[20]

If reason-giving involves pro-attitudes and pro-attitudes can be equated with the phenomenon Collins calls 'compensation', then there is a stronger case for considering that ordinary action usually involves compensation:

> "Thus the concept of reason-giving explanations of actions reproduces the essential features of teleological organization that we found to account for the intelligibility of explanations that cite effects rather than causes. An explained action is referred to its objective or goal, and reason-giving explanation implies the kind of compensatory plasticity upon which the analysis of teleology was found to depend."[21]

---

[19] Collins, A.W., "Action, Causality and Teleological Explanation", p. 362.
[20] Davidson in "Actions, Reasons and Causes" p.6.
[21] Collins, A.W., "Action, Causality and Teleological Explanation", p. 363.

3. The problem

What is important for Collins' thesis is the role of the constant outcome (constancy is the element that shows the importance of the outcome). The importance of the outcome (which is identified with the purpose) justifies the assertion that the explanation of an action normally does not appeal to causes, but to purposes and, therefore, can be rightly called 'teleological'.

The constancy of the outcome in Collins' examples contrasts with the variation of events exterior to the given system under discussion. This is why a pattern might be said to emerge in the case of what Collins calls the compensatory character of certain events: variability of input (events exterior to the system but affecting the system) connected with constancy of the output (the pursued outcome). Can we call any event corresponding to this pattern or scheme a compensatory event? The answer is not mysterious because in order for Collins' theory to work it must be 'no'. One may find many examples of mechanisms that correspond to the above scheme without displaying anything that might seem 'compensatory' in an obvious manner.

But how can one justify this negative answer? What is the difference, in other words, between an event with compensatory characteristics and one that merely corresponds to the above scheme? What is the difference between a body that keeps its temperature constant and some other sort of system that keeps its temperature constant because it is not affected in any way by the surrounding temperature?

An approximate direction for an answer can be found by merely looking at the paradigmatic examples: the difference between homeostasis and a mechanism that gives the same outcome is that, in the case of homeostasis, there has to be a connection between the variety of external events and the constant outcome. Moreover, only certain external events are relevant for the constant outcome.

My proposal is that, in homeostasis, we assume that the constant outcome is not a coincidence but a *response to* certain exterior events (variations in temperature). Consequently, something must link the exterior event with the internal event (of the body maintaining constant temperature). Something must show that the body *reacts* to the exterior

change in temperature. The only obvious thing connecting the two is the *causal chain*: this is how we assume that the constant outcome is not a coincidence but a response to the outer variations in temperature. Therefore, in the case of homeostasis, Collins has a ready answer to the challenge that mere coincidence is possible (even if that answer is not explicitly stated in his article).[22] But, as it is clear from the previous section, *the causal relation is the exact point of dissimilitude with the case of the helmsman or of switching on the lights:* we cannot assume, says Collins, that there is a relation of causality between the direction of the wind and the helmsman's actions.

Does the problem not re-emerge, threatening the 'compensatory' character of actions by our inability to distinguish it from coincidental constant outcome? The point of my research is not to cast doubt on the teleological explicability of common actions; for the purposes of this essay I will assume that Collins' argumentation in this sense may be regarded as successful. I aim to analyze how, on this account, a teleologically describable system differs from a system that is not teleologically explicable, as well as how this difference is articulated by the notion of "response" I have employed so far.

This might be just a methodological issue, but I find it an illuminating one: accepting Collins' arguments against the standard view does not give much information about his own view. It might be true that there is no causal relation between wind change and what the helmsman does. Perhaps another relation obtains, as our own use of counterfactuals indicates. But this is just to presuppose that some relation obtains, not to

[22] He explicitly agrees with this account in an earlier article, "Teleological Reasoning": "How does S [a system] manage to be goal-directed? We will be unable to answer this unless we can trace the causal connection between threats to G and the occurrence of the compensatory event B in S. To call B "compensation" is to assume that there is such a causal connection. As we saw in the case of neural sweat-control centers destroyed by disease, this assumption is indispensable. If it fails, the teleological explanation will be withdrawn. But a fair correlation of goal-threats and B-like events in S suffices to assure us that a causal connection must be responsible for the observations. In real cases, not just science fiction, we rightly rely on the conviction that causal connections exist though we are unable to trace them." pp. 458–9.

describe that relation, what role it plays or how the whole teleologically organized system is supposed to work.

As the problem of causal *versus* teleological account of action is a well-known and widely discussed one, I should make clear several points about the difference in framework between the problem proposed here and other concerns present in the literature. I am not concerned here with the so called "Davidsonian challenge" (in Arthur Mele's terms), where the main question is "In virtue of what is it true that a person acted in pursuit of a particular goal?".[23] Mele argues that the teleologist cannot answer this challenge and that only a causal story could account for an action having a particular reason and not another. Schueler has a nice account of this challenge:

> "The Davidson–Nagel point here is that, unless we say that my desire to see my friend caused me to head for the coffee house, we can't make sense of the thought that this is what moved me, that this was my real reason for going, rather than, say, my desire to get out of my office. Likewise, we can't makes sense of the fact that I went rather than stayed, since, after all, I had reasons for staying too; e.g., I wanted to get some more work done. I have this whole set of desires, some of which will be satisfied by going to the coffee house and some of which will be frustrated by this action (and some neither, of course). So the explanatory tools available at the level of reasons don't seem sufficient to actually explain my action. So if, as we are supposing, it is only my desire to see my friend that is my real reason for going, there must be something different about this desire that provides it with the explanatory force it has, and what can that be except that it caused me to act where the others did not? To find an explanation, we seem forced to say that this was the cause of my action." (Schueler, *Reasons and Purposes*: 51)

Perhaps, without a causal chain, we do not know which one was the reason a person really acted upon. This might be a legitimate theoretical concern or a thorny problem for the teleologist (Schueler's argues it is not). My concern is a different one. In my setting of the problem, the causal

[23] In Mele, A.R., "Goal-Directed Action: Teleological Explanations, Causal Theories, and Deviance", *Noûs, 34* (supplem.), 2000, p. 280.

connection in teleologically explicable events does not play the role of identifying reasons for action but the role of a differentiating trait: it seems to distinguish compensatory mechanisms like homeostasis from non-compensatory events that have contingently (i.e. merely lucky) recurring constant outcomes. The problem is that in actions where the causal connection is missing, no obvious trait presents itself to differentiate between compensatory and non-compensatory systems.

Collins touches upon the problem of lack of causal relation in case of common action and his thesis, I think, can be summarized in three points:

a) That lack of causal connection is no problem because compensatory character is non-causal:

> "Doubts about this causal relationship are not a serious threat to the teleological interpretation of reason-giving in any case. We required a causal relationship in the case of physiological compensatory activity in order to understand how the right compensation event manages to accompany the right threat. Without a causal relation, compensatory behavior would appear either miraculous or coincidental and, in that case, not really compensation at all. That is why we posit a causal connection between environmental changes and compensatory responses though we are ignorant of the details. In the context of action doubts about the causal character of the relation between environmental changes and compensating actions do not have the same basis at all."[24]

b) Learning what relation holds between environment and the compensating action is unimportant as long as the natural use of counterfactuals indicates that there has to be one.

c) That we can naturally distinguish between a relation of *causing* and a relation of *prompting* (even though he does not elaborate on the latter):

> "We do not suspect that if a compensating action was not caused by events threatening a goal then it is not compensation. Those who hold that actions are not caused do not mean that actions do not help to bring about objectives. Whatever the relation between environmental change and

[24] Collins, A.W., "Action, Causality and Teleological Explanation", p. 360.

> compensating act that relationship will support counterfactuals such as "Had the wind not shifted as it did, the helmsman would not have done what he did". Even in the setting of physiology it is the support of counterfactuals like this and not an actual causal story that is crucial for the recognition of the homeostasis. ... Compensation and teleology could be systematically characterized by substituting supporting counterfactuals for causal connection between threat and compensation. We could then distinguish two ways in which this condition could be satisfied, since both activities *caused by* threats to homeostasis and actions *prompted by* threats to homeostasis offer the needed counterfactual support."[25]

I believe that a closer, systematic look at the paradigmatic examples central to the discussion is a precondition for discussing these points.

4. A critique of Collins' argument

Even though Collins' argument rests on the resemblance of common action with homeostasis, there is no clear description in the text for the characteristics or structure of homeostasis. I think several stages of homeostasis can be distinguished. Some of them can be found in the ordinary action of turning on the lights, while some others seem to be missing. Comparing them might shed some light on the claim that both homeostasis and ordinary action may be regarded as compensational.

Keeping the constant temperature of a body may be said to involve the following distinctive steps in the happy, normal functioning, cases:

a) an initial point t1 where the body is in the "normal" range of temperature, i.e. the temperature that allows it to function optimally;
b) the *repeated* occurrence of an event that threatens to destabilize the system of the living body in this respect, of constant temperature (which would affect its viability): a significant decrease or increase of temperature in the environment;

[25] Collins, A.W., "Action, Causality and Teleological Explanation", p. 360.

c) a subsequent point t2 where the body's temperature starts to be modified (each time) by the exterior event, i.e. it occurs a start towards destabilizing the system;
d) point t3 where the body reacts or responds (each time) with its own mechanisms at the exterior (gradually becoming interior) threat;
e) point t4 where the system is stable again by returning to the constant temperature.

Observation: in order to decrease the chances of a random match between t1 and t4, the process needs to be repeated for various exterior threatening events.

It is easy to see that the case of the helmsman can be made to correspond point by point to these steps: the constant course, the wind, the destabilizing wave, the action and the return to the stable course of the boat.[26]

Most likely, the poignancy of homeostasis comes from its circular structure: coming back to the same state surely underlines that state as the pursued outcome and makes its attaining unsusceptible of being a mere contingency. But this exact feature, of "coming back to the same state", seems to be lost when one analyzes the more common example of switching on the lights (which is supposed to be the paradigm case for common action in general). Surely, in this third case, there is no observable temporal line like the one outlined above, with one initial state reinstated in the end: there is no switching on the lights to which we come back to. But there is, one might say, a search for the constancy of the outcome even if not a circular one. More precisely, if we try to fit this example into the previous scheme then steps a) and b) are missing and we might have an analogy for steps c), d) and e): the initial state and the threatening event are missing but we might have the signaling of a lack

[26] Consequently, a short general scheme can be issued:
a) Stability of the system
b) Potentially threatening exterior event (repeatedly)
c) Beginning of destabilizing the system (repeatedly)
d) Activity of the system in counter-reaction to the destabilizing threat (repeatedly)
e) Reinstating the initial state of stability

(sitting in the dark), the activity (the hand moving) as a remedy and the desired outcome (the light is on).

If this looks artificial, then one might take the re-modeling one step further (in a direction indicated by Collins) and say that the common action can be seen as a cycle where the initial and the final state are a match, but they are not both actual states; one of them should be a merely possible state, the one that the agent wants to attain. Therefore, on this picture, the agent should be seen as registering the difference (or the 'lack' if we want to keep the compensation vocabulary) between the state of affairs that she or he wants to bring about and the actual state of affairs; and after registering this difference, acting to reduce it to null, i.e. acting to make the actual state look like the potential one. The scheme might look[27] like this:

a) possible state to be attained – destabilizing event
b) the lack of a match between the actual state and the one to be attained in various possible scenarios – destabilizing the system
c) action to bring about a) in accordance with each possible scenario from b) – reinstating the stability of the system
d) the result: the actuality of a)

Of course, this scheme differs from the one for homeostasis, but one can still make the case for the presence of some kind of "compensatory" mechanism, one that compensates for a lack and more importantly, one where the match between point a) and point d), in spite of the variations present in b) and d), seems to make a convincing case for calling this "pursuit of a goal".

And now the main question can be asked more clearly: can the above scheme describe a coincidence? In homeostasis, coincidence was excluded by stepwise causation: the threatening state turned from "exterior" into "interior" to the organism, and the organism's activity was

[27] The scheme might look outlandish but I think it is not at all unusual or unheard of. Representing human action as the result of lack, of something missing, was such a powerful notion in the ancient Greek philosophy that the whole cosmos was modeled after it: the perfect being was immobile because movement would have been the sign of a lack or "want" and therefore an imperfection.

removing it clearly in *response* to it. How do we know in this case, of switching on the lights, that the movement of the hand is a *response* to states a) and b) (i.e. the lack of a wanted state) above? Collins's answer seems to be that the "link" we are looking for is the pro-attitude of the agent naturally and readily described by anyone with counterfactuals of the type "Had the obstacle not been where it was, the agent would not have done what he did".[28] This pro-attitude makes sure that there is no random connection between the various environmental conditions and the constancy of the outcome of the action; it is the thing that makes sure that the relevant aspects of the environment are chosen as the ones threatening the outcome and ensures their removal.

One may grant that the constant outcome is not a coincidence because it is clear to us that it is a response to factors in the environment rather than some lucky companion that happens to be present whenever those factors are present. But how did this fact become clear to us and how does this response work? Is it different or not from the kind of response we were counting on in the case of homeostasis? Let us distinguish between the three types of situations that have been discussed until now:

A) the situation in which the co-presence of various stimuli in the environment and a certain constant event is a mere contingent one or depends upon a convention (e.g. a measuring mechanism that would indicate the same figures no matter where or what it measures);
B) the situation in which the co-presence of various stimuli and a certain constant event is not contingent, a causal mechanism is involved;
C) the situation in which the co-presence of various stimuli and a certain constant event is not contingent, a certain agent is involved.

My point here is that it is true that we readily distinguish between the first situation on the one hand and the second and the third on the other hand (like Collins maintains), but we do not exactly know how we

[28] The immediately recognizable truth of the counterfactuals, Collins claims, is enough to convince us that there is a link, there is a connection such that the constant outcome is not an accident (even though we do not know what kind of connection or how it happens).

do that because no clear criteria for distinguishing between compensatory and non-compensatory systems have been provided.[29] My suggestion here was that, in cases B) and C), we regard the constant event as a *response* to the environment and so we exclude contingency. But this hardly alleviates the difficulty. For how do we identify a response? Needless to say, that is an interesting problem even if we accept that we often do distinguish correctly among the three situations.

The notion of response itself might prove problematic because it might cover quite diverse situations. The situation B) might resemble situation C) (as Collins insists) but there are significant differences between them too. The scheme provided earlier for homeostasis made clear the following structure of the entire system: two states ongoing by default collide, one being the initial state of the system, of equilibrium (constant temperature) and the other being the ongoing state of the environment (let us pick increase in temperature). These two states cannot both persist and therefore one starts changing the other, the equilibrium being affected.

No such collision need be registered in the case of action: the initial state from the scheme (i.e. the desired possible state) does not have to be changed or threatened in its continuous existence by the other ongoing state, of the environment that fails to correspond to the wish (i.e. the possible state of maintaining the course of the boat is not necessarily affected by the wind or waves). It would seem that there is an unbridgeable gap between states in this case, but we still connect the two states and call one a "response" to the other.

Collins might claim that the gap is bridgeable by the purpose connecting the two. However, in this way, we return to a somehow mysterious connection. The fact that we affirm the existence of this connection by using counterfactuals does not seem to diminish the problem. Despite this dissimilarity of the two cases, I have called (maybe inappropriately) both reactions against the environment "responses". It

[29] Again, this is not meant to be, properly speaking, a criticism of Collins' text, because he does not seem to be concerned with this problem. It is more an indication of where the research might go.

might be the case that for B) one should rather use the term "reaction" (to the environment) than "response" and that the term should be saved only for cases of actions. Schueler's distinction[30] between function and purpose might be useful at this point. The body keeping its temperature constant might be said to have the function, rather than the purpose, of keeping the temperature constant. Schueler argues that the difference consists in the fact that a purpose is not intrinsically given by the inner structure or causal history of a system. A function, however, is "readable" from the inner physical organization of the system or from its components. The function of the body keeping its constant temperature may be read from the body's organization, but there is nothing in the inner structure of a hand that would give us a clue about turning on the lights.

To illustrate how wide the difference between these two kinds of responses could be (i.e. the difference between the two ways of pursuing an outcome), one can imagine a completely unfamiliar realm, with different natural laws and unknown forms of life. Finding out in this scenario which repeating events are coincidences, which are causally triggered and which are results of an action properly performed by an agent, makes it very clear that the difference is readily registered and important. Once it would be clear that a certain recurring event is not a coincidence, the tests that one would employ to see if an agent is involved in responding or not would be relevant in establishing the criteria we employ for detecting such differences.

In conclusion, I think that even if we accept that most common actions are compensatory and therefore teleologically explicable, this result also needs an account of what exactly can be called 'compensatory' and what not. The contrast may provide a more illuminating account of what 'compensatory' is supposed to be. I have argued for the presence among the traits of a compensatory kind of activity of something like a response to the disadvantages presented by the environment at a certain moment with respect to a certain system, but the notion of "response" seems to be quite a heterogeneous one when taking into account the differences between homeostasis and action.

[30] In *Reasons and Purposes*, pp. 7–8.

REFERENCES

Collins, A. W. (1978). Teleological reasoning. *The Journal of Philosophy*, 75(10), 540–550. https://doi.org/10.2307/2025843

Collins, A. (1984). Action, causality, and teleological explanation. *Midwest Studies in Philosophy*, *9*(1), 345–369. https://doi.org/10.1111/j.1475-4975.1984.tb00067.x

Davidson, D. (1963). Actions, reasons, and causes. *Journal of Philosophy*, 60(23), 685-700. https://doi.org/10.2307/2023177

Dumitru, M., & Kroon, F. (2008). What to say when there is nothing to talk about. *Crítica (México, DF)*, *40*(120), 97–109. https://doi.org/10.22201/iifs.18704905e.2008.1002

Dumitru, M. (2019). Lumi ale gândirii: zece eseuri logico-metafizice. Iaşi, Polirom.

Mele, A. R. (2000). Goal-directed action: Teleological explanations, causal theories, and deviance. *Philosophical Perspectives*, *14*, 279–300. https://doi.org/10.1111/0029-4624.34.s14.15

Schueler, G. F. (2003). *Reasons and purposes: Human rationality and the teleological explanation of action*. Oxford, Clarendon Press. https://doi.org/10.1215/00318108-114-3-411

Von Wright, G.H. (1971/2018) *Explanation and Understanding*, Routledge.

Revista Română de Filosofie Analitică,
Romanian Journal of Analytic Philosophy
Vol. XVI, Iulie-Decembrie 2022, Nr. 2, p. 67-81
ISSN (ediția electronică): 1843-9969,
ISSN (ediția tipărită): 1844-2218
DOI: 10.62229/rrfaxvi-2/4

THE DERIVABILITY GAP – A PROBLEM OF INTUITIONS

SONIA MICU[1]

Abstract: This paper addresses our difficulty in grasping how something like consciousness could emerge from the processes of our brain, aiming to explain the intuitions that underlie this struggle. I start with an investigation of materialism, revealing that its supporters often have deep-seated dualist intuitions implicitin some of the language they use. I then question whether we can warrant the claim that conscious experience is fundamentally different from non-conscious phenomena with respect to causal powers and causal roles. I identify the derivability gap as the reason behind our intuitive struggles, explaining how these intuitions make it difficult to accept a materialist view of consciousness. Lastly, I explain why we hold onto these intuitions. Instead of seeking a socio-cultural origin for our dualist intuitions about the mind-body problem, I suggest we examine our intuitive grasp of the physical world, arguing that our tendency to see things in binary on/off forms extends beyond our grasp of consciousness. I propose that we consider whether the same simplifying principle that aids our understanding of unobservables might also influence how we conceive consciousness. I conclude that our lack of direct exposure to its complexity may underlie our binary understanding of life and non-life, which we extend to the contrast between material and non-material.

Keywords: Consciousness; Hard Problem; Explanatory Gap; Derivability Gap; Illusionism; Meta-Problem

[1] Sonia Micu recently defended her doctoral dissertation in Philosophy at the University of Bucharest.

Introduction

Several theories of consciousness have been developed during the last few years, and despite a clear consensus on which theory has the strongest arguments, materialism seems to be the prevailing view among most contemporary philosophers. In this paper, I will look into the difficulties of intuitively accepting such a view and propose a possible explanation.

David Chalmers (1995) asserted that if any problem can be considered the problem of consciousness, it is the problem of experience. He acknowledged that, when we engage in cognitive activities and perceive the world around us, there is a subjective dimension, a first-personal perspective. Chalmers referred to this subjective aspect as experience. Examples of such experiences include visual sensations like the perception of colors and depth, auditory experiences such as the sound of a clarinet, olfactory experiences like the smell of mothballs, bodily sensations ranging from pain to pleasure, mental imagery, emotional experiences, and the continuous stream of conscious thoughts. All these states share the common characteristic of there being something that it is like to be in them, they are states of experience.

While there is a consensus that experience has a physical basis, there is currently no satisfactory explanation for why and how it emerges from physical processes. This problem of consciousness comes with an 'explanatory gap', a term introduced by Joseph Levine in 1983, according to which an explanation of the physical processes doesn't contribute to our understanding of a subjective experience, such as the way pain feels. As David Papineau (2019) observes, our knowledge of the relation between the physical states and what we subjectively experience doesn't seem to help us overcome the 'dualist' intuitionthat they are simply different states which accompany each other. Our struggles with understanding consciousness are reflected in what Chalmers calls

> " 'phenomenal reports': the things we say about consciousness (that is, about phenomenal consciousness). More specifically, many people make

> problem reports expressing our sense that consciousness poses a hard problem. I say things like 'There is a hard problem of consciousness', 'It is hard to see how consciousness could be physical', 'Explaining behaviour does not explain consciousness', and so on." (Chalmers, 2018, p. 7)

I start by investigating a few comments on materialism which reveal the hard-wired dualist intuitions of both materialists *and* of those arguing against it.

2. Materialists as Disguised Dualists

A particular flaw in the materialist stance that has been highlighted by several philosophers, including Saul Kripke, Joseph Levine and David Papineau. Despite arriving at this conclusion using different arguments, they bring light to the same difficulty in overcoming our own intuitive dualism.

As Kripke (1980) construes it, materialism asserts that a comprehensive (and true) understanding of the world can be achieved through a physical description alone. Mental facts are believed to be inherently and necessarily dependent on the underlying physical facts. According to Kripke, no identity theorist has presented a compelling argument against the intuitive view that this strict ontological dependence of mental facts on physical facts may not be accurate. The thorny dialectic surrounding how essences relate to ontological dependence is chronicled in the exchange between Kit Fine and Jessica Wilson in Fine (2020).

Levine (1983) presented a response to Kripke's argument, acknowledging that his own version of the argument does not directly claim the falsity of materialism, making it a less forceful critique compared to Kripke's. Nevertheless, Levine maintained that if his interpretation was accurate, it still presented a challenge to materialism and better captured the discomfort that many philosophers experience in relation to that belief. From Levine's point of view, there seems to be only one practical route for a materialist to confront this dilemma: it involves rejecting the

fundamental intuition upon which the argument is built. Levine suggested that this would require taking a more radical stance in eliminating the concept of qualia, a step that might be too bold for many materialists.

Papineau's (2011) brought in his own proof that even the most materialist of philosophers are, in fact, intuitively resistant to mind-brain identities, and the proof lies in the terminology they use. In his words, "brain processes are standardly said to 'generate', or 'yield', or 'cause', or 'give rise to' conscious states. These expressions are common currency among many thinkers who will insist that they are no dualists. But the phraseology gives the lie to their denial."(Papineau, 2011, p. 12) He uses the example of water-H2O identities, in which one doesn't 'generate' or 'give rise' to the other, to show that such words give away a different type of perceived relation between two entities, one closer to the one between fire and smoke. In other words, once we state that brain processes 'give rise' to conscious states, we see them as ontologically different.

Each of these three authors puts forward a criticism of materialism, whether it is in invoking type-identity theory without fully applying its rules, in disregarding qualia or in utilizing a dualist-like language. The dialectic is clearly summarized by Mircea Dumitru (Dumitru 2019, pp. 100–116), along with the lucid diagnostic that much of the literature seems to pertain to shifting the burden of proof from dualists to materialists or the other way around.

Before proceeding to discuss the derivability gap, let us first clarify what relationship a standard version of materialism bears to such a concept. Recently, Papineau (2019), in a response to Chalmers (2018), elaborated on illusionism as part of the conversation on materialism. According to him, since most philosophers are materialists who hold that conscious states are either identical to or fundamentally realized by material states, they are inclined to dismiss the problematic intuitions as false. On their views, people are simply wrong in believing that consciousness possesses non-physical attributes. Undoubtedly, there may be an explanation for why people hold these mistaken beliefs, but those

beliefs are ultimately false. Papineau concludes from this that consciousness is indeed acknowledged to exist by materialist philosophers, but people tend to have numerous misconceptions about its nature. Since Chalmers (2018) used the term "illusionist" to classify anyone who rejected the problematic intuitions, he put standard materialists in the same category as those who claim that consciousness itself is an illusion. For Papineau, the views of the two groups are different. The standard materialist perspective is that the intuitions may be illusory, but that consciousness itself is real. Illusionists not only deny the validity of the intuitions but also deny the existence of consciousness altogether. Papineau sees a dichotomy in Chalmers' classification: either one accepts the problem intuitions and holds a non-physicalist realist view of consciousness, or one rejects the intuitions and is labeled an "illusionist" alongside those who deny the existence of consciousness.

Papineau also claims that the essence of the hard problem, as understood in the context described by Chalmers, revolves around the intuitive claims that consciousness is non-physical. For Papineau, the weak illusionist position provides a viable resolution by asserting that these claims stem from false intuitions, and that the main challenge lies in empirically explaining why these intuitions are so strong. One possible interpretation of Chalmers' stance that Papineau offers is that he might be considering the "hard problem" not as a matter of anti-physicalist intuitions per se, but rather as synonymous with the derivability gap. Papineau notes that weak illusionists do acknowledge the existence of the derivability gap, but they do not perceive it as a problem in its own right.

We can conclude from this that the derivability gap is a relevant concept only for those materialists who embrace a standard materialist view, namely that consciousness is real, of purely physical nature. In this case, addressing the derivability gap can serve as proof, as a solution to the hard problem of consciousness. If one classifies as an illusionist and thus claims that consciousness isn't real, then to them there is no derivability gap and no hard problem of consciousness to solve.

2. The Challenge of Identity Statements

Kripke (1980) wrote that

> "when Descartes, and others following him, argued that a person or mind is distinct from his body, since the mind could exist without the body. He might equally well have argued the same conclusion from the premise that the body could have existed without the mind. [He added:] let 'Descartes' be a name, or rigid designator, of a certain person, and let 'B' be a rigid designator of his body. Then if Descartes were indeed identical to B, the supposed identity, being an identity between two rigid designators, would be necessary, and Descartes could not exist without B and B could not exist without Descartes." (Kripke, 1980, p. 145)
> "The final kind of identity, the one which I said would get the closest attention, is the type-type sort of identity exemplified by the identification of pain with the stimulation of C-fibers. These identifications are supposed to be analogous with such scientific type-type identifications as the identity of heat with molecular motion, of water with hydrogen hydroxide, and the like." (Kripke, 1980, p. 148)

Levine (1983) argues that Kripke's Cartesian argument against materialism rests on two claims:

> "first, that all identity statements using rigid designators on both sides of the identitysign are, if true at all, true in all possible worlds where the terms refer; second, that psycho-physical identity statements are conceivably false, and therefore, by the first claim, actually false. [Levine's objective being] to transform Kripke's argument from a metaphysical one into an epistemological one. My general point is this. Kripke relies upon a particular intuition regarding conscious experience support his second claim. I find this intuition important, not least because of its stubborn resistance to philosophical dissolution. But I don't believe this intuition supports the meta-physical thesis Kripke defends – namely, that pyscho-physical identity statements must be false. Rather, I think it supports a closely related epistemological thesis." (Levine, 1983, p. 354)

He argues that the reason we can imagine psycho-physical identity to be true is because we lack the epistemological ground which could deny such an intuition. We have an epistemological challenge, or, as Levine calls it, "an explanatory gap" which makes it difficult to know which statements are true. Levine starts his argument with the following assumption:

> "To begin with, let us assume that we are dealing with a physicalist type-identity theory. That is, our materialist is committed to statements like:
> (1) Pain is the firing of C-fibers.
> On Kripke's general theory, if (1) is true at all it is necessarily true. The same of course, is the case with the following statement:
> (2) Heat is the motion of molecules.
> That is, if (2) is true at all it is necessarily true. So far so good."
> (Levine 1983, p.354)

He then proceeds to describe what he calls 'a felt contingency' about these statements. One can indeed imagine a world in which they are false, but this would have to be a logically impossible world, if such identities are deemed to be necessarily true. Therefore, we would need to explain away our felt contingency, even if it appears coherent to us. This seems achievable for (2), since we seem to be able to imagine heat without the underlying motion of molecules but perhaps produced by a different mechanism. Contingency could rather apply to a statement (2') such as:

> "The phenomenon we experience through the sensations of warmth and cold, which is responsible for the expansion and contraction of mercury in thermometers, which causes some gases to rise and others to sink, etc., is the motion of molecules" (Levine, 1983, p. 355).

Such a solution would satisfy our felt contingency, but as Levine observes, it would not work for (1). The difference that he remarks between (1) and (2) is that what counts as pain is the experience, the sensation of pain, which makes it impossible to separate the phenomenon

from the sensation, as was the case for heat. Therefore, since in the case of (1) our felt contingency cannot be explained away, the only remaining option is to renounce the truth of (1).

What we can conclude from these arguments is that although materialism claims a type of identity, it isn't a typical one. Comparing identity statements regarding pain and other phenomena doesn't prove helpful. If such an approach had been useful, we could gain an understanding of consciousness by comparing it to other phenomena. The challenge we face is precisely that consciousness isn't like any other phenomenon. Pain and other subjective experiences are a special kind because 'what they feel like' is the phenomenon, not just a by-product.

It bears mention that appreciating the felt intuitiveness of how pain might differ from its physical correlates is typically done while presupposing that holism about phenomenological vocabulary is not well-suited to account for how we use words like "pain". For otherwise it would be questionable to draw inferences from identity statements involving the word "pain" to real identities involving real pain. For discussion, cf. (Quine and Ullian, 2007).

3. The Explanatory Power of Functionality

Let us go back to the claims put forward by Kripke. The difference between claims (1) and (2) is underlined by another more significant difference between the two, one that Levine (1983) puts forward as follows. Nothing of fundamental value is left out from the explanation of the identity of statement (2). The same does not hold for statement (1). Levine successfully captures the explanatory force of statement (2) by formulating the statement (2') above. With it, he shows which mechanism brings about the causal functions of heat, explained by our knowledge of chemistry and physics. For Levine, the two statements (2) and (2') exhaust all there is to be understood about the notion of heat: both its essential nature and its causal role.

A functionalist story would claim that statement (1) does the same for pain. It explains the causal role of pain by referencing an avoidance mechanism that goes into effect when C-fibers are excited by certain nerve endings which are in turn excited by an interaction with the environment such as the penetration of skin with a sharp object. The challenge that Levine sees here is that this explanation does not exhaust the notion of pain. He remarks that the qualitative character of pain is an essential part of the concept of pain, while its connection with C-fiber firing remains mysterious.

Chalmers (1995) also addressed the topic of functional explanation when he pointed out that the distinction between the easy problems and the hard problem of consciousness lies in the nature of the questions they pose.

> "The easy problems are easy precisely because they concern the explanation of cognitive abilities and functions. To explain a cognitive function, we need only specify a mechanism that can perform the function. The methods of cognitive science are well-suited for this sort of explanation, and so are well-suited to the easy problems of consciousness. By contrast, the hard problem is hard precisely because it is not a problem about the performance of functions. The problem persists even when the performance of all the relevant functions is explained." (Chalmers, 1995, p. 202)

I argue that the conversation on functionality raises two issues. First of all, the assumption that 'what it feels like' is a defining part of the pain, without which we cannot fully explain it, also starts from a dualist perspective, that there is a physical and a separate non-physical nature to pain. This is an intuition still to be validated. *As with other ideas from this paper, one can either start from the intuition and attempt to explain it, or start from explaining the phenomenon and show that an intuition is not justified. If we decide to use intuitions, we must name them for what they are.*

Second, if we say that other processes can be explained through their functional role while for consciousness such an explanation isn't possible, we might seem to assume that qualia have no causal role in

phenomenal consciousness. I propose we consider a scenario in which what an experience is like plays a functional role in that experience, a role perhaps still to be uncovered. Such an assumption might lead us to the conclusion that an agent might not be fully functional without qualia, that philosophical zombies might lack some functionalities by lacking consciousness.

4. An Alternative View on Dualist Intuitions

Even if given a full scientific account of the functionality of any phenomenon attributed to consciousness, we would still lack an explanation of the experience associated with that phenomenon. This is known as the explanatory gap. Chalmers (1995) acknowledges Levine's (1983) "explanatory gap" between cognitive functions and conscious experience, a gap that needs bridging in order to gain a comprehensive understanding of consciousness. He agrees that merely describing the functions does not provide an explanation for the subjective experience that accompanies them. While conscious experience may play a cognitive role, any functional explanation alone falls short in accounting for the phenomenon. It is possible that exploring functions in depth might lead to insights that contribute to understanding experience, but such discoveries would be additional explanatory rewards. Chalmers suggests that the conventional methods of cognitive science and neuroscience, developed to explain cognitive functions, are insufficient for addressing the hard problem of consciousness.

We might argue that when it comes to the topic of consciousness, the dualist intuition is everpresent. Papineau (2019) argues that Chalmers (2018) already presupposes, by his phrasing, that consciousness is of a non-physical nature, when he describes the hard problem as "why and how do physical processes in the brain give rise to conscious experience?" (Chalmers, 2018, p. 6) and, later, when he introduces a category of gap intuitions as those by which "there is an explanatory gap between physical processes and consciousness" (ibid., p. 12). Papineau explains

that, if one entity 'gives rise to' another, they must possess distinct ontological statuses. For instance, fire gives rise to smoke, but H2O does not give rise to water. To Papineau, these psychological responses are brute intuitions that stem from a preexisting commitment to dualism as the alternative explanation. People's initial conviction that the mind is separate from the brain leads them to be naturally perplexed by the capacity of brain processes to generate conscious phenomena. They are dissatisfied with physiological explanations and wonder why the brain 'gives rise to consciousness'.

Papineau proposes that we consider what he calls the "derivability gap" as the underlying cause for the hard problem. Along with its associated explanatory gap, it is, in his view, the evident explanation for the perplexity we experience in regards to consciousness. As Papineau remarks, Chalmers has consistently advocated that the 'hard problem' and the 'explanatory gap' are both caused by the absence of a priori derivability. He has been arguing that the 'easy problems' of consciousness are 'easy' precisely due to their reliance on functional concepts specifying roles. Processes such as learning or memory can be accounted for because we have an understanding of the functional roles that they play, which enables us to identify corresponding physical mechanisms. Chalmers has also emphasized that the 'hard problem' emerges precisely because phenomenal states are not subject to a priori analysis, and the apparent "explanatory gap" stems from our incapacity to a priori derive the existence of phenomenal facts from our knowledge of physical mechanisms.

Papineau (2020) agrees with Chalmers that explaining 'problem intuitions' is key to a satisfactory account of consciousness. He comments:

> "According to the mainstream view, we think of salt as the stuff that is white, crystalline, granular, with a distinctive taste, that dissolves in water, and is found in the oceans. Now imagine someone who has a fully detailed account of the physical make-up of the world, in terms of the distribution of matter, arrangement of elementary particles, the deployment of fields, and so on. In principle, such a person could arguably put this knowledge together with their prior conceptual grasp

> of salt to figure out that salt must be NaCl, on the grounds that NaCl is the stuff that fits the conceptual requirements for salt—white, crystalline, ... However, we can't do this with pain, say, or with visual experiences of red" (Papineau, 2020, p. 18)

We think of pain in terms of the feelings it generates, not in terms of some role it plays. And so, connecting physical facts with the phenomenon of pain is something that doesn't come easy to us. We can't derive mind-brain identities and this is what creates a feeling of puzzlement about them. We conceive scientific properties and conscious properties differently.

I conclude from Papineau's approach that, if the explanatory gap is an epistemological gap, the derivability gap doesn't seem to be of epistemic kind, but of a deeper intuitive nature. I argue that such a gap that can never be closed. A scientific explanation of how and why consciousness arises from physical processes *might solve the explanatory gap for us. It might, however, have no impact on the derivability gap*, if this gap is as deeply routed in the way we conceive the world as Papineau states.

Is our intuitive dualism a proof against materialism as one might assume from Kripke's argument or rather a chance to solve the hard problem of consciousness as in Chalmers' hypothesis? For Levine (1983), the fact that this deep-seated intuition about our subjective experiences proves to be so resistant to philosophical dissolution shows that the enduring puzzle of the mind/body problem will stay with us for as long as its corresponding intuition does. Papineau invokes an inability to free ourselves from "an implicit commitment to dualism" (Papineau, 2011, p.8) as the cause for our feeling that something is left unexplained in this mind-brain problem. We have a strong belief that our pain cannot be just some C-fibers firing, that our conscious states cannot be reduced to brain states. For Papineau, there would be no explanatory gap if we only overcame our intuitive resistance and accepted mind-brain identity. Instead, the dualist nature of our thoughts on mind and brain makes it difficult to identify some phenomenal kind with a material kind. In his view, the presence of the dualist intuition does not pose an argument

against materialism itself. He suggests that the difficulty faced by materialism is *primarily psychological rather than theoretical*.

Papineau's main argument is that the intuitive implausibility of materialism does not pose a problem for the philosophical position itself. Materialism, according to Papineau, is a well-supported and coherent standpoint. The fact that many individuals find materialism difficult to believe is not sufficient to discredit its validity, since many truths are challenging to accept. He suggests that if the intuitive implausibility of materialism presents a problem, it is a problem for materialists to address, rather than a fundamental flaw in materialism. Materialists should recognize and examine the influence of dualistic intuitions on their thinking, adjusting their perspectives accordingly. He acknowledges that some may view the dualist intuition as evidence against materialism and expect materialists to explain or dismiss this intuition by demonstrating why it persists despite being false. However, *he refuses to concede that the dualist intuition inherently supports dualism* or undermines materialism. Instead, he asserts that the superiority of dualism's explanation for the dualist intuition over materialism's explanation should not be assumed without thorough investigation and evaluation. While acknowledging the significance of the dualist intuition and the desire to explain it, he argues against taking it as conclusive evidence against materialism or in favor of dualism and emphasizes the need to critically assess whether dualism can provide a more compelling explanation for the psychological phenomenon of the dualist intuition than materialism can.

I propose we look into an explanation of this intuitive dualism. We may notice the pattern we have of seeing the world through a binary approach when we think of how we understand living organisms. We are incapable of intuitively accepting a correspondence between some unnoticeable chemical processes and what is happening at macro level in a living organism. The only way in which we realized that life isn't something of a different nature, separated from chemistry, was to create very simple chemical structures and observe the passing from inorganic to organic: the same chemical components arranged in a certain structure that we call 'life', molecules organized in a way that forms 'living'

organisms. We learned that a complex animal is nothing but a reapplication of the same principle, that what we call 'life' is a form of organization of molecules.

A Final Word

The derivability gap is based on our incapacity to intuitively accept that something like our consciousness can derive from the processes in our brain. We might blame this on our intuitive dualism. If our intuitive dualism is standing in the way of our understanding of consciousness, it is instrumental to know the origin of this intuition. I offer a possible route. Rather than look into a socio-cultural source for our mind-body dualist intuition, I propose we look into how we perceive the world around us. Our binary view seems to go beyond the problem of consciousness.

Everything we learned from science about the continuum between inert matter and living organism, which teaches us of the gradual chemical changes that make the transition to life in an organism, doesn't seem to liberate us from the binary view of life and non-life. We cannot observe with a naked eye these micro-phenomena happening in the world, and so we use a binary classification. I propose that we look into whether the same simplifying principle that has helped us make sense of the world we couldn't micro-observe and understand is a principle hidden in our perception of what consciousness is. Not being directly exposed to this complexity might create our binary conception of life/non-life that we have taken further and associated with the difference between material and non-material. If we cannot accept that life isn't of a different separate nature then these processes, but rather 'derives' from chemistry, it is understandable that we also cannot *intuitively* accept that consciousness 'derives' from brain processes.

If this assumption is explored and validated by further research, I believe it can offer an answer to the problem of phenomenological reports and why we find consciousness so puzzling, if not dissolve the hard problem itself.

REFERENCES

Chalmers, D. (1995). Facing up to the Problem of Consciousness, *Journal of Consciousness Studies*, Volume 2, pp. 200-219.

Chalmers, D. (2018). The Meta-Problem of Consciousness, *Journal of Consciousness Studies*, Volume 25 (9–10), pp. 6-61.

Dumitru, M. (2019). *Lumi ale gândirii. Zece eseuri logico-metafizice*. Iaşi: Polirom.

Fine, K. (2020). Comments on Jessica Wilson's "Essence and Dependence". In: *Metaphysics, Meaning and Modality: Themes from Kit Fine* (pp. 471–475). Edited by Mircea Dumitru. Oxford, US: Oxford University Press. https://doi.org/10.1093/oso/9780199652624.003.0033

Kripke, S. A. (1980). *Naming and Necessity*, Harvard University Press.

Levine, J. (1983). Materialism and Qualia: The Explanatory Gap, *Pacific Philosophical Quarterly*, Volume 64, pp. 354-361.

Papineau, D. (2011). What Exactly Is the Explanatory Gap?, *Philosophia*, Volume 39, pp. 5–19.

Papineau, D. (2019). Response to Chalmers' 'The Meta-Problem of Consciousness', *Journal of Consciousness Studies*, Volume 26, Numbers 9–10, 2019, pp. 173–181(9).

Papineau, D. (2020). The Problem of Consciousness. In *The Oxford Handbook of the Philosophy of Consciousness*. Oxford, UK: pp. 14–36.

Quine, W. V., and J. S. Ullian. (2007). *Ţesătura Opiniilor. (*Romanian edition of *Web of Belief)*. Translated by Mircea Dumitru. Pitești, Romania: Paralela 45.

Revista Română de Filosofie Analitică,
Romanian Journal of Analytic Philosophy
Vol. XVI, Iulie-Decembrie 2022, Nr. 2, p. 83-118
ISSN (ediția electronică): 1843-9969,
ISSN (ediția tipărită): 1844-2218
DOI: 10.62229/rrfaxvi-2/5

THINNING THE JUNGLE OF "UNCONCEIVED ALTERNATIVES." STANFORD'S ANTIREALISM MEETS EXPECTED UNIFICATIONS AND AVOIDABLE INCONSISTENCIES

IOAN MUNTEAN[1]

Abstract: K. Stanford (2006, 2009) has offered an antirealist argument (the "problem of unconceived alternatives", *PUA*) based on the argument that scientists are not able to grasp alternatives to a current scientific theory *T*. According to *PUA*, the mere existence of some epistemically inaccessible alternatives (*T'*, *T''*, …) weakens our trust in *T* and shakes the foundations of mainstream scientific realism. The realist may entertain the inkling that inter-theoretical relations (both existing and expected or 'hoped-for') play a role in accepting or rejecting *PUA*. The most celebrated intertheoretical relations, such as consilience, reduction, realization, emergence, equivalence, or approximation—whether prospective, expected, or realized—bear relevance to the conceivability of their alternatives. This paper presents an 'eliminative inference' based on an 'unification posit' that weakens the *PUA*. We employ first a minimal model of inter-theoretical unification couched in terms of the 'term identification' of the theoretical terms of two initially different theories, T_1 and T_2. Then we rethink unification as an 'ideological identification' where predicates in different theories are identified. Finally, we can envision a more sophisticated unification as entailment relations among T_1 and T_2 and their

[1] University of Texas, Rio Grande Valley and University of Illinois, Urbana-Champaign. Email: ioan.muntean@utrgv.edu. This material has benefited from the feedback of the audience and reviewers from the Philosophy of Science Association, the Society for Exact Philosophy, the Indiana Philosophical Association, and reviewers of the *RRFA*. Special thanks to Otavio Bueno, Ken Long, Andrei Mărăşoiu, and Connor Wilson for suggestions and support.

empirical grounds. In all these cases, we propose scenarios of inconceivability based on a minimal consistency requirement run against the "syntactic view" of scientific theories. The upshot of this mechanism is that *some* alternatives to T_1, which remain unconceived within the conceptual and ideological space of T_1, can be eliminated because they are inconsistent with empirical constraints on T_2. The overall space of 'serious' alternatives to both theories is 'thinned.' Consistency is a requirement that conditions inter-theoretical relations, mainly when the overlapping evidence supports theories. This argument illustrates in what sense *PUA* is lessened when scientists or scientific communities operate based on theoretical posits.

Keywords: *unconceived alternatives, K. Stanford, antirealism, conceptual space, B. Van Fraassen, P. Gärdenfors*

1. Introduction

Over the last five decades, several successful challenges to scientific realism have been formulated, including "the pessimistic meta-induction," the "underdetermination of theories by evidence," and the "problem of unconceived alternatives" (PUA). This paper is focused on the latter, which comes in several flavors. L. Sklar states that for any scientific theory T there are always incompatible alternatives (in what follow, alternatives to T are designated as a space of theories between which no rational choice can be made based on a priori "plausibilities, strength, parsimony, inductive confirmation, and so forth, relative to present empirical evidence" (L. Sklar 1981). Sklar's underdetermination of T by its alternatives is recurrent and transient. All alternatives are transient because new data may render some less attractive or make others more preferred, depending on the evidence we gather in the future (L. Sklar 1981; Lawrence Sklar 2000).

K. Stanford has advocated a different version of PUA and presented it as a comprehensive argument, supported by some historical evidence.[2]

[2] *PUA* is exposed in the 2006 book and promoted as "the new induction" (*NI)* in contrast with the mainstream pessimistic metainduction (PMI) (Stanford 2006). Other references to the *PUA* and NI are (Bhakthavatsalam and Kidd 2019; Chakravartty 2008; Devitt 2011;

PUA is premised on the idea that the most significant challenge to scientific realism arises from our inability to consider the full range of serious alternatives to a given hypothesis we seek to confirm and ultimately accept. Stanford's argument is an induction from the history of science (including all domains and all historical periods) to present and future science:

> We have every reason to believe that there are theoretical alternatives remaining unconceived by us whose grasp will be regarded by future scientific communities as absolutely fundamental and/ or a necessary precondition for conceiving of or even understanding the further accounts of nature that they themselves embrace. (Stanford 2006, 131).

In his book-length argument, Stanford puts forward a very general claim: scientists, in *every* scientific field and *at any time* in history, have found themselves in an "epistemic position in which [they] could conceive of only one or a few theories that were well confirmed by the available evidence, while subsequent inquiry would routinely (if not invariably) reveal further, radically distinct alternatives as well confirmed by the previously available evidence as those [they] were inclined to accept on the strength of that evidence" (Stanford 2006, 19). The book discusses at least three cases from the 19th-century history of biological inheritance: Ch. Darwin's "pangenesis theory", Fr. Galton's "stirp theory," and Weismann's "germ-plasm theory." The view ignored by all these scientists was gene regulation, so each of these scientists failed to envisage a relevant alternative to their theory about inheritance, the alternative that would have been accepted later by the scientific community (Stanford 2006, 132–33).

First, there is an essential caveat under which Stanford's argument is operational: because he focuses on unconceivability, *PUA* refers to scientists, not scientific theories. This sharply contrasts with the standard underdetermination problem, which emphasizes the falsehood of theories. Relevantly, Sklar's 'transient underdetermination' refers to both

---

Forber 2007; Godfrey-Smith 2007; Kukla 2010, 2010; Rowbottom 2019; Saatsi, Psillos, and Winther 2009; Stanford 2021, 2009, 2015b; Zamora Bonilla 2019).

the truth of theories and the scientists inability to conceive of alternatives to existing theories. The present paper offers an analysis of the inconceivability of radically different alternatives to a given theory T, which is arguably distinct from the mere existence of logical alternatives, which may not be different relevantly from T. Here, $\mathcal{T}_T = \{T', T'', T''' \ldots T^{(n)}\}$ is the countable set of alternatives of T, all of them being supported by the same evidence E as T but most of them being incompatible with T. This raises the immediate issue of the identity of theory T when compared to its unconceived alternatives. How do they differ from other members of $\mathcal{T}_T$ and from T? The fundamental difference here is epistemic, as given by all scientists at a moment t, none of the scientists can conceive any of the $T', T'', T''' \ldots$

From the history of science discussed in the book, Stanford infers that future scientists would see the space of our "theoretical grasp" as limited as we see old theories nowadays (*e.g.,* Weismann's) through the lenses of molecular cell biology:

> We have every reason to believe that there are theoretical alternatives remaining unconceived by us whose grasp will be regarded by future scientific communities as absolutely fundamental and/or a necessary precondition for conceiving of or even understanding the further accounts of nature that they themselves embrace (Stanford 2006, 131).

The current molecular genetics has been arguably supported empirically by the experimental data available to Weismann, whereas his theory has been rejected by the data made available subsequently (in the 1920s).

1.1. Conceivable problems with the Problem of Unconceived Alternatives (*PUA*)

There are probably several ways to reject Stanford's *PUA*. First, a realist can "trivialize" this "problem" by insisting it is not entirely different from the underdetermination of theories by evidence or the pessimistic metainduction. Hence, it is fundamentally vulnerable to the same

arguments marshaled usually against antirealist arguments. If *PUA* is nothing more than *UTE* and *PMI* in disguise, then it can be addressed by the realist in the same way. *PUA* is important in many ways, but it generates new consequences for neither the scientific realism nor the antirealism positions.[3]

Second, one can reject the range to which *PUA* applies. Even if *PUA* is plausible in the case studies discussed in the book, what makes us think that the schema can be generalized beyond that historical context? As some reviewers have noted, Stanford's historical initial base is relatively small and restricted to a relatively short period in the history of biology (Ruhmkorff 2019; Votsis 2007). *PUA* is then a weak inductive argument, either because the sample used by Stanford is small, or because the sample is atypical.

However, one way to reconsider Stanford's *NI* is by applying it to other disciplines or other theories outside the scope of evolutionary biology. P.D. Magnus questioned the validity of Stanford's schema in a different context by addressing this question: were classical mechanics and the special theory of relativity equally confirmed in, let us say, the year 1780 (Magnus 2006). One can admit that carefully selected data from astronomical and terrestrial observation made before the 1780s would corroborate the special theory of relativity as an unconceived alternative to Newtonian mechanics.

Third, one can weaken Stanford's *NI* by showing that his predicament is fundamentally sound for a large enough class of examples from the *past* history of science (included in the book), albeit current scientific theories and the way scientists think about science differ in some fundamental respects from historical cases. We have become better at approaching truth, using scientific standards, and acknowledging the schema of falsificationism in our current theories than scientists in the past have done. As some realists have noticed, our current and future theories are becoming more sophisticated than the historical cases at hand.

---

[3] This argument is expressed in (Chakravartty 2008; Ruhmkorff 2019) and by J. Saatsi in (Saatsi, Psillos, and Winther 2009; Saatsi 2019).

1.2. The maturity question: science or scientists?

In a more concrete sense, one can talk about the maturity of scientific communities and the maturity of their respective theories. According to this argument, most cases discussed by Stanford are not 'mature' theories, in the sense that some of our present theories are mature. Similarly, others would consider that scientific communities have become increasingly efficient in discovering alternatives to a given theory. P. Godfrey-Smith pointed out that scientific communities in our days are bigger, better connected, better organized, and can better explore the alternatives to mainstream theories (Godfrey-Smith 2007). The current communities are less vulnerable to the problem of "unconceived alternatives" than past communities of scientists because of some fundamental differences in their "epistemic status." Others would claim that *PUA* is less likely to be applied to the present and future of science, because something in the methodology, metaphysical commitments, and the general organization of science has changed radically since, let us say, the biology of the 19th century. As M. Devitt stated:

> […] we have very good reason to believe that we have been getting better and better at learning about the unobservable world; good reason to think that, aided by technological developments, there has been, over recent centuries, a steady improvement in the methodology of science. That's why our present theories are more successful" (Devitt 2011, 292).

Yet another key ingredient is the claim that the changes in the general advancement of science will not be so "dramatic," to put it that way, in the future. Stanford coined this attitude as "scientific catastrophism" (Stanford 2015a). The catastrophist postulates that the future of science will be quite different from its past or present and that the historically significant changes in science are mostly a thing of the past. Catastrophists weaken the power of the *PUA* from the past to the future by postulating that the history of science is not uniform. To the other camp, the "uniformitarians," significant changes will occur in the future at roughly the same rate as in the past: "In the course of further inquiry, those theories will ultimately be overturned, supplanted, or

transformed in the manner of their historical predecessors" (Stanford 2015a, 877).

The realists usually retort by pointing out that the history of science is not uniform and that irreversible changes affect how we conduct the scientific endeavor from now on. When realists such as Godfrey-Smith and others emphasize that communities of scientists are more important (as opposed to individual scientists), they provide such a mechanism against *PUA* (Dellsén 2019; Godfrey-Smith 2007). To attack Stanford's *NI*, it is sufficient to show that when a community of scientists is larger, communicates more effectively, adopts a more effective methodology, separates relevant information from noise, etc., the scientific endeavor is less vulnerable to the problem of unconceived alternatives. Stanford shows that, on the contrary, the current situation is fundamentally different: theoretical orthodoxy, deep-rooted bias of the present institutionalized science, fosters scientific *conservatism* which, as his argument goes, was not present during the Scientific Revolution and the one or two centuries to follow:

> Today's scientific communities are almost certainly more effective vehicles for testing, evaluating, and applying theoretical conceptions of various parts of the natural world than were their historical predecessors, but I have argued that we have compelling reasons to believe that they are less effective than those same predecessors in conceiving, exploring, or developing fundamentally novel theoretical conceptions of nature in the first place (Stanford 2015b, 3931).

1.3. Assumptions of the current argument

This paper proposes a new framework for Stanford's debate with P. Godfrey-Smith, A. Chakravartty, J. Saatsi, and F. Dellsén. In the present framework, we set aside the sociological and historical differences between the past and the future of science *per se*. Separately from the value-laden historical context, one can see that science has been periodically controlled by certain 'posits': theoretical virtues or theories,

ideals, background information, methodological maxims, metaphysical assumptions, and worldviews, among others. The posits are conceptualized as "external posits" by P. Vickers, "presuppositions" by Ph. Kitcher, or "idle elements" by St. Psillos (Philip Kitcher 1995; Psillos 1999; Vickers 2013). Sometimes these posits are adopted explicitly, sometimes presupposed tacitly. They do not show up explicitly in scientific inferences, so they can be called "indirect." Nevertheless, at first glance, some posits can serve as eliminative entailments that decrease the number of unconceived alternatives. At various points in the history of science, these posits can be more or less effective in diminishing the strength of Stanford's inductive argument.

We take an 'unification posit' as a sought-for epistemic virtue of a community of scientists and, by extension, a virtue of a scientific discipline (hopefully constituted by more and more unificatory theories). The main aim of this paper is to investigate how unificatory posit imposed on the theories of one discipline can reduce the pool of alternative theories, even if these alternatives remain unconceived.

Stanford argues that today's scientists are not more creative or efficient in their ability to exhaust the logical space of alternatives to a given theory. Regarding background information, Stanford believes that auxiliary hypotheses, although they may improve over time, are typically overlooked despite being equally well-supported by the available evidence. In other words, in such cases, the totality of evidence available at the time of an earlier theory's acceptance characteristically offers equally compelling support for the combination of a later accepted alternative to that theory, together with the requisite alternative auxiliary hypotheses that would themselves later come to be accepted (Stanford 2006, 20). This paper argues that theories of an scientific discipline *D* at an earlier moment could be subsumed under an 'expected unification' posit, and this in itself would reduce space of reciprocally *consistent* unconceived alternatives in *D*.

The present argument does not attempt to show that contemporary scientific communities are better than past ones or worse than future ones: they profess different posits. The argument here is premised on the idea that communities of scientists may operate at different epochs, on a

different set of theoretical posits, and seek different epistemic goals, which can be couched in terms of posits. Nevertheless, some of these posits endure over time or are more stable than methods, experimental setups, evidence, or theory choice standards. Each scientific discipline can undergo different stages, but one or more of its posits remain relatively unchanged. It is assumed here that posits such as parsimony, unification, explanatory power, predictive power, empirical adequacy, and coherence play a particular stable and enduring role. If the present argument is sound, then theoretical posits offer a general argument against the *PUA*, which weakens it, independently of the uniformity of the history of science, of different sociological factors, or the empirical success or failures of *D*. These theoretical posits are, it is assumed here, general enough and change relatively infrequently in the history of science. S. Schindler suggests that we should believe some of our theories because they (some only!) possess 'virtues' (non-ad hocness, consistency, unification, parsimony, or fertility) that extend beyond the evidence (Schindler 2018). This is the realist posit about scientific theories, and according to some statistical analysis, posing theoretical virtues is popular among scientists (Mizrahi 2022; Schindler 2022).

Middle positions between realism and antirealism have friendly attitudes towards such posits. In M. Massimi's perspectivism, standards of scientific conduct are relatively stable when compared to changes of perspectives (Massimi 2021). The present argument can be read as a conditional statement: "if this and that standard *S* is present and endures at some epoch in the history of *D*, that period is less vulnerable to the *PUA*, compared with other epochs in which *S* is abandoned or replaced frequently." The current paper identifies at least one standard *S* that may weaken the power of the *PUA*. This argument can be categorized as an argument for convergent realism, albeit weaker than arguments based on the need for diversity in the sampled history. It is not an argument based on the uniformity of the history of science, but can be used in conjunction with it.

2. Space of Unconceived Alternatives to Theories (UAT) and theoretical posits

What does Stanford mean by inconceivability and how do we identify elements in $\mathcal{T}_T$? Stanford and his critics express only informally the requirements on $\mathcal{T}_T$.[4] Stanford runs on a skeptical ticket against scientific knowledge by emphasizing that the *PUA* is an undisputable and unavoidable aspect of scientific life, at any moment in the history of science, contemporary scientists are no exception. As Stanford's *PUA* is primarily branded as an argument against scientific realism, its cognitive and epistemic dimensions are often overlooked. On the contrary, by situating his argument on the realist-antirealist map, we take Stanford's argument to tackle the cognitive inability of scientists to conceive alternatives directly. The present paper frames cognitive inability in terms of syntactic constraints on how scientists use the vocabulary of existing theories. Based on Stanford's own characterization, "unconceived alternatives" to our current theories are the result of our cognitive inability to conceive a different theory equally well confirmed by the existing and known empirical evidence. The failure to conceive alternatives to a theory is couched in the following sections in a first-order logic formalism. Imagine we have a finite set of objects and a finite set of conceivable predicates. One way to delineate the inability to conceive alternatives to a given theory that pairs some objects with some predicates is to keep the vocabulary of the language fixed and set a limit on our ability to pair the same objects with existing predicates. The other approach is to depopulate the signatures of relevant objects or predicates that are currently not used in our science but may become helpful in the future. Yet another, even more radical way is to introduce entirely new constants and predicates in $\mathcal{T}_T$ nonexistent in T. In the following, we aim to clarify these scenarios.

One suggestion, repeated in different contexts, is that inconceivability is linked to scientists' inability to imaginatively exhaust a *space* of plausible, scientifically serious, and reasonable candidate

[4] See the *Synthese* volume on *PUA* and its introduction in (Bhakthavatsalam and Kidd 2019).

theories for a given set of phenomena before "proceeding to eliminate all but a single contender" (Stanford 2006, 29, 31, 32). Let us assume, informally, that alternatives to *T*, referred here as the space of 'Unconceived Alternative Theories' (*UAT*), form a space that scientists can, in principle, explore. One task of the present approach is to develop a more precise formulation of this *UAT* space and ways to reduce it. It is often repeated by Stanford and Sklar that the space of important, and serious, alternatives to a theory is "indefinitely large:" at a given moment, a scientist or even a large community of scientists can embrace only a subspace of such a space. Genuine conceptual improvements in the present, compared to past science, mean that:

> we enjoy the luxury of conceiving of and considering an ever-larger space of serious theoretical alternatives. Of course, even if the space of unconceived alternatives contained only a finite number of well-defined possibilities, we would seem to have little reason to believe that we are presently at the end of an exhaustive search of it and have finally reached the point at which serious unconceived possibilities no longer pose any real danger to our theoretical science in a given domain (Stanford 2006, 133).

Stanford warns us that the space of alternatives has a "vague and indefinite character, with members that are difficult if not impossible to individuate sharply or unequivocally: an indefinite number of alternative possibilities are neglected" (Stanford 2006, 133). But the argument for such a claim, according to Stanford, is mainly historical:

> While there are certainly cases of eliminative inferences in which we can justify restricting our attention to some small part of the space of possibilities […] our historical investigation will suggest that in the case of fundamental theoretical science it is often a consequence of our failure to conceive of the serious alternative possibilities that do in fact exist that we embrace the substantive assumptions needed to restrict the space of theoretical alternatives under consideration to a comparatively small and/or well-behaved set (Stanford 2006, 41).

Stanford presages us to accept that the space of alternatives has a "vague and indefinite character, with members that are difficult if not

impossible to individuate sharply or unequivocally: an indefinite number of alternative possibilities are neglected" (Stanford 2006, 133).

2.1. Downsizing the *UAT* space

It is part and parcel of the present argument to provide a formal framework for a version of the eliminative argument that can reduce the impact of the *UAT* spaces on the practice of science by scientists and on scientific realism. We restrict the present analysis parsimoniously to a first-order logic, a rather orthodox form of the "received view" on scientific theories and the theoretical language (TL_T) of the theory, without insisting too much on the observational language (OL_T) or their correspondence functions (CF_T).[5]

When and how do theories have unconceivable alternatives? There is a trivial answer to this question: always! At any time *t*, in any scientific discipline *D* and any theory *T* within it, there are alternatives (conceived or not by the scientists practicing *D*), supported at *t* by the same known evidence *E* that supports *T*, just because any *T* is underdetermined by *E*. This trite answer is well heeded by Stanford: the underdetermination of any *T* by evidence *E* is not the same as *PUA*. Although there is an infinity of trivial, conceivable alternatives to *T* equally well supported by *E*, not all of them are 'serious,' distinct enough from each other (and from *T*), and truly unconceivable. Nevertheless, what are the *serious* and *distinct* alternatives to *T* still unconceivable within *D* (i.e., not by the individual scientist, but all the scientists practicing *D*)?

Relating the informal approach of the *NI* to the syntactic view and scientific unification requires some conceptual clarifications and definitions that both proponents and critics of *PUA* could accept. For the present purposes, we choose a minimal formalism suited for bringing together Stanford's *PUA* and the unification posit in the received view framework.

[5] We follow here the notations and standard syntactic conventions from (Carnap 1995 (1966); Hempel 1966).

To challenge *PUA*, any realist argument must articulate clearly the "truly inconceivability" of *T*'s alternatives. Domain *D* and a definite empirical evidence *E* to *D*, one can imagine a set $\Theta_E$comprised of all theories in *D* empirically supported by *E*. In simple probabilistic terms:

$$\forall T(T \in \Theta_E \rightarrow (pr(T|E) > pr(T))$$

Within Θ_E there is a 'field' Y_E defined as a set of theories accepted, known, and grasped by scientists practicing that discipline, together with the accepted body of evidence *E*. The complementary set $\overline{Y}_E$ is simply the set of unknown and inconceivable theories supported by E.[6] We assume now these rather inconspicuous claims:

(1) Any theory *T* in Y_E is known and accepted by the scientists practicing in *D* at *t*. In a formal notation, $\forall T(T \in Y_E \rightarrow (K_t(T, D) \wedge A_t(T, D))$, where $A_t(p,C)$ is a two-place predicate that formalizes the acceptance of a proposition *p* by the community *C* at *t*, and $K_t(p,C)$ formalizes that *p* is known by the community *C* at *t*.[7]

Acceptance of a scientific theory must meet some conditions, even if, as Quine points out, "acceptability depends on a weighing of the total evidence" (Quine and Ullian, 2007). Minimally, they are at least:

(2) Acceptance in *D* that *E* supports *T* at *t* in a probabilistic term: it is true that the probability of 'T given E', $pr(TE)$, is higher than the probability of *T*: $\forall T(T \in Y_E \rightarrow A_t('pr(T|E) > pr(T)',D)$;

(3) Acceptance in *D* that *T* is 'true' at t: $A_t('T',D)$. This may mean in a strong form that if *s* is a scientist, then she believes '*T* is true' or, in an inferential way, that the negation of T entails a contradiction given rules of inference in D: $\forall s(s \in D \rightarrow A_t('T', s)$. In terms of entailment: $\forall s(s \in D \rightarrow A_t(\sim T \vdash_D \perp, s)$.

---

[6] We assume that $\overline{Y}_E \cup Y_E = \Theta_E$ is a partition of Θ_E.

[7] We do not assume here any particular relation between predicates *A* and *K* except that accepted theories are "known" and that "to be known" does not entail "to be accepted."

2.2. A functional definition of unconceivability

What is exactly the field of theories Y_E?At a more refined level of analysis, other requirements for accepting T may include predictive and/or explanatory powers, conformity with current scientific standards (methodologically, epistemically, etc.), fecundity, usefulness, etc. However, for the present purpose is adamant to see that the *D* community can, in principle, expect T to enter into some *future* inter-theoretical relationship with another accepted theories from Y_E. Then T is accepted as part of a larger theoretical field composed of different theories, models, etc. in Y_E and within this theoretical field some theories may become reduced, eliminated, or… unified. This is a holistic acceptance that constitutes a theoretical posit relevant to *NI* and *PUA*. In this paper, we focus on the unifying inter-theoretical posit that can potentially reduce the power range of *PUA*.

DEFINITION 1. At any moment t, for any theory T in Y_E there is a non-null unconceivability function $U_t: Y_E \rightarrow \overline{Y}_E$ relating one theory in $Y_E$to a countable set of unconceived alternative theories $\mathcal{T}_T \in \overline{Y}_E$,such that $\mathcal{T}_T = U_t(T)$:

- I) At t, E supports all alternative theories in the inconceivability co-domain of T:

$$\forall T'\big(T' \in \mathcal{T}_T \rightarrow (pr(T'|E) > pr(T')\big)$$

- II) At t, none of the theories in $\mathcal{T}_T$ is known (grasped) or accepted in *D*:

$$\forall s \forall T'\big((s \in D \wedge T' \in \mathcal{T}_T) \rightarrow (\sim K_t(T', s)\big)$$

- III) Unbeknownst to all scientists in *D*, there is at least one theory in $\mathcal{T}_T$ that E supports better than it supports T:

$$\exists T'\left(T' \in \mathcal{T}_T \wedge \big(pr(E|T') > pr(E|T)\big) \wedge \sim K_t(T', D)\right)$$

- IV) The degree of the inconceivability of T is not related to one alternative T', but to the size of the minimal inconsistent set of T and $\mathcal{T}_T$: MI($T \wedge \mathcal{T}_T$).

We do not envisage inconceivability as a characteristic of T, but as a function $U_t: Y_E \rightarrow \overline{Y}_E$ which relates at t some accepted T in Y_E (by all its practitioners $\forall s \in D$) and empirical evidence E that supports T at t (a moment or an interval of time) to a subset $\mathcal{T}_T$ from the set of the unconceived theories supported by E: $\mathcal{T}_T \in \overline{Y}_E$.

The minimal inconsistent set of A is defined as a subset of A that is inconsistent, but any sub-subset of the latter is consistent:

$$\mathrm{MI}(A) = \{A' \in A; A' \vdash \perp; \forall A'' \in A'; A'' \nvdash \perp\}$$

(Benferhat, Dubois, and Prade 1997).

Removing one element from MI(A) makes it consistent but MI(A) is inconsistent. A consistent set A has $\mathrm{MI}(A) = \varnothing$, while for an inconsistent set $A', 2 \leq \mathrm{card}(\mathrm{MI}(A')) \leq \mathrm{card}(A') - 2$. We expect U_t to change in time, as E changes, as well as how E supports T. The identity of theories in $\mathcal{T}_T$ is vital for *PUA*, as at $t_1 > t_0$, one or more of them will replace T as newly accepted theories in D, once new evidence E_1 is acquired at t_1. We correlate the "seriousness" of the alternatives to T with the MI($\mathcal{T}_T \wedge T$), but not with the MI($\mathcal{T}_T$). In fact, we impose the condition:

$$T \in \mathrm{MI}(\mathcal{T}_T \wedge T)$$

This means that T has to be incompatible with members of $\mathcal{T}_T$, whereas the inconsistency of $\mathcal{T}_T$ is not relevant here.

An unconceived alternative to T in D is a member of the set $\mathcal{T}_T = U_t(T)$, all empirically supported by E at t. Members of Y_E, the set of known and accepted theories in D is supposed to be as consistent as possible, whereas $\overline{Y}_E$ is not, and neither is $\Theta_E = Y_E \cup \overline{Y}_E$. Crucially, Stanford thinks that $\mathcal{T}_T$ and T have a non-trivial degree of logical inconsistency, albeit they are at t supported by the same empirical evidence. However, the

definition above is not sufficient for the present argument. The framework in which the present argument is couched includes these extra components:

- Delineation of the 'inconceivability of alternatives' $\mathcal{I}_T$, given T;
- Simplified versions for the (future) unificatory posit of two theories T_1 and T_2 using the "syntactic view".
- A requirement for consistency and the delineation of this requirement from the lax presence of inconsistencies among sets Y_E and $\overline{Y}_E$.

The result is an eliminative inference that weakens the *PUA* by discarding classes of alternatives to T. The eliminative inference proposed here removes parts of UAT from consideration and restricts the attention to its significant subspaces. The assumption used here is a 'unificatory' posit, which qualifies in Stanford as "substantive assumption" (Stanford 2006, 40). The unification posit relates two theories from D, rather than a single theory (as does, for example, the parsimony posit), and for the present goal, it would weaken in some cases *PUA*. Unification is more akin to a normative and prescriptive posit and not as an *actual* and *effective* inter-theoretical relation: statistically, we have robust reasons to believe that contemporary scientists adopt it as an ideal (Mizrahi 2022; Schindler 2022). Instead of focusing on one theory and its alternatives and successors, the present approach relates two different theories simultaneously, mostly known or potentially known to the scientific community. The unification is the hypothesis that in the future they could be unified by a simple process of identification of theoretical terms, and it probably should be called the "austere unification" which is in this approach expected, sought-for, or just hoped-for.

Two caveats are in order here. First, we use a simplified version of scientific unification, decoupled from explanation, prediction, or understanding.[8] We proceed this way to better relate to Stanford's claim

[8] In this sense, the present approach does not follow the more orthodox approaches to unification of Friedman, Kitcher, Morrison, or Schupbach (Friedman 2001, 1974; Philip Kitcher 1981; P. Kitcher 1999; Morrison 2000; Schupbach 2005). Unification as theoretical

that his new induction is about scientists, their goals, ideals, and practices of science, not about scientific theories *per se*. It appears that neither explanation (including inference to the best explanation), nor prediction plays a central role in *PUA*. What is central in *PUA* is being realist about the theoretical terms of our current accepted theories.

Second, we use a basic version of the "syntactic view" of scientific theories. There are two main reasons for employing the syntactic view: first, it seems that Stanford's own argument is couched more in terms of the syntactic view (*aka* "Received View"), and less in terms of the more popular "Semantic View" of scientific theories. *PUA* is about how scientists conceive and formulate their theories as collections of statements about the world, and not as models. Second, even if this present approach is provisional and probably simplistic, it is unclear whether a semantic view approach would fundamentally change the eliminative inference argument against *PUA*.[9] The present choice reflects a general trend to view the syntactic and semantic views as alternative descriptions of scientific theories, rather than opposing ones. Some authors talk about the peaceful coexistence of these two views (Lutz 2017, 2014), while others dismiss the semantic view altogether and favor the syntactic view (Halvorson 2013). Without further ado, a syntactic approach based on first-order logic is assumed to be sufficient for the present purpose. Our aim no is to formalize scientific theories, but to provide an eliminative inference that weakens *PUA* in specific cases. An alternative approach, based on the semantic view of science (models or partial structures rather than theories), or any other alternatives to the syntactic view, may or may not illuminate interesting aspects, but is not followed here.

---

virtue is discussed more recently in (Kao 2019; Patrick 2018; Roche and Sober 2017; Schindler 2022). Pluralism about scientific theories and the 'dis-unity posit' in science is disucssed in (Cartwright 1999; Dupré 1993; Hartmann 2001).

[9] This can be an interesting venue for research, not addressed here. We use the idea of conceptual spaces, similar in spirit to the syntactic view, as discussed in (Gärdenfors 2000). An even more attractive option, belonging to the semantic view, is to use the formalism of "partial structures" and quasi-truth, advanced by French and da Costa (Costa and French 2003). See a comparison in (Bueno 2015).

Practitioners of science in D are employing a first-order language $\mathcal{L}$, formalized as a vocabulary $\mathcal{V}$, containing logical terms such as some quantifiers $(\exists, \forall, \ldots)$, connectives $\rightarrow \leftrightarrow \vee \wedge, \ldots$, the identity symbol (= used between constants or variables), and a signature Σ which includes the constants c_1, c_2, c_3 ... (representing theoretical or unobservable terms), predicates $P_1, P_2 \ldots . P_n$ with arbitrary arity, and functions $F_1 \ldots F_p$, each with an arbitrary number of arguments $F_i(x_1, x_2 \ldots .)$.[10] To all these we can add rules of entailment for $\mathcal{L}$ accepted by scientists in D.[11] We assume that the signature Σ of $\mathcal{V}$ has a *finite* number of theoretical terms, predicates, and functions. This model's UAT space comprises all possible combinations between logical terms, constants, predicates, and functions as defined by T and all theories in $\mathcal{T}_{\mathcal{T}}$. In this sense, the space of possible combinations is countable, given the infinite number of combinations among logical terms (even if the number of constants, predicates, and functions is finite). We focus here exclusively on the countable case, in which the signatures of our vocabulary are countable. The *PUA* idea is that even if the number of possible combinations is finite, scientists cannot grasp at t the combinations between, let us say, the theoretical objects of their theories and the possible (but conceivable) set of predicates.

In this simplified view, a scientific theory consists of a set of theoretical terms and predicates that can be attributed to these terms, along with all the logical consequences that can be inferred. For example, a theory T can quantify over two theoretical terms c_1 and c_2 with several predicates of any n-arity: $\{P_1, P_2, \ldots P_n\}$ will have a simple signature $\sigma_T = \langle \{c_1, c_2\}, \{P_1, P_2 \ldots P_n\} \rangle$.[12]

---

[10] The signature is the part of the vocabulary that contains all the constants c, predicates P, and functions F. In our approach the scientific domain D uses the same vocabulary, but theories may have different signatures $\sigma_{T_1}, \sigma_{T_2}, \ldots$.

[11] We take here a syntactic view about entailment in D and assume simple forms of entailment in $\mathcal{L}$ from a theory T such as: $T \nvdash_D \perp$ for 'T is true' and $T \vdash_D \perp$ for 'T is not true.' This implies that the language of D comes with rules of entailment.

[12] Here, we ignore the observational terms of T as they do not play a clear role in *PUA*. We also ignore the correspondence functions that relate theoretical terms to observational terms, unless otherwise stated. We characterize the relation of a theory with evidence in Bayesian probabilistic terms.

3. Inconceivability of theories: three scenarios

The main question is: how do we define the inconceivability of T over the space spanned by $\mathcal{T}_T$? The space-inspired accounts (B. van Fraassen, P. Gärdenfors, F. Zenker, *i.a.*) are helpful to our approach: the space of $\mathcal{T}_T$ is larger than what the scientific community can conceive.[13] This paper presents several scenarios for the inconceivability of theories, ranked from strong to weak, and applies the eliminative inference to a weak form of inconceivability.

3.1. **Scenario 1:** ***terminological & ideological*** **(full) inconceivability**

The strongest model for what Sklar and Stanford might imply about our (recurrent) inability to exhaust the space of alternatives to a theory is based on our failure to exhaust the space of theoretical terms c_i, the space of the predicates P_j, and that of functions F_k. Given E at t, the scientists fail to connect the correct theoretical terms with the appropriate predicates or functions. In this scenario, the scientist(s) may have limited access to the space of theoretical terms. They cannot imagine enough relevant, serious, and meaningful alternatives to T in $\mathcal{T}_T$ because the alternatives to a given theory may use different theoretical entities.

In a very simplified version, if a theory T with a signature $\sigma_T = \langle c_1, P_1 \rangle$ is composed of a constant c_1 and a predicate P_1 that applies to c_1, then one of its alternative $T' \in \mathcal{T}_T$ has a different constant c_2, a different predicate P_2 and another signature $\sigma_{T'} = \langle c_2, P_2 \rangle$. Theory T may claim that $P_1(c_1)$ is true, whereas T' claims that $P_2(c_2)$ is true. T and T' are consistent in this simple case as they have disjunctive signatures and any combination of them can be conducive to truth. But the conceivability of T' is a difficult epistemic process. Scientists need to "jump" in the conceptual space from the point c_1 with property P_1 to a different point c_2

[13] Van Fraassen's and Gärdenfors' 'state-space' and 'conceptual space' are different and may serve different purposes, but we use both approaches here. See (Bueno 2015; Gardenfors 1990; Gärdenfors 2014; Van Fraassen 2008).

with a new property P_2. The inconceivability of T' relative to T means that given the posits of D at t and various constraints on the epistemic reach of the scientists practicing D, this 'jump' is unlikely at t. Together with P. Gärdenfors, we assume that points in conceptual spaces have properties, and regions of points with the same property form a subspace. The distance between areas of the extensions of P_1 and that of P_2 is a central concept in this approach, as it correlates with the probability of T' being conceivable (although not accepted yet) coming from the space of T. We also need to postulate that T and T' are empirically equivalent in that they are supported by the same empirical evidence E.[14] Let us call this the "terminological&ideological inconceivability".[15] We can assume that this scenario is a case of a serious UAT for Stanford.

3.2. Scenario 2: *terminological* inconceivability

A weaker inconceivability scenario occurs when scientists currently use a set of predicates and functions, but lack the correct theoretical term(s) to be predicated of. Scientists can still conceive the relevant predicates P_i or functions F_i, although they are predicated about the improper theoretical terms. This scenario can be called a "term inconceivability." For example, if T is the wrong theory at present, with signature $\sigma_T = \langle c_1, P_1 \rangle$ and it claims $P_1(c_1)$ (c_1 is in the extension of P_1), then the correct, alternative theory T' with $\sigma_{T'} = \langle \{c_1, c_2\}, P_1 \rangle$ will state correctly that a different term c_2 is in the extension of the predicate P_1: $P_1(c_2)$ and state that $\sim P_1(c_1)$. T and $T'$are inconsistent: $T \wedge T' \vdash_D \perp$, although T cannot assign truth values to $P_1(c_2)$: $T \wedge P_1(c_2) \nvdash_D \perp$ and $\sim (T \wedge P_1(c_2) \nvdash_D \perp$.

[14] We do not state here what it means to be empirically supported by evidence. Still, in general, this can be couched in terms of correspondence functions between theoretical and observational terms. It is possible that T and T' need to share a set of observational terms and have their own correspondence functions.

[15] We use ideology here in a restrictive sense inspired by Quine: the ideology of a theory is the list of n-place predicates used by that theory (Quine 1951).

3.3. Scenario 3: *ideological* inconceivability

There is another way in which scientists cannot conceive an alternative to T by operating with a different set of predicates and functions on the same set of theoretical terms. Here, scientists do not have the whole ideology available when considering all possible alternatives to a theory, although the same set of theoretical terms is used. This is a form of "ideological inconceivability." For example, if $\sigma_T = \langle c_1, P_1 \rangle$, then an alternative theory T′ in $\mathcal{T}_T$ uses the same theoretical term c_1 with a different predicate P_2: $T' = \langle c_1, P_2 \rangle$. In this case, the truth value of $P_2(c_1)$ is true which, according to *PUA* is equally supported by E at t and will be accepted and known at at t_1>t.

Think of a Kuhnian example: the properties of mass in Newtons and Einsteins theories of gravity. The same theoretical term has different properties, but these meanings can be compared and contrasted with one another. Although the geometric and topological properties of the Newtonian and relativistic spacetimes overlap at the lower velocity limit, some of their properties are fundamentally different, the most obvious being the mass dependence on the velocity. As before, the theories T and T' can be empirically equivalent, for a given set of data (in this case, for low velocities compared to the speed of light).

3.4. Scenario 4: *predication* inconceivability

Last, the weaker form of inconceivability is determined by the inability to predicate a known (conceivable in principle) property about a known theoretical entity. The scientists possess the proper theoretical terms c_i, the proper predicates P_j and functions F_k, but cannot make the appropriate predication. A community of cognitive agents may have limited ability to relate the predicates to theoretical terms correctly. In this sense, this is an inconceivability of alternatives due to the incomplete set of possible predications, when the theoretical terms are the right ones, as well as the predicates and functions. One potential way of expanding the extension of a given predicate is by conjecturing that two different

theories refer to the same theoretical term, which therefore obeys the same set of rules of inference. Imagine we have a set of theoretical terms and a set of predicates and two theories that operate on them, but attribute different truth values to the same claims: $\sigma_T = \langle c_1, c_2, P_1, P_2 \rangle$, such that T states that $P_1(c_1) \wedge P_2(c_2)$ is true or in syntactic notation: $P_1(c_1) \wedge P_2(c_2) \nvdash_D \perp$ and one of its alternatives T' with the same signature, $\sigma_{T'} = \langle c_1, c_2, P_1, P_2 \rangle$ states that, syntactically, $P_1(c_2) \wedge P_2(c_1) \nvdash_D \perp$. Scientists in D are unable to conceive (or understand) $P_1(c_2)$ and $P_2(c_1)$, so they do not attribute meaning or truth values to these statements.

We believe that all the scenarios above accurately characterize *PUA*, but the following section focuses on a version of scenario 4, specifically the weak "predication inconceivability." To expand their predictive capabilities, scientists can identify the theoretical terms used by two theories and envision possible unification, whether ideal or real. However, a quick remark about scenarios 1-3 is in order. The cases in which new terms or predicates are needed are probably too strong for the eliminative inference proffered here. This eliminative inference does not apply in cases of major conceptual and nomological revolutions in science. This paper suggests that certain situations, including those discussed by Stanford, are susceptible to eliminative inferences and are closer to scenario 4 than to stronger forms of inconceivability, such as scenarios 1, 2, or 3.

4. *PUA* in the intertheoretical mill: unification, fragmentation, effective fields, and inconsistency

Unification is the virtue of a new scientific theory, T_σ (or hypothesis), to represent multiple phenomena that seemed unrelated before the introduction of T_σ. This new theory, T_σ, is created by combining two existing theories, T_1 and T_2, which differ in their theoretical terms, predicates, functions, and empirical support, E_1 and E_2, and each has its own U function that creates two sets of *UAT*: $\mathcal{T}_{T_1}$ and $\mathcal{T}_{T_2}$.

The possibility that two different theories can be unified acted both as a theoretical posit and as a concrete accomplishment in the history of

science after the Scientific Revolution. One can read the history of science as a partial history of successive unifications. Still, the story of physics in the 20th and 21st centuries can hardly be told without stressing the desire for unification: Einstein's unified field theory, various Grand Unified Theories (*GUT*), Supersymmetry, Superstring Theory, Canonical Quantum Gravity, and many more. Other theories, such as statistical mechanics, quantum mechanics, or quantum field theory, also have unification as one of their motivations.[16]

There are cases of unification akin to scenario 4 in biology. Foremost, the Modern Synthesis in biology brought together previously fragmented biological subfields into a coherent framework centered on the concept of evolution. Previous areas, such as genetics, paleontology, systematics, and embryology, operated largely independently, with conflicting theories about how life evolved and radically different empirical support. The progress between the 1930s and 1950s integrated Mendelian genetics with Darwinian evolution, solving disputes about the mechanisms of inheritance and natural selection.[17] The modern synthesis has eliminated contradictions between genetics and evolution, demonstrating that mutations and recombination provide the raw material for natural selection, and has linked microevolution to macroevolution, thereby bridging the gaps between genetics and paleontology. As B. Mayr suggested, the synthesis straightened out conflicts and disagreements between genetics and evolution, so "a united picture of evolution emerged" (Smocovitis 1992). This unified framework remains the foundation of modern evolutionary biology.

In physics, even given possible troubles in the paradise of unification, most physicists would endorse an architectonic representation of known interactions that can be read as a progressive history toward unification. After confirming the existence of four fundamental physical

[16] A book-length analysis of unification in physics and biology is (Morrison 2000). Recent analyses based on explanation, theories of truth, and Bayesianism are: (Bangu 2017; Blanchard 2018; Patrick 2018; Schupbach 2005).

[17] Key figures like Th. Dobzhansky, E. Mayr, J. Huxley, and G. Simpson played crucial roles in demonstrating how genetic variation and selection drive evolutionary change. See (Morrison 2000; Plutynski 2005; Smocovitis 1992).

forces—all the other forces being merely apparent or derivative from these: electromagnetism (being already unified), gravity, the strong nuclear force, and the weak nuclear force, in the first half of the 20th century and developing accurate theories of these forces for each of them: "the aim of physics is now to produce theories which unify these forces, which show, ultimately, that there is at base only one fundamental force in the universe, which has come to display itself as if it were many different forces" (Maudlin 1996, 129). This is the intuition that disparaged empirical phenomena E_1 and E_2 may be explained by a common 'structure' for which scientists strive to find a representation within theory *Tu*.

Within theoretical physics itself, unification can be understood in several ways. For example, some unificatory programs were designed to unify fundamental fields, while others aimed to unify matter with fields, and yet others were premised on even stronger assumptions and endeavored to unify gravity with all the other known quantifiable fields. One can see successful unifications in physics and biology. S. Glashow suggested that in the 1950s, after the massive success of quantum field theories, physics was "patchy":

> The study of elementary particles was like a patchwork quilt. Electrodynamics, weak interactions, and strong interactions were clearly separate disciplines, separately taught and separately studied. There was no coherent theory that described them all. Developments such as the observation of parity-violation, the successes of quantum electrodynamics, the discovery of hadron resonances and the appearance of strangeness were well-defined parts of the picture, but they could not be easily fitted together (Glashow 1980, 539).

However, is unification a general principle in science? Today, enthusiasm for unification is less common among biologists and chemists, where fragmentation in specialized fields may be more pronounced in the sciences. There is significant fragmentation in specific disciplines, such as molecular genetics or oxidative metabolism, partly because the same processes do not operate uniformly across all orders of life or in the same manner. Nonetheless, some believe that biology has reached a level at

which a steady consolidation process will replace the fragmentation process. The most enthusiastic scientists see 'consolidation' as a sign of unification:

> Scientific progress is based ultimately on unification rather than fragmentation of knowledge. At the threshold of what is widely regarded as the century of biology, the life sciences are undergoing a profound transformation. They have long existed as a collection of narrow, even parochial, disciplines with well-defined territories. Now they are undergoing consolidation, forming two major domains: one extending from the molecule to the organism, the other bringing together population biology, biodiversity studies, and ecology. Kept separate, these domains, no matter how fruitful, cannot hope to deliver on the full promise of modern biology. They cannot lead to an appreciation of life in its full complexity, from the molecule to the biosphere, nor to the generation of maximal benefits to medicine, industry, agriculture, or conservation biology (Kafatos and Eisner 2004, 1257).

4.1. The 'austere unification' by term identification

Pairs of theories in Y_E with two sets of empirical support (E_1 and E_2), are potential candidates for unification, reduction, equivalence relation, approximation, and so forth, even if this is not yet accomplished. But each of these known theories has their own set of $\mathcal{T}_{T_1} = U_t(T_1)$ and $\mathcal{T}_{T_2} = U_t(T_2)$ raise an important question: what if in the future T_1 will enter into a inter-theoretical relation? Following the line of F. C. Kafatos and T. Eisner, the current proposal investigates the potential future unification of two unrelated theories, especially when no new predicates are needed to account for alternatives to a theory. When two theories are unified the new theory contributes to scientific progress, even when the unification is not fully realized (P. Kitcher 1999).

Assume that in *D,* there are two known and accepted theories with these signatures $\sigma_{T_1} = \langle c_1, P_1 \rangle$ and $\sigma_{T_2} = \langle c_2, P_2 \rangle$, with their two UAT of unconceived alternatives spaces $\mathcal{T}_1 = U_t(T_1)$ and $\mathcal{T}_2 = U_t(T_2)$. Within scenario 4, scientists can conceive the predicates needed for an alternative, but they cannot link them to the appropriate theoretical terms. Unification

through identification is a mechanism that relates two different domains of inquiry when the theoretical term c_1, quantified by T_1, is essentially the same as c_2, quantified by T_2:$c_1 = c_2$. In this way, the predicate P_1 from T_1 can now be used by T_2.

Originally, before we identified c_1 with c_2, T_1 and T_2 differed in their theoretical claims. T_1 states that $P_1(c_1)$ and T_2 states that $P_2(c_2)$. In this sense, both P_1 and P_2 are conceivable ideologies, and c_1 and c_2 are conceivable terms, but what is not conceivable are $P_1(c_2)$ and $P_2(c_1)$.

In this toy example, the number of alternatives is limited. It is also essential to see that other alternatives to T_1 are: $T_1' = \langle c_2, P_1 \rangle$ stating that $P_1(c_2)$ and $T_1^{"} = \langle c_2, P_1 \rangle$ stating that ~$P_1(c_2)$.

The eliminative inference is based on the idea that the unification achieved by identifying c_1 with c_2 eliminates alternatives to T_1 and T_2. Remember that both terms c_1 and c_2 are conceivable at this moment, and by this identification, we enlarge the space of our predication with these two new sentences: $P_1(c_2)$ and $P_2(c_1)$. However, we now have an inconsistency between pairs T_1' and T_2'. This inconsistency indicates that the space of alternatives to T_1 and T_2 is reduced after unification by identification.

4.2. PUA, UAT, the 'lush unification,' and minimal inconsistency with evidence

As *PUA* depends on the existence and the 'size' of *PUA* space $\mathcal{T}_T$ the antirealist can point to the inconsistency of the T with any (all?) of its alternatives: $\exists T'(T' \in \mathcal{T}_T \wedge \sim (T \wedge T')$. As we have seen in the previous examples, consistency in itself is not enough to restrain *PUA*, even when T and its alternatives are supported empirically by the same evidence. The case in which $Y_E$includes two known theories, which potentially can be unified, is worth exploring further. However, it is the role of empirical evidence within expected unification posit that can diminish the strength of the *PUA*. In a more idealistic case, the scientists hope a new theory T_σ, different from T_1 and T_2 will reduce the inconsistency between them and unify their empirical bases $E_1 \wedge E_2$ better than them taken individually:

$$pr(T_\sigma|E_1 \wedge E_2) > pr(T_1|E_1) \times pr(T_2|E_2)$$

In the least problematic case, the empirical evidence of the unifying theory T_σ should be the union of E_1 and E_2, and we do not assume that unification is triggered or conditioned by the occurrence of new empirical data beyond E_1 and E_2. Unification is the emergence of T_σ with its own new signature.

This is referred to here as the 'lush unification' of T_1 and T_2, when scientists hope that T_σ will subsume vastly different phenomena under the same theoretical framework. Although T_σ may or may not contain identification of theoretical terms, it is expected, hoped for, and deemed as a new theory. It comes with its own range of unconceived alternatives $\mathcal{T}_{T_\sigma}$, but this does not preclude us from comparing the $\mathcal{T}_{\mathcal{T}_1}$ and $\mathcal{T}_{\mathcal{T}_1}$ before and after the emergence of T_σ.

In this case of "lush unification," empirical evidence "cross-pollinates" into the space of alternative theories of the pair T_1 and T_2. Let us put this in the form used before: we can conditioned the minimal set of T_1 by evidence E_2 because of T_σ, and compare it with the minimal set of T1 without evidence E_2. This is a result that we do not demonstrate here:

$$\mathrm{MI}(T_1, \mathcal{T}_{T_1}, E_2) \supset \mathrm{MI}(T_1, \mathcal{T}_{T_1}) \text{ and } \mathrm{MI}(T_2, \mathcal{T}_{T_2}, E_1) \supset \mathrm{MI}(T_2, \mathcal{T}_{T_2})$$

therefore, the cardinality of the minimal inconsistency set of any of the two theories is reduced by the present of the evidence used by the other theory, when the two are unified by T_σ. The cardinality of the $\mathrm{MI}(\{T, \mathcal{T}_T\})$ is taken here as being strongly correlated with what Stanford means by "serious" UAT. A high $card(\mathrm{MI}(\{T, \mathcal{T}_T\}))$ means that given E, there is a significant number of unconceived alternatives to T. The "lush unification" creates a new unification theory T_σ that brings together previously unrelated theories and use the evidence of all these theories "collectively."

4.3. Austere and lush unification in search of some case studies

This paper is particularly focused on the unificatory ideal of science, arguing that *when* and *where* the unificatory ideal operates in a community

of scientists, the semantic eliminative inference is more powerful. As stated here, the unificatory posit is not the same as the logical positivist "unity of science" as we operate in *PUA* at the local level in D, rather than as a whole science. We know that any *D* has several theories that share some common features and some empirical evidence. The claim is that a community of scientists who adopts a local unificatory posit, as local as limited as it may be, is less prone to the problem of unconceived alternatives. A community of scientists who see every theory in D as insular and isolated will be more vulnerable to the issue of unconceived alternatives. Excessive semantic pluralism is conducive to stronger antirealism instrumentalism and a stronger *PUA* (Ruhmkorff 2019).

The paradigmatic case that comes to mind is the unification of the theory of light and the theory of electromagnetic waves. One way to unify these two domains is to postulate an identity between the light wave and the electromagnetic wave, a hypothesis advanced by Maxwell through the introduction of the displacement current, a theoretical term. A step further, Maxwell could identify two other theoretical terms: the luminiferous ether and the electromagnetic ether. This introduced new "ideologies" to the previously unrelated theories of electromagnetic waves and optics, in the sense that optical concepts were applied to electromagnetic waves and vice versa. More importantly, none of the previous theories were surveyed unscathed: Ampere's law was modified, and some aspects of interference in optics were adjusted accordingly. This is a case of non-reductive unification, in which neither of the two theories reduces the other (Morrison 2000, 78). We can imagine that this process eliminates some alternatives to theories in optics and electromagnetism, such as theories about the transfer of energy in the two ethers, the speed of propagation in the two media, and ultimately the very idea of polarization of light and EM waves. In all these cases, constraints from optics "shaved off" alternatives in electromagnetism and the other way around, such that the *UAT* space of the new theory was reduced.

The second example comes from the debate about variation in a biological population. Before the 1900s, there were mainly two theories, each with its own followers. The Darwinians, a group that included A. Weismann and Fr. Galton, believed that selection alone produced the

change from one generation to the next. The Mendelians (a much smaller group, represented by W. Bateson and his competitor W.F.R. Weldon) believed that something other than selection was the leading cause of this change: mutation was a possible candidate. However, heritable traits and natural selection, as theoretical terms, were considered. Staunch Darwinians were gradualists and rejected the idea that mutation can play any role in evolution. The synthesis of these two theories was made possible much later, when people began to consider the genetic basis of evolution. This likely occurred with the work of K. Pearson, S. Wright, and R. Fisher in the late 1920s. The unification posit was to conceive that at least a slight selection pressure and heredity could contribute *together* as explanatory factors of change in population. Therefore, more recent theories of inheritance would attribute predicates such as generation, inheritance, growth, and development, which are all present in the Darwinian pangenesis theory, to a different theoretical term: a shared germinal source, or "hereditary particles," and not to the development of an organism's tissues (Stanford 2006, 68).

More importantly, once Weldon started to disagree with Bateson on the foundations of Mendelianism, he built his alternative on Francis Galton's 'ancestral heredity,' in which hereditary information from a distant ancestor is reduced by half with each generation and mixed during mating. Unlike the discrete, binary Mendelian traits, Weldonian traits vary continuously and follow a normal distribution. This constituted a 'conceived but unaccepted' alternative to Batesons dominant view of genetic determinism. Even more enticing is to consider Weldon's theory, following his untimely death in 1906, as an unexplored alternative to Bateson's theory. Gr. Radick has recently explored this counterfactual history (Radick 2022). Had Weldon lived, he might have produced a different synthesis of evolution and genetics. The fictional "Weldonian genetics" would have been more unifying than Bateson's genetics because it would eliminate alternatives to Mendel's theory that are more incompatible with evolution.

Finally, one can summon a third example of 'clipping' the *UAT* space from the quantum gravity program. As both quantum physics and general relativity lack the necessary generality, physicists frequently conjecture that the two theories will eventually be unified in an integrated

quantum gravity theory. This unificatory posit is more often marshalled than the instrumentalist view about the theories or the sheer search for their 'corrective' alternatives. One plausible venue here is scenario 1, where a new conceptual basis and ideology will be needed. The discrete geometry of space, the separation of space and time, the emergence of spacetime, dualities, and holographic principles could potentially be components of such a new theory. However, one can also speculate that we are closer to scenario 4 in current physics, which endorses more conservative unification and a less disruptive revolution. In such cases, the theoretical terms of quantum mechanics and general relativity "work together" but are predicated on different principles. Therefore, one can think that quantum and relativity elements would generalize well in a subsequent program and explore how those elements function together to generate the whole structure of the new theory: "Next one can explore the more general structures that can be obtained by loosening the constraints imposed in the current theory on one or another of the components that goes to make up the theoretical framework" (Lawrence Sklar 2000, 112).

If scenario 4 applies to quantum gravity, then one can see that, based on constraints from the other domain, unconceived alternatives, such as quantum theory, are constrained by requirements from general relativistic considerations, and vice versa. When quantum physicists can identify theoretical terms from quantum theories with theoretical terms from general relativity (e.g., entropy, energy, information, etc.), the quantum alternatives are constrained and restricted, even if quantum alternatives are still unconceived.

4.4. Coda on language dependence and conceptual spaces

There are several loose ends to this argument. First, scenarios 1, 2, and 3 are more frequently encountered in the history of science. Second, one can only speculate that the more pluralist fragmentation posit, somewhat the opposite of the unificatory posit, if dominant in an epoch, would enlarge the number of possible alternatives to a theory and make science more vulnerable to Stanford's PUA. Whether we live in the fragmentation of

the scientific epoch or a more unificatory one is an empirical question, tricky to address here. This paper has no issue with the fragmentation of science, but it shows that *NI* is weakened when other perspectives dominate a scientific discipline. If the current science is dominated by the disunity posit, the *PUA* space is augumented by fragmentation.

Second, another problem with this approach is its dependence on the first-order language, including its vocabulary and semantics. Moving to a semantic view may solve this problem, as models are not linguistic units of analysis. The 'conceptual spaces' mentioned before and their recent incarnations are promising candidates in this respect (Gärdenfors 2000, 2014; Zenker and Gärdenfors 2015). Theoretical terms are not only regions of conceptual space but they can be equipped with a geometry (convexity) and a metric. Geometric, non-linguistic representation can represent knowledge and the inconceivability of alternatives to theories. Gärdenfors and his collaborators claim that the qualities of objects (mainly their theoretical terms) can be represented without presuming an internal language. As a prospective alternative to the semantic approach, the conceptual space approach enables a more robust evaluation of what it means to embed and, ultimately, unify two theories within a larger and richer theoretical structure.

5. Conclusion

The present proposal attempts to weaken K. Stanford's problem of unconceived alternatives (*PUA*) by showing that some posits (called in the antirealist literature "standards" by M. Massimi or "substantive assumptions" by Stanford), such as unification and consistency, when adopted even tacitly in domain *D*, reduce the relevance and number of alternatives to *D*'s accepted theories. To do so, we need to think more holistically and see a theory T_1 as part of a field of theories in its unconceived alternatives enter in an intertheoretical relationship based on how T_1 relates to other accepted theory (T_2) in *D*. In this proposal, alternatives to T_1 are associated with alternatives to T_2, at least in the case of the most conservative case where T_1 and T_2 share the vocabulary but have different signatures (scenario 4 above). In the case of a "lush

unification," scientists posit the hypothesis that the two theories, T_1 and T_2, will be unified and replaced with a new theory T_σ that will bring the empirical support E_1 and E_2 together. This will reduce the space of alternatives to both T_1 and T_2. The overall goal of the present argument is to ease the PUA for cases where scientists in domain *D* endorse some (normative) 'expected posits' such as unification, consilience, or parsimony of future theories in *D*.

6. References

Bangu, S. (2017). Scientific Explanation and Understanding: Unificationism Reconsidered. *European Journal for Philosophy of Science* 7 (1): 103–26. https://doi.org/10.1007/s13194-016-0148-y.

Benferhat, S., Dubois, D., & Prade, H. (1997). Some Syntactic Approaches to the Handling of Inconsistent Knowledge Bases: A Comparative Study Part 1: The Flat Case. *Studia Logica* 58 (1): 17–45. https://doi.org/10.1023/A:1004987830832.

Bhakthavatsalam, S., and Kidd, I. J. (2019). Science, Realism, and Unconceived Alternatives: Introduction to the Special Issue on Unconceived Alternatives. *Synthese* 196 (10): 3911–13. https://doi.org/10.1007/s11229-019-02353-z.

Blanchard, T. (2018). Bayesianism and Explanatory Unification: A Compatibilist Account. *Philosophy of Science* 85 (4): 682–703. https://doi.org/10.1086/699157.

Bueno, O. (2015). Belief Systems and Partial Spaces. *Foundations of Science* 21 (1): 225–36. https://doi.org/10.1007/s10699-015-9416-0.

Carnap, R. (1995/1966). *An Introduction to the Philosophy of Science*. New York: Dover Publications.

Cartwright, N. (1999). *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.

Chakravartty, A. (2008). What You Don't Know Can't Hurt You: Realism and the Unconceived. *Philosophical Studies* 137 (1): 149–58. https://doi.org/10.1007/s11098-007-9173-1.

Da Costa, N. C., & French, S. (2003). *Science and Partial Truth: A Unitary Approach to Models and Scientific Reasoning*. Oxford University Press. 273 pp.

Dellsén, F. (2019). Should Scientific Realists Embrace Theoretical Conservatism? *Studies in History and Philosophy of Science Part A* 76 (August): 30–38. https://doi.org/10.1016/j.shpsa.2018.09.005.

Devitt, M. (2011). Are Unconceived Alternatives a Problem for Scientific Realism? *Journal for General Philosophy of Science* 42 (2): 285–93. https://doi.org/10.1007/s10838-011-9166-9.

Dupré, J. (1993). *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, Mass.: Harvard University Press.

Forber, P. (2007). Forever Beyond Our Grasp?: Review of P. Kyle Stanford (2006), Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives. *Biology & Philosophy* 23 (1): 135–41. https://doi.org/10.1007/s10539-007-9074-x.

Friedman, M. (1974). Explanation and Scientific Understanding. *Journal of Philosophy* 71 (17): 5–19.

Friedman, M. (2001). *Dynamics of Reason*. Center for the Study of Language and Inf.

Gardenfors, P. (1990). Induction, Conceptual Spaces, and AI. *Philosophy of Science* 57 (1): 78–95.

Gärdenfors, P. (2000). *Conceptual Spaces the Geometry of Thought*. A Bradford Book. Cambridge, Mass: MIT Press.

Gärdenfors, P. (2014). *The Geometry of Meaning: Semantics Based on Conceptual Spaces*. The MIT Press. https://doi.org/10.7551/mitpress/9629.001.0001.

Glashow, S. L. (1980). Towards a Unified Theory: Threads in a Tapestry. *Reviews of Modern Physics* 52 (3): 539. http://dx.doi.org/10.1103/RevModPhys.52.539.

Godfrey-Smith, P. (2007). Recurrent Transient Underdetermination and the Glass Half Full. *Philosophical Studies* 137 (1): 141–48. https://doi.org/10.1007/s11098-007-9172-2.

Halvorson, H. (2013). The Semantic View, If Plausible, Is Syntactic. *Philosophy of Science* 80 (3): 475–78. https://doi.org/10.1086/671077.

Hartmann, S. (2001). Effective Field Theories, Reductionism and Scientific Explanation. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 32 (2): 267–304. https://doi.org/10.1016/S1355-2198(01)00005-3.

Hempel, C. (1966). *Philosophy of Natural Science*. 1st ed. Pearson.

Kafatos, F. C., and T. Eisner. (2004). Unification in the Century of Biology. *Science (New York, N.Y.)* 303 (5662): 1257.

Kao, M. (2019). Unification Beyond Justification: A Strategy for Theory Development. *Synthese* 196 (8): 3263–78. https://doi.org/10.1007/s11229-017-1515-8.

Kitcher, P. (1999). Unification as a Regulative Ideal. *Perspectives on Science* 7: 337–48.

Kitcher, P. (1981). Explanatory Unification. *Philosophy of Science* 48 (December): 507–31.

Kitcher, P. (1995). *The Advancement of Science: Science Without Legend, Objectivity Without Illusions*. Oxford University Press.

Kukla, A. (2010). A. Kukla's Review of 'Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives', by P. Kyle Stanford. *Mind* 119 (473): 243–46.

Lutz, S. (2014). What's Right with a Syntactic Approach to Theories and Models? *Erkenntnis* 79 (S8): 1475–92. https://doi.org/10.1007/s10670-013-9578-5.

Lutz, S. (2017). What Was the Syntax-Semantics Debate in the Philosophy of Science About? *Philosophy and Phenomenological Research* 95 (2): pp. 319–352. https://www.jstor.org/stable/48578792.

Magnus, P. D. (2006). What's New about the New Induction? *Synthese* 148 (2): 295–301. https://doi.org/10.2307/20118691.

Massimi, M. (2021). Realism, Perspectivism, and Disagreement in Science. *Synthese* 198 (S25): 6115–41. https://doi.org/10.1007/s11229-019-02500-6.

Maudlin, T. (1996). On the Unification of Physics. *Journal of Philosophy* 93 (3): 129–44.

Mizrahi, M. (2022). Theoretical Virtues in Scientific Practice: An Empirical Study. *British Journal for the Philosophy of Science* 73 (4): 879–902. https://doi.org/10.1086/714790.

Morrison, M. (2000). *Unifying Scientific Theories: Physical Concepts and Mathematical Structures*. Cambridge University Press.

Patrick, K. (2018). Unity As An Epistemic Virtue. *Erkenntnis* 83 (5): 983–1002. https://doi.org/10.1007/s10670-017-9923-1.

Plutynski, A. (2005). Explanatory Unification and the Early Synthesis. *The British Journal for the Philosophy of Science* 56 (3): 595–609. http://bjps.oxfordjournals.org/cgi/content/abstract/56/3/595.

Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth.* Routledge.

Quine, W. V. (1951). Ontology and Ideology. *Philosophical Studies* 2 (1): 11–15. https://doi.org/10.1007/BF02198233.

Quine, W. V., and J. S. Ullian. 2007. *Ţesătura Opiniilor. (*Romanian edition of *Web of Belief)*. Translated by Mircea Dumitru. Pitești, Romania: Paralela 45.

Radick, G. (2022). *Disputed Inheritance: The Battle over Mendel and the Future of Biology*. Chicago, IL: The University of Chicago Press.

Roche, W. and Sober, E. (2017). Explanation = Unification? A New Criticism of Friedman's Theory and a Reply to an Old One. *Philosophy of Science* 84 (3): 391–413. https://doi.org/10.1086/692140.

Rowbottom, D. P. (2019). Extending the Argument from Unconceived Alternatives: Observations, Models, Predictions, Explanations, Methods, Instruments, Experiments, and Values. *Synthese* 196 (10): 3947–59. https://doi.org/10.1007/s11229-016-1132-y.

Ruhmkorff, S. (2019). Unconceived Alternatives and the Cathedral Problem. *Synthese* 196 (10): 3933–45. https://doi.org/10.1007/s11229-015-0947-2.

Saatsi, J. (2019). Historical Inductions, Old and New. *Synthese* 196 (10): 3979–93. https://doi.org/10.1007/s11229-015-0855-5.

Saatsi, J., Psillos, S., Winther, R. G., & Stanford, P. K. (2009). Grasping At Realist Straws. *Metascience* 18 (3): 355–90. https://doi.org/10.1007/s11016-009-9299-1.

Schindler, S. (2018). *Theoretical Virtues in Science: Uncovering Reality Through Theory*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108381352.

Schindler, S. (2022). Theoretical Virtues: Do Scientists Think What Philosophers Think They Ought to Think? *Philosophy of Science* 89 (3): 542–64. https://doi.org/10.1017/psa.2021.40.

Schupbach, J. N. (2005). On a Bayesian Analysis of the Virtue of Unification. *Philosophy of Science* 72 (4): 594–607.

Sklar, L. (1981). Do Unborn Hypotheses Have Rights? *Pacific Philosophical Quarterly Anc Personalist (The) Los Angeles, Cal.* 62 (1): 17–29. http://cat.inist.fr/?aModele=afficheN&cpsidt=12557219.

Sklar, L. (2000). *Theory and Truth: Philosophical Critique Within Foundational Science*. Oxford; New York: Oxford University Press.

Smocovitis, V. B. (1992). Unifying Biology: The Evolutionary Synthesis and Evolutionary Biology. *Journal of the History of Biology* 25 (1): 1–65. http://www.jstor.org/stable/4331201.

Stanford, P. K. (2006). *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. New York: Oxford University Press, USA.

Stanford, P. K. (2009). Scientific Realism, the Atomic Theory, and the Catch-All Hypothesis: Can We Test Fundamental Theories Against All Serious Alternatives? *Br J Philos Sci* 60 (2): 253–69. https://doi.org/10.1093/bjps/axp003.

Stanford, P. K. (2015a). Catastrophism, Uniformitarianism, and a Scientific Realism Debate That Makes a Difference. *Philosophy of Science* 82 (5): 867. https://doi.org/10.1086/683325.

Stanford, P. K. (2015b). Unconceived Alternatives and Conservatism in Science: The Impact of Professionalization, Peer-Review, and Big Science. *Synthese* 196 (August): 3915–32. https://doi.org/10.1007/s11229-015-0856-4.

Stanford, P. K. (2021). Realism, Instrumentalism, Particularism: A Middle Path Forward in the Scientific Realism Debate. In *Contemporary Scientific Realism: The Challenge from the History of Science*, edited by Timothy D. Lyons and Peter Vickers, 0. Oxford University Press. https://doi.org/10.1093/oso/9780190946814.003.0010.

Van Fraassen, B. C. (2008). *Scientific Representation: Paradoxes of Perspective*. Oxford; New York: Clarendon Press; Oxford University Press.

Vickers, P. (2013). A Confrontation of Convergent Realism. *Philosophy of Science* 80 (2): 189–211. https://doi.org/10.1086/670297.

Votsis, I. (2007). J. Votsis's Review of 'Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives' by K. Stanford. *International Studies in the Philosophy of Science* 21 (1): 103–6.

Zamora Bonilla, J.P. (2019). Realism Versus Anti-Realism: Philosophical Problem or Scientific Concern? *Synthese* 196 (10): 3961–77. https://doi.org/10.1007/s11229-015-0988-6.

Zenker, F., & Gärdenfors, P., eds. (2015). *Applications of Conceptual Spaces The Case for Geometric Knowledge Representation*. Cham: Springer.

Revista Română de Filosofie Analitică,
Romanian Journal of Analytic Philosophy
Vol. XVI, Iulie-Decembrie 2022, Nr. 2, p. 119-141
ISSN (ediția electronică): 1843-9969,
ISSN (ediția tipărită): 1844-2218
DOI: 10.62229/rrfaxvi-2/6

THE DOUBLE EMPATHY PROBLEM AS A DIALOGIC SENSE-MAKING STYLE ASYMMETRY

RADU NEDESCU[1]

Abstract: The occurrence of social comprehension difficulties when people living with autism, henceforth: autistics,[2] interact with neurotypicals motivates the re-emergence of key questions about the mind and its interaction with other minds; what are minds and how do they relate to the world and others? The disruption of smooth social interaction brings forth the question of *how is a mind able to socially interact* and this question motivates one to tacitly provide a definition of *what a mind is.* This is visible in Simon-Baron Cohen's exposition of the *theory of mind* theory, henceforth: ToM, in his book *Mindblindness*. In this book, Baron Cohen states that autistics have at least a degree of mindblindess and that mindreading[3] is the means through which the mind relates to other minds.[4] His tacit descriptions of ontological properties of the mind, henceforth: ontological descriptions or assertions, are utterly different from those provided by enactivists and by those who contribute to the 20th century tradition of phenomenology, henceforth: phenomenologists.[5] The tension between the ontological descriptions of ToM Theory and those provided by enactivists and phenomenologists has led to a thriving battle ground.

This article's key aim is to provide descriptions that facilitate enactivist or phenomenological analyses that engage with the double empathy problem

[1] Radu Nedescu has received a Master degree from KU Leuven, Institute of Philosophy. His research focuses on phenomenology and autism.

[2] Quirk, "Results and Analysis of the Autistic Not Weird 2022 Autism Survey - Autistic Not Weird"; Chapman and Bovel, "Neurodiversity, Advocacy, Anti-Therapy."

[3] Baron Cohen's version of ToM theory is the key cognitive model used for explaining social interaction difficulties in autism.

[4] Baron-Cohen, 1–5, 21–22, 26–30.

[5] Gallagher, "Understanding Interpersonal Problems in Autism."

hypothesis. To bring its aim to fruition, I follow three steps. Firstly, I define the approaches and concepts I use: phenomenology, enactivism, and the double empathy problem. Secondly, I argue in favor of using phenomenology and enactivism for explaining social difficulties in autism by presenting two, at least prima facie, disadvantages of Baron Cohen's articulation of ToM theory; one disadvantage stems from the ethical implications of his ontological assertions and the other stems from his ontological assertions. Thirdly, I describe autistic-neurotypical social interactions in a non-pathologizing manner by performing an enactivist analysis of the double-empathy problem surrounding autistic-neurotypical social interactions.

Keywords: double empathy problem, enactivism, dialogic sense-making.

1. Introduction

1.1. Defining phenomenology

The continental 20st century tradition of phenomenology has been created by the mathematician and philosopher Edmund Husserl whose goals were similar to those of the mathematician and philosopher Gottlob Frege.[6] Both of them have aimed to use precise abstract descriptions for providing foundations for STEM activities and both have vehemently rejected psychologism.[7] These two pioneers diverged regarding their analysis object and methodology. Frege's analysis object was mathematics, he aimed at establishing rigorous conceptual foundations for mathematics. Husserl's analysis object was the mind and he aimed at establishing rigorous conceptual foundations for scientific practice in general.[8] Frege developed the contemporary core methodological tools for formal logical and linguistic analysis.[9] Husserl invented a method for detecting and analyzing those series of interconnected traits, henceforth: structures, that

[6] McIntyre, "Husserl and Frege."
[7] Mohanty, "Husserl and Frege."
[8] Gelan, "The Idea of Rigorous Science in Husserl's Phenomenology and Its Relevance for the Other Sciences."
[9] Cook, "Frege's Logic."

define the necessary, henceforth: invariant, traits of the mind or of specific mental states and acts.[10]

Husserl was not an introspectionist because he was not interested in a specific person's experience, but in finding those structures that specify the invariant structures of the mind, its mental states, and its connection to the world.[11] Phenomenologists name mental states intentional acts because any conscious state, any *what is it like to be a mind S* is *a what is it like for S to be directed towards an object P*. Phenomenologists hold that mental states are intrinsically connected to the world; this applies even to brains in a vat. For a phenomenologist, a mind in a vat is a mind connected to the quasi-world produced by the stimulus given to that brain. Phenomenologists state that one's conscious connection to the world is a direct connection to it and not one mediated through representations. Indeed, in conscious experience, the mind's intentional connection to an object is given as a connection that is not mediated through representation. This is clearly so by noticing how one experiences the world; to give an example, one's experience of sitting in a café is not one of siting in a mental representation, i.e., in a series of signs that denote a café that is not consciously experienced.[12]

The seemingly bizarre assertion that one's mind is directly connected to the world becomes clear once one explains how phenomenologists define the term *world* and the locution *accessing the object as such*. The impact of how phenomenologists define them shapes the phenomenologists' description of how the gap between objects and the mind is traversed. Phenomenologists describe the gap between an object and the mind as traversed by the interaction between that object's manner of expressing itself to that mind and that mind's manner of receiving that object.[13] In contrast, classical philosophy of mind describes this gap as being quasi-traversed by the mind's act of generating a representation that connects that mind to that object.

---

[10] Husserl, *Experience and Judgment: Investigations in a Genealogy of Logic*.
[11] Zahavi, *Phenomenology*, 15, 34–38.
[12] Zahavi, 16–23.
[13] Husserl, *Ideas*; Husserl, *Analyses Concerning Passive and Active Synthesis*.

The phenomenological definition of the world, if made explicit, is that the *world* is the set that must always include all the following sets of members: minds, mental experiences, and objects that are not mental experiences. They are tacitly defined in this way because phenomenologists assert that minds cannot exist without at least one object and that objects cannot exist without at least one mind. For analyzing the reasons behind the aforementioned assertion, I recommend reading the chapter "Internalism, externalism, and transcendental idealism" from *Husserl's Legacy*.[14] Therefore, for phenomenologists, if a set has no member that is a mental experience or a mind, then that set cannot be equivalent to the world. The reverse is equally the case, if a set has no member that is not a mental experience or a mind, then that set cannot be equivalent to the world. Phenomenologists define the mind's *access to the object in itself as* that mind's access to the *object in itself* as that object expresses itself to that mind in a manner *receivable by that mind*. The explanations provided above are meant as mere clarifications of points often stressed by phenomenologists.[15]

1.2. Defining enactivism

Enactivism is a subfield of radically embodied cognitive science. Radically embodied cognitive science is that branch of cognitive science that describes cognitive processes without relying on mental representations, but on dynamic process occurring between an organism and its environment. Enactivism is that subfield of radically embodied cognitive science that took its starting point from the book *The Embodied Mind*.[16]

For enactivists, the mind is not *an empty name*, and they do not make *negative existential*[17] claims about theory of mind (ToM), they assume that

[14] Zahavi, *Husserl's Legacy*, 127–29.
[15] Merleau-Ponty, *Phenomenology of Perception*; Husserl, *Ideas*; Zahavi, *Phenomenology*.
[16] Varela, Thompson, and Rosch, *The Embodied Mind*; "Enactivism | Internet Encyclopedia of Philosophy," sec. 1. Core Commitments.
[17] For more about negative existentials, see Dumitru, M., and Kroon, F. (2008). What to say when there is nothing to talk about. *Crítica* (México, DF), 40(120), 97-109.

the mind results from "multiple kinds of physiological, sensorimotor, and interpersonal" processes that influence and are influenced by the environment.[18] Enactivists hold that an agent, i.e., an organism, is *coupled* to an environemnt if and only if that agent and that environment mutually influence each other.[19] In enactivist cognitive science, *enaction* is the organism's act of transforming the environment such that it fulfils its needs.[20] Organisms create meaning when they shape, couple with, and adapt to their environment.[21] This activity of creating meaning is named *sense-making*.[22] When at least two organism interact, their interaction becomes a quasi-autonomous system that generates new meaning; this manner of meaning production is named *participatory sense-making*.[23] To argue for the ontological adequacy of enactivist cognitive science is beyond this paper's scope.

1.3. Defining the double empathy problem hypothesis

The double empathy problem hypothesis states that one's low social comprehension success degree is not caused only by one's mindreading faculty, but also by how other participants socially interact with you. In other words, success in understanding the other's mental state is a result of a process that is co-authored by all participants. According to Damian Milton, "social subtext is never fully given as a set of a priori circumstances, but is actively constructed by social agents engaged in material and mental production."[24] He elaborates on this as follows:

> There is a tendency in the application of positivist methodologies in cognitive psychology and science to incorrectly assume that there is a set

[18] Di Paolo, Cuffari, and Jaegher, *Linguistic Bodies*, 17–18.
[19] Di Paolo, Cuffari, and Jaegher, 17–18, 21, 46,.
[20] Di Paolo, Cuffari, and Jaegher, 21–22, 46, 109–11.
[21] Di Paolo, Cuffari, and Jaegher, 17–18, 21, 32-36,.
[22] Di Paolo, Cuffari, and Jaegher, 32–36.
[23] Di Paolo, Cuffari, and Jaegher, 73–75.
[24] Milton, "On the Ontological Status of Autism," 884.

> of definable social norms and rules that exist for people to follow. […] The 'theory of mind' and 'empathy' so lauded in normative psychological models of human interaction refers to the ability a 'non-autistic spectrum' (non-AS) individual has to assume understandings of the mental states and motives of other people. When such 'empathy' is applied toward an 'autistic person', however, it is often widely inaccurate in its measure. Such attempts are often felt as invasive, imposing and threatening by an 'autistic person', especially when protestations to the contrary are ignored by the non-AS person doing the 'empathizing'.[25]

2. The two, at least prima facie, disadvantages of Baron-Cohen's ToM Theory

2.1 Baron-Cohen's tacit ontological descriptions

Before explaining the two disadvantages of using Baron Cohen's version of ToM model for analyzing social interaction in autism, I need to make explicit his ontological description of the mind and of its access to other minds. For him, the mind is that kind of agent that has states such as wanting, knowing, planning, and recognizing,[26] and the mind is situated "inside one's own head," namely, inside one's brain.[27] He asserts in his thought experiments that human minds occur in such manner that they cannot directly access other minds.[28] For him, social comprehension is the result of using the mental act named by him *mindreading*; this mental act is that of *interpreting* the actions of others as those of beings endowed with mental states.[29] Therefore, the mental act of interpreting the other assigns to that other mental states and these assigned mental states are implicit representations. Also, mindreading is most often performed unconsciously. That mindreading is a mental act of assigning mental states to the other, implies that the other exists as another for oneself exclusively due to one's

[25] Milton, 884.
[26] Baron-Cohen, *Mindblindness*, 1–5.
[27] Baron-Cohen, 27.
[28] Baron-Cohen, 21–24.
[29] Baron-Cohen, 1–5, 26–30.

own interpretative activity.[30] In other words, given his assumption that human minds have no direct access to the world and to other minds, for him, social comprehension also takes place exclusively inside one's mind.

2.2. The ethical disadvantage of classical ToM Theory

Parallel to the ontological tension between an enactivist and a ToM Theory description of autism, there is also an ethical tension surrounding classical ToM Theory and autism. The later tension is between ToM Theory's assertion that autistics suffer from mindblindess and the neurodiveristy movement's non-pathologising assertion that autism is not a disorder, but only a non-neurotypical neurological configuration. Pathologising is the act of implicitly or explicitly asserting that one suffers from a deficiency when in fact one just navigates the world differently. The neurodiversity movement is the movement that argues for the empowerment and social inclusion of people whose neurological configuration is highly different from that of neurotypicals. This movement's key claim is that conditions such as autism, ADHD, dyslexia, etc., are not disorders, but merely different manners of navigating the world.[31] I argue that ToM indeed faces the aforementioned ethical tension and that this tension's existence favors the usage of enactivist and phenomenological descriptions of autistic-neurotypical social interactions instead of those provided by Baron Cohen's ToM Theory.

The neurodiversity movement holds that during autistic-neurotypical social interactions, both autistics and neurotypicals face social comprehension difficulties because, according to this movement, the autistics' decreased social comprehension ability is not the result of the autistics' mindreading faculty in itself, but that of the decreased compatibility between the autistic and neurotypical manners of

[30] Baron-Cohen, 1–5, 21–30, 32–58.

[31] Milton, "On the Ontological Status of Autism"; Walker, *Neuroqueer Heresies*, sec. Throw Away the Master's Tools: Liberating Ourselves from the Pathology Paradigm; Neurodiversity: Some Basic Terms & Definitions; Defining Neurodiversity.

socializing. This description of autistic-neurotypical social interactions has been explicitly brought in the neurodiversity movement's discourse by Damian Milton's double empathy problem hypothesis.[32]

Baron-Cohen's model of the ToM describes autism as a disorder because his model asserts that the lower social comprehension degree found in autistics—unless accidental factors occur—is caused only by a deficient mindreading faculty, within the autistics' minds.[33] The ethical, at least prima facie, disadvantage of his model that I argue for is that its metaphysical commitments facilitate the pathologization of autistics. My argument involves answering the following questions: **1.** Is his ToM Theory right when stating that their lower social comprehension degree is caused only by their deficient mindreading faculty? **2.** Is his ToM Theory's metaphysical description of the mind forcing classical ToM Theory to affirm the just aforementioned assertion? **3.** Is his model of the ToM right in describing autism as a social interaction disorder or is classical ToM Theory pathologising autistics?

1. His ToM Theory is not flawed by stating that the lower social comprehension degree of autistics is caused only by the autistics' mindreading faculty because social comprehension involves the interpretation of other participants' expressions—it can be both verbal or non-verbal—and this interpretative activity's success depends on the compatibility between one's *interpretative schema* and the expression to be interpreted. I call an *interpretative schema* a social agent's set of tacit inference rules that tell that social agent how to convert the other social agent's expressions into an interpretation of that specific expression, this interpretation can also be a prediction of the other agent's actions or intentions. If the other agent's expression is not compatible with one's interpretative schema, then the interpretative act fails, but neither due to the interpreter alone nor due to the other agent alone.

Autistics not requiring substantial support are certainly able to understand that other people have their own mental states; in such cases, the mindblindess attributed to them is partial and it is considered to only

[32] Milton, "On the Ontological Status of Autism."
[33] Baron-Cohen, *Mindblindness*, 1–7, 59–63, 69–71.

decrease the quality of their social comprehension.[34] In such cases, it cannot be stated that only the autistic interpreter is at fault because no interpretation can happen without the bidirectional influence between one's interpretative schema and the content to be interpreted; for there to be a successful interpretation, there needs to exist an active interpretative schema that derives an interpretation from the expressed content and there needs to exist an expressed content that is in such a manner that the interpretative schema derives from it a successful interpretation. Because of this bidirectional influence in any interpretative activity, and especially during social interpretative activities between highly different manners of interpreting what is relevant during social interactions, no agent alone can be the cause of one's low interpretative success degree. However, strictly speaking, no agent can be the cause of any low interpretative success degree because interpretative success is always the result of the interaction between agents; therefore, it cannot result from each agent's actions taken in isolation.[35]

2. His ToM Theory's ontological assertion does not force ToM Theory to assert that mindreading cannot be influenced by the content that has to be interpreted by that mindreading agent. This is so because his ToM Theory's assertion that social comprehension is an act entirely performed inside one's mind without access to others' mental states in themselves does not entail that the mind performing that mindreading act cannot be influenced by the content it has to interpret. Therefore, his model of ToM can assert, without contradicting its ontological assertions, that not only an autistic's mindreading faculty causes that autistic's low social comprehension degree.

However, classical ToM Theory's model of social comprehension, by focusing only on an agent's interpretative acts in isolation, incentivizes interpretations according to which the mindreading faculty is the only cause for those social comprehension difficulties seen in autistics. This is the ethical disadvantage of at least his ToM Theory, namely, that it

[34] Fuchs, "Pathologies of Intersubjectivity in Autism and Schizophrenia," 197–98.
[35] Heasman et al., "Towards Autistic Flow Theory"; Milton, "On the Ontological Status of Autism."

incentivizes interpretations that place the fault for an autistic's social comprehension difficulty on the autistic person alone. In fact, the fault is not to be ascribed to the agents, but to the lower compatibility degree between each agent's social practices.

3. Given the answers to points 1 and 2, it cannot be stated that autism is a social interaction disorder because an autistic's lower interpretative success degree cannot be due to that autistic's social interpretation faculties in themselves, but only due to the decreased compatibility between autistic and neurotypical social practices. In the realm of social interaction between agents with highly different practices and needs, the occurring difficulties in social comprehensions cannot be found inside an agent.[36]

The aforementioned ethical tension favors the usage of enactivist and phenomenological descriptions of autistic-neurotypical social interactions, instead of those provided by his model of ToM, because enactivist and phenomenological descriptions already have a conceptual apparatus that emphasizes the aforementioned bidirectional mutual influence. Enactivist cognitive science, by centering around dynamic processes, provides a framework highly suitable for analyzing states of affair such as the social comprehension difficulties caused by the lower compatibility degree between social agents with highly different social practices. Due to the aforementioned reasons, I hold that it is more economical to use enactivist and phenomenological concepts to analyze autistic-allistic social interactions than adapting ToM Theory such that it acquires the fidelity needed for properly explaining and analyzing such social interactions.

2.3. The disadvantage of his ToM Theory's ontological assertions

The disadvantage of his ToM Theory's ontological description of the mind as lacking any direct access to other minds and as confined inside one's

[36] Milton, "On the Ontological Status of Autism"; Lynch, "Invisible Abuse: ABA and the Things Only Autistic People Can See."

brain is that his ToM Theory is less compatible with phenomenological and enactivist descriptions of autistic-allistic social interactions. The aforementioned decreased compatibility is a disadvantage for his ToM Theory because it decreases his ToM Theory's ability to analyze autistic social interaction from the perspective of autistics' conscious experience. Analyzing autistic social interactions from this perspective, by being able to make visible the perspective of autistics themselves, facilitates an exploration of autistic social interactions that does not pathologies them.

However, one can object that his ToM Theory's low compatibility with other relevant theories is not a disadvantage because his ToM Theory, unlike phenomenological and enactivist theories, provides a true ontological description of how minds are able to understand and predict the actions of other minds. My reply to this objection is that even if his ontological descriptions are true, his ToM Theory's low compatibility with enactivist and phenomenological descriptions is a disadvantage because rejecting such descriptions hampers one's comprehension of that part of the mind for which social interactions are meaningful, namely, that mind's conscious part. [37] It hampers it even if phenomenological and enactivist descriptions were false ontological descriptions of the asubjective world;[38] this is so since one can use phenomenological and enactivist approaches without granting their descriptions the status of objective ontological assertions.

For properly understanding autistic social interactions, it is necessary to also use phenomenological and enactivist concepts. Regarding enactivist concepts, this is so because they enable the analysis of the dynamics involved in autistic-neurotypical social interactions without reducing these interactions' complexity; this complexity is reduced when focusing only on the peculiarities of how autistics socially interact. Regarding phenomenological concepts, this is so because they were especially developed to and tailored for capturing the features of mental acts as consciously experienced by a mind.

---

[37] i.e., even if such approaches fail to provide descriptions that denote the ontological state of affairs.

[38] I do not believe that phenomenological and enactivist approaches lead to false metaphysical description.

For understanding social comprehension in autism, it is important to understand autistic mental acts as they are consciously experienced by autistic minds because understanding them decreases the distortions caused by neurotypical interpretations of autistic social interactions; such interpretations often risk to be distorted by the prevalent interpretative neurotypically informed frameworks.[39] It is only natural for misunderstandings to occur when agents with different social practices and interpretative frameworks socially interact.[40] Because an epistemic gap is involved during such interactions, it is important to take into account how other agents navigate the world.

3. An enactivist description and analysis of the double empathy problem

In this section, I will use enactivist concepts from the book *Linguistic bodies* to analyze Damian Milton's double empathy problem hypothesis.[41] Before using these concepts, I have to explain and define these concepts and present how the authors of this book describe the difficulties that autistics face during social interactions. First, I will use these concepts and descriptions to analyze Milton's articulation of autistic-neurotypical social interactions. Afterwards, I will use these concepts and descriptions to analyze those traits of autistic-social interactions influenced the most by the double-empathy problem.

3.1. Presenting the enactivist concepts I will use

The book *Linguistic Bodies* creates multiple concepts to describe mutually influencing dynamics that together form one's mental states and one's interaction with the world. Their descriptions involve both sense-making

[39] Lynch, "Invisible Abuse: ABA and the Things Only Autistic People Can See"; Walker, *Neuroqueer Heresies*, sec. Throw Away the Master's Tools: Liberating Ourselves from the Pathology Paradigm; Chapman and Bovel, "Neurodiversity, Advocacy, Anti-Therapy."
[40] Milton, "On the Ontological Status of Autism."
[41] Di Paolo, Cuffari, and Jaegher, *Linguistic Bodies*.

and participatory sense-making.[42] In the following, I am interested in those concepts created by the book's authors for describing intersubjective processes, i.e., processes that involve participatory sense-making. Below I will present these concepts by providing a unitary description of how they are interconnected. If I were to present each concept separately, I would risk reducing the enactivist interdependent dynamic descriptions to atomistic elements. The aforementioned enactivist dynamic descriptions are presented below.

When organisms successfully interact with each other by repeatedly performing the same participatory sense-making acts across time, these organisms' interaction dynamic form a stable pattern. The authors of *Linguistic Bodies* named this type of patterns *partial acts*. For multiple organisms to mutually apply the same *partial acts*, these partial acts need to become *normative partial acts*, they need to tell all the involved agents how to responds. The book's authors named strongly normative partial acts *interlocking social acts*.[43] All the types of participatory sense-making mentioned above do not require mindreading; they can be viewed as precursors of ToM. In their enactivist picture, mindreading is enabled by the most complex type of participatory sense-making; they named this type of participatory sense-making *dialogical sense making*.[44] Dialogical sense-making involves the turned-based verbal or non-verbal information exchange between participants and it can be performed only by linguistic bodies.[45] They define *a linguistic body* as an organism's set of "embodied and material patterns" through which that organism expresses itself to others either through speech or any other modality.[46] They named these patterns *utterances*.[47] However, to avoid confusion, I now rename them *communication patterns*.

Dialogical participatory sense-making involves the turn-based expression of *communication patterns*. This type of sense-making produces

[42] Di Paolo, Cuffari, and Jaegher, 32–204.
[43] Di Paolo, Cuffari, and Jaegher, 139–59, 150–51, 159.
[44] Di Paolo, Cuffari, and Jaegher, 172–75, 191, 195.
[45] Di Paolo, Cuffari, and Jaegher, 172–75, 191.
[46] Di Paolo, Cuffari, and Jaegher, 193.
[47] Di Paolo, Cuffari, and Jaegher, 173–75.

a more stable interaction pattern that solves those intersubjective tensions unsolvable by any other type of participatory sense-making act. Dialogical sense-making solves these tensions by better organizing the participants' sense-making production, by allowing only an agent per turn to emit a *communication pattern*. This creates an asymmetry between the agent emitting the *communication pattern* (the turn holder) and the rest of the participants (those who receive the turn holder's *communication pattern*).[48] This asymmetry enables a linguistic body to perceive the other as a linguistic body with distinct mental states and intentions. This is so because, when the turn-holder has a strong regulator role, the turn-holder leads the participants to recognize her/him as an autonomous agent. For the dialogue to continue, the audience also has to recognise the turn-holder as an autonomous agent.[49]

However, even in a dialogue, there is not guarantee that one knows how to produce adequate *communication patterns* and that others will adequately interpret these communication patterns. To increase the chances of enacting a smooth dialogue, linguistic bodies must resort to social interaction patterns that "precoordinate the expectations of producers and audience." These patterns are named *participation genre*, par example: "cooking together, eating together, finding seats at the theatre, coordinating labour, playing, etc."[50] To prevent a participation genre from failing, the turn holder has to strategically modify his/her *communication pattern*. To do so, a linguistic body has to apply *self-control*, namely, to act both as a producer and an interpreter. More precisely, *self-control* involves interpreting one's own *communication pattern* before producing it in order to increase one's social success.[51] In a dialogue, for ensuring that the participants are on the same page, they can use *reported communication patterns*, namely, they can repeat or modify a participant's previous *communication pattern*. By using *reported communication patterns*, linguistic bodies make their interpretations explicit, and, by doing so, they

[48] Di Paolo, Cuffari, and Jaegher, 169–75.
[49] Di Paolo, Cuffari, and Jaegher, 169–76, 193.
[50] Di Paolo, Cuffari, and Jaegher, 178.
[51] Di Paolo, Cuffari, and Jaegher, 184–86.

can coordinate their interpretations.[52] By coordinating them, linguistic bodies create dialogues that are both stable and dynamic. This coordination process is named *frame building*.[53] By interacting with each other, linguistic bodies can transform themselves in two ways, either by idiosyncratically adopting the other's *communication patterns*, i.e., by *incorporating* them, or by being changed by these patterns, i.e., by *incarnating* them. Too much incarnation leads to decreased autonomy and not enough incarnation makes one to be too different from others.[54]

3.2. Autistic-neurotypical dialogic sense-making as described in *Linguistic Bodies*

In chapter 10 from the *Linguistic Bodies*,[55] its authors state that the core participatory sense-making difficulty found in autistic-neurotypical social interactions is that of "co-construct[ing] and coregulate[ing] an interactive dissonance together with other participants."[56] More specifically, the key not adequately managed coregulation tension is that between "the regulator and regulated role."[57] This tension is not adequately managed "because of [a] clash between the autistic self-organization and embodiment" and the neurotypical "cultural habitus."[58] The coregulation difficulties occurring during autistic-neurotypical social interactions, by affecting the production of dialogic acts, lead to shortcomings in recognizing other participants "as autonomous sense-makers."[59] The book's authors explicitly state that neurotypicals have difficulties in recognizing autistics as autonomous sense-makers and that autistics have difficulties in recognizing neurotypicals as autonomous

[52] Di Paolo, Cuffari, and Jaegher, 186–90.
[53] Di Paolo, Cuffari, and Jaegher, 190.
[54] Di Paolo, Cuffari, and Jaegher, 191–94, 211–12.
[55] Di Paolo, Cuffari, and Jaegher, 261–77.
[56] Di Paolo, Cuffari, and Jaegher, 266.
[57] Di Paolo, Cuffari, and Jaegher, 266.
[58] Di Paolo, Cuffari, and Jaegher, 266.
[59] Di Paolo, Cuffari, and Jaegher, 266.

sense-makers. Based on these descriptions, they assert two hypotheses on the key participatory sense-making difficulties characterizing autistic-neurotypical social interactions:

1. Autistics would try to "cope with the inherent tensions of participatory sense-making between individual and interactive norms [either] by" regulating a social interaction too much (*over-shooting*) or not enough (*undershooting*).[60] In other words, autistics would either "attempt to resolve a particular tension as individual agents rather than in a joint act" or "withdraw momentarily to allow others to resolve the tension" instead of participating in the process.[61]
2. Autistics would better tackle the *pragmatic aspects* of a social interaction than the social interaction's *expressive aspects*. In other words, autistics would tackle those aspects that directly impact the interaction better than those that depend on "the relations between the participants."[62] They hypothesize this because the autistic-neurotypical social interaction involves coregulation difficulties.

3.3. The enactivist conceptualization and analysis of double-empathy problem

I start by quoting Damian Milton's definition of the double empathy problem. I do so to unpack his definition by using enactivist concepts from *Linguistic Bodies*. By unpacking his definition, I can analyze it by using enactivist concepts. His definition of the double empathy problem is the following:

> The 'double empathy problem': a disjuncture in reciprocity between two differently disposed social actors which becomes more marked the wider the disjuncture in dispositional perceptions of the lifeworld– perceived as

[60] Di Paolo, Cuffari, and Jaegher, 266.
[61] Di Paolo, Cuffari, and Jaegher, 266.
[62] Di Paolo, Cuffari, and Jaegher, 269.

> a breach in the 'natural attitude' of what constitutes 'social reality' for 'non-autistic spectrum' people and yet an everyday and often traumatic experience for 'autistic people'.[63]

The double empathy problem is a lived experience, is a series of intentional acts. These intentional acts are directed to the other sense-makers and to the social interaction as such; this interaction is a quasi-autonomous participatory sense-making dynamic. All intentional acts involve the bidirectional dynamic co-influence between the subject and the objects, between the embodied sense-maker and a specific part of the environment (including other sense-makers). The double empathy problem is a lived experience in which linguistic bodies have difficulties in co-constructing a shared sense because of unsolved participatory and dialogical sense-making tensions. To solve such tensions, linguistic bodies need to influence each other without inhibiting the other's shared sense-making production; this inhibiting occurs when the participatory and dialogic sense-making agents have incompatible sense-making and embodiment styles. These incompatibilities lead to the disruption of each linguistic body's expectations about the possible meaningful communication patterns of other linguistic bodies. This disruption is a disruption of the natural attitude that is experienced by both autistics and neurotypicals. However, according to Damian Milton, this disruption is "more sever for the non-autistic" because the disruption itself is an "unusual" experience for the neurotypical, but it is "a common experience" for "the autistic."[64] The autistics' familiarity with the disruption does not make it less traumatizing, on the contrary. Therefore, the intensity of the disruption is not proportional with its traumatic intensity. Interestingly, neurotypicals are often unaware that such disruptions occur because their lifeworld, by being hegemonic, is habitually believed by them as the only one; because of this, the disruption is "healed perceptually."[65] To be *healed perceptually* denotes the

[63] Milton, "On the Ontological Status of Autism," 884.
[64] Milton, 885.
[65] Milton, 885.

following state of affairs: the disruption of S's natural attitude is unnoticed by S because S tacitly interprets S's social interaction with J as an interaction whose hindrances **1.** must be caused by J's peculiar interaction manner and **2.** cannot be caused by S's social interaction manner. Therefore, *perceptual healing* is a process that alters one's interpretation of one's own social interaction. In Milton's own words, "a person who sees their interactions as 'normal' […] can apply the label on the 'other' locating the problem in them"[66]

The key factors that hinder the empowerment of autistic people, for Milton, are the following: **1.** "the normalization agenda" and stigmatization motivated by perceptual healing, **2.** "internalized oppression," **3.** the exclusion of autistics from producing knowledge on autistics.[67] In the rest of this sub-section, I will supplement Milton's analysis of these factors with enactivist concepts from *Linguistic Bodies*.

1. Perceptual healing, by placing the social interaction difficulty's cause entirely within the sense-maker that deviates from the majority's natural attitude, it makes it more likely for neurotypicals to deem autistics as "abnormal" and, therefore, to also "stigmatise" or "sanction" them. Perceptual healing, by effacing the fact that autistics are fully developed, yet different, linguistic bodies, it encourages neurotypicals to over-regulate the interaction dynamic by trying to convert autistic sense-making practices into neurotypical ones.[68] In other words, perceptual healing incentivizes the normalization of autistics.

2. Internalized oppression is the process through which an autistic alters his/her interpretation of oneself by adopting the neurotypical's tacit or explicit belief that autistics suffer from a disorder. In other words, autistic starts to believe, like many neurotypicals do, that autism is a pathology. According to Milton, this process "lead[s] to a self-imposed psycho-emotional disablement." In other words, internalized oppression destabilases how autistics apprehend their own abilities and needs.[69]

[66] Milton, 885.
[67] Milton, 885.
[68] Milton, 885.
[69] Milton, 885.

When internalized oppression takes place, autistics introject neurotypical sense-making practices and communication pattern. By introjecting them, autistic sense-makers start to apprehend situations through the neurotypical natural attitude, through a natural attitude that is neither their own nor compatible with their own manner of sense-making. I conclude, based on the aforementioned, that the neurotypicals' difficulty in adequately recognizing autistics as autonomous sense-makers does not emerge only due to participatory sense-making asymmetries, but also due to perceptual healing.

3. Knowledge production is the process of acquiring information about something or somebody, in this case, about autism and autistics. Milton's critique is that the production of knowledge about autism is not mainly done by autistics, but by neurotypicals that, instead of allowing autistics to contribute, they place them as "the 'product' of the industry, the thing' that is 'intervened' with."[70] The exclusion of autistics from knowledge production also entails the exclusion of autistics from the material production of practices that empower autistics. The coregulation difficulties occurring during autistic-neurotypical social interactions lead to the decreased occurrence of smooth dialogic sense-making. This decreased occurrence, by affecting the genuine recognition of autistics as autonomous sense-makers, impacts, at the macro level, the knowledge production about autistics.

3.4. Analyzing those autistic social interaction dialogic participatory sense-making practices that are most often involved in the double empathy problem

The dialogical participatory sense-making acts most often involved in the double empathy problem are those that often involve the internalisation of communication pattern originating from agents that inhabit a different lifeworld than one's own. This is so because the double empathy problem results from "the asymmetry" between how the "social actors" involved

[70] Milton, 885.

make sense of each other's social interaction style.[71] From an enactivist perspective, social comprehension is both performed through and the result of participatory sense-making acts. The dialogical participatory sense-making acts most often involved in the double empathy problem are the following: **1.** self-control, **2.** frame building, **3.** incarnating communication patterns.

1. Regarding the dialogic participatory sense-making act of self-control, there is a disjunction between the autistic persons' *communication pattern* producer role and the same autistic person's communication pattern interpreter role. The autistic's producer role is more inclined to express the autistic's own sense-making while the autistic's interpreter role is more inclined to enact neurotypical sense-making patterns. This is so because production, unlike interpretation, does not incentive one to focus on the other. The consequence of this disjunction is that autistics often cannot rely on their own sense-making style to apply self-control and this leads to an inner alienation, to a high tension between the content to be expressed and the manner of expressing it. To manage this tension, the autistic has to consume more energy and incorporate neurotypical *communication patterns*; in the neurodiversity community, this manner of tension management is named masking, i.e., acting in a neurotypical manner instead of being yourself.

2. Frame building is the dialogical practice of coordinating the participants' interpretations of their previous communication patterns; these interpretations are made explicit by using reported communication patterns. Because the sense-making styles of autistics and neurotypicals are not synchronized, they lead to disjunctions in how "the social world" is experienced.[72] Because of their difficulty in mutual coordination, the participants are less able to reach a mutually agreed upon interpretation of their previous utterances. The consequence of this is that their reported utterances are less able to be united into an intersubjectively shared interpretation frame and, therefore, the social interaction becomes less stable.

[71] Milton, 884.
[72] Milton, 884.

3. Regarding linguistic bodies' act of incarnating the communication patterns of others, the autistic, due to its upbringing in a neurotypical world, is incentivized to incarnate neurotypical communication patterns, namely, communication patterns dissonant to their autistic sense-making and participatory-sense-making style. There is also a notable difference regarding autistic and neurotypical participation genres.[73] While neurotypical participation genres change faster and tend to not focus on a singular theme, autistic participation genres tend to center around special interests or activities not requiring a fast co-regulation of "interactive dissonance together with [the] other participants".[74]

4. Conclusion

This article has presented two apparent disadvantages of Simon Baron-Cohen's description ToM Theory and analyzed the double empathy problem hypothesis by applying enactivist concepts to this hypothesis and to those intersubjective dynamics most often involved in autistic-neurotypical social interactions. I have explored the context surrounding the battle between using enactivism and phenomenology or ToM Theory for analyzing and describing autistic-neurotypical social interactions. There is an ethical concern surrounding the usage of ToM Theory for analyzing or describing autism and that this concern has to do with the clash between the neurodiversity movement and the clinical conceptualization of autism. My aim has been to provide descriptions that facilitate enactivist and phenomenological analyses which use the double empathy problem hypothesis.

References

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. MIT press. https://doi.org/10.7551/mitpress/4635.001.0001.

[73] Heasman et al., "Towards Autistic Flow Theory," 480–81.
[74] Di Paolo, Cuffari, and Jaegher, *Linguistic Bodies*, 266.

Chapman, R., & Bovell, V. (2022). Neurodiversity, advocacy, anti-therapy. In *Handbook of autism and pervasive developmental disorder: Assessment, diagnosis, and treatment* (pp. 1519–1536). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-88538-0.

Cook, R. (2004). Frege's Logic. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Summer 2024. Metaphysics Research Lab, Stanford University, 2024.https://plato.stanford.edu/archives/sum2024/entries/frege-logic/.

Di Paolo, E. A., Cuffari, E. C., & De Jaegher, H. (2018). *Linguistic bodies: The continuity between life and language*. MIT press.

Dumitru, M., & Kroon, F. (2008). What to say when there is nothing to talk about. *Crítica* (México, DF), 40(120), 97–109.

Hutto, D. D. (2025). Enactivism. Internet Encyclopedia of Philosophy. ISSN 2161-0002. Accessed April 13, 2025. https://iep.utm.edu/enactivism/.

Fuchs, T. (2015). Pathologies of intersubjectivity in autism and schizophrenia. *Journal of Consciousness Studies*, 22(1-2), 191–214.

Gallagher, S. (2004). Understanding interpersonal problems in autism: Interaction theory as an alternative to theory of mind. *Philosophy, Psychiatry, & Psychology*, 11(3), 199–217. https://doi.org/10.1353/ppp.2004.0063.

Gelan, V. E. (2015). The Idea of Rigorous Science in Husserl's Phenomenology and Its Relevance for the Other Sciences. In *Proceedings of the International Conference Humanities and Social Sciences Today. Classical and Contemporary Issues* – Philosophy and Other Humanities, edited by Mihai-Dan Chiţoiu Ioan-Alexandru Tofan, 141–56. Pro Universitaria, 2015.

Heasman, B., Williams, G., Charura, D., Hamilton, L. G., Milton, D., & Murray, F. (2024). Towards autistic flow theory: A non-pathologising conceptual approach. *Journal for the Theory of Social Behaviour*, 54(4), 469-497. https://doi.org/10.1111/jtsb.12427.

Husserl, E. (2012). *Analyses concerning passive and active synthesis: Lectures on transcendental logic* (Vol. 9). Springer Science & Business Media.

Husserl, E. (1973). *Experience and Judgement: Investigations in a Genealogy of Logic.* Vol. 4. Routledge and Kegan Paul.

Husserl, E. (2012) *Ideas: General Introduction to Pure Phenomenology*. Translated b W. R. Boyce Gibson. Routledge Classics. London New York: Routledge. https://doi.org/10.4324/9780203120330.

Lynch, C. L. (2019). Invisible Abuse: ABA and the Things Only Autistic People Can See. Neuroclastic (blog). Accessed April 13, 2025 https://neuroclastic.com/invisible-abuse-aba-and-the-things-only-autistic-people-can-see/.

McIntyre, R. (1987). Husserl and Frege. The Journal of Philosophy 84, no. 10 (1987): 528–35.

Merleau-Ponty, M. (2012). Phenomenology of Perception. Translated by Donald A. Landes. Abingdon, Oxon New York: Routledge, https://doi.org/10.4324/9780203720714.

Milton, D. E. (2012). On the ontological status of autism: The 'double empathy problem'. *Disability & society*, 27(6), 883-887. https://doi.org/10.1080/09687599.2012.710008.

Mohanty, J. N. (1977). Husserl and Frege: A new look at their relationship. In *Readings on Edmund Husserl's Logical Investigations* (pp. 22-32). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-010-1055-9_3.

Bonnello, C. (2022). Results and analysis of the autistic not weird 2022 autism survey. Autistic Not Weird. https://autisticnotweird.com/autismsurvey/.

Varela, F. J., Thompson, E., & Rosch, E. (1993). *The embodied mind: Cognitive science and human experience* (14. print.).

Walker, N. (2022*). Neuroqueer heresies: Notes on the neurodiversity paradigm, autistic empowerment, and postnormal possibilities*. Tantor Media, Incorporated.

Zahavi, D. (2017). *Husserl's legacy: Phenomenology, metaphysics, and transcendental philosophy*. Oxford University Press.

Zahavi, D. (2019). *Phenomenology: The Basics*. The Basics. London New York: Routledge, Taylor & Francis Group. 2019.

Revista Română de Filosofie Analitică,
Romanian Journal of Analytic Philosophy
Vol. XVI, Iulie-Decembrie 2022, Nr. 2, p. 143-170
ISSN (ediția electronică): 1843-9969,
ISSN (ediția tipărită): 1844-2218
DOI: 10.62229/rrfaxvi-2/7

ESSENTIALISM AND ENVIRONMENTAL CRISIS: AN INEVITABLE (RE)INTRODUCTION

MICAH THOMAS PIMARO, JR.[1]

***Abstract*:** Recent conversations in environmental studies tilt towards the imperative for local knowledge systems. This knowledge is often held by non-experts and outside formal institutional settings. Lived experiences offer alternative perspectives on environmental crises. The challenge, however, remains: how might alternate knowledge be integrated into broader environmental action conversations? In response, metaphysical coherentism, according to which reality consists of a network of independent elements, where every component is grounded in relation to others, is proposed. Such grounding could accommodate the plurality of perspectives that are inherent in the environmental crisis and address the top-down approach in policy frameworks. Metaphysical coherentism argues that greater clarity is needed in the ontological categories of environmental studies.

Keywords: essentialism, environmental crisis, alternate knowledge, metaphysical coherentism, epistemological pluralism

1. Introduction

The current environmental crisis, as marked by a 90% rise in CO_2 emissions since 1970, and the ambitious need for reducing the current warming to 1.5 degrees Celsius, demands new ways of thinking and

[1] Micah Thomas Pimaro, Jr. is a lecturer in Philosophy at the University of Calabar, Nigeria, and doctoral student within the Faculty of Philosophy at the University of Bucharest.

approaches to collective action (Maurya et al., 2020; Burelli & Pala, 2021). While global climate agreements such as the Paris Agreement emphasize the imperative of collective actions, they are often called for in situations where there is high awareness of a problem but there are no meaningful actions, overlooking the local, non-expert knowledge that is held by communities most affected by climate change. This exclusion increases inequalities and could limit effective solutions. But how can we integrate alternate knowledge or different perspectives into broader environmental action conversations? By appealing to metaphysical coherentism, a framework according to which reality consists of a network of independent elements, where every component is grounded in relation to others, I offer a way to embrace alternate knowledge and promote inclusive action.

Globally, different communities are increasingly being affected by the impact of the triple planetary crisis unequally, either in areas where resources are scarce or where there are poor governance mechanisms or capacity; in all these instances, people do face environmental degradation simultaneously along with systemic inequality, especially pertaining to how these challenges are addressed (Abbass et al., 2022; UNDP, 2023). Even when research has shown that air and water pollution can harm people both physically and mentally, especially in those communities that live in industrial zones that are at the same time economically disadvantaged and politically marginalised (Pratt et al., 2015; Mansalidis et al., 2020; Siddiqua et al., 2022; EEA, 2022), the questions of responsibility, justice and collective actions remained contested (Ranniger, 2020; Aneesh et al., 2020; Carlo and Davide, 2021).

The 1997 Kyoto Protocol attempted to tackle such a challenge by introducing the concept of accountability in cutting down emissions; however, poorer nations are still carrying the burden of environmental degradation as well as the mental and health consequences that come with that (Babiker et al., 2000; Kronlid, 2003; Birkmann et al., 2022; Rentschler and Nadezda, 2023). The implication of such tendencies remains that the effectiveness of the Kyoto Protocol is interrupted by minimal participation and implementation mechanisms, as recorded in the UN Treaty Collection (1998) and reiterated by Barrett (1998). In 2016, the Paris Agreement was adopted with a broader view to ensure that

nations reduce their emissions based on the principle of "common but differentiated responsibilities and respective capabilities" as a means of balancing equity with collective action (Falkner, 2016; Annalisa, 2016; Kennedy and Pauw, 2016).

A growing consciousness and concern arose from such analysis on climate justice. Since climate change and other environmental issues affects us unequally, its burdens and benefits should be fairly shared (Schlosberg and Collins, 2014). This is a top-down solution, strategies being designed by central authorities. Such strategies could lead to the "crisis of paralysis", where people know what is wrong, but existing policies are inadequate and unable to engage people who are most affected by the crisis (Zvobgo et al. 2022). This makes the call for the integration of local knowledge a necessity. Local knowledge simply is the lived experiences of communities or knowledge they have held on to from time immemorial as they interact with their local ecosystems, also including how they have adapted to change and protected their environments; these are in general regarded as non-expert knowledge (IPBES, 2019; Mustonen, 2021; IPCC, 2022; Mustonen et al., 2022). According to the IPCC report as recorded by Portner et al. (2021), the triple planetary crisis encompassing climate change, pollution and biodiversity loss requires new ways of reasoning and inclusive approaches that should reflect the complexity of human societies and nature. This would imply the need for integrating community views in making climate policy relevant, just, equitable and workable (Turner et al., 2022).

Nevertheless, the challenge to this proposal would remain how to identify, legitimize and include laypeople's knowledge or non-expert knowledge and what role that can play in shaping broader environmental action. But since human beliefs are interrelated and could gain meaning contextually, metaphysical coherentism might champion the necessity of not stripping away complexity inherent in the adoption of local knowledge but clarify why diverse views should be embraced to make space for understanding people living through the environmental crisis; a shift that could aid in reimagining environmental governance as well as a shared understanding of collective action (Swiderski, 2024).

2. Dominant Epistemologies in Environmental Crisis

This section sets out to clarify the way our understanding and responses to environmental crises are shaped by dominant epistemologies. By dominant epistemologies, I mean knowledge systems that uphold certain ways of seeing and explaining the crisis that marginalize other people, even if unintendedly. Such approaches might offer vital insights into environmental issues. However, in so far as they are unable to account for, and ignore, the lived experiences of people most affected by these crises, dominant epistemologies could limit the effectiveness of environmental policies in such regions. The global environmental crisis characterized by the triple planetary crisis calls for a comprehensive understanding and response, but the dominant epistemologies, particularly those rooted in positivism, scientism, and neoliberalism, have profoundly influenced how such crises are understood and, by extension, addressed, especially by key international bodies such as the IPCC and the UNFCCC.

Positivism has been central to the methodologies of the IPCC and other scientific and policy bodies. IPCC's reports are based on rigorous empirical research and data-driven models that provide crucial insights into the causes and consequences of environmental degradation. Even though such an approach must have contributed in terms of raising environmental consciousness at international forums like the COP, it has its drawbacks, especially as the reliance on quantitative data often marginalizes qualitative insights such as indigenous knowledge systems, which offer valuable perspectives on environmental management and sustainability. More so, the emphasis on scientific consensus can sometimes obscure the ethical and social dimensions of environmental management issues. For instance, while the IPCC provides projections and scenarios, it does not address the questions of inequalities, justice and collective actions that are before us (Barnett et al., 2008).

The scientist *credo* that scientific and technical expertise can solve all problems has significantly influenced environmental policy making. The IPCC sometimes proposes technocratic solutions commonly found in concerns about renewable energy technologies. While these technological innovations are essential, the science underlying their promotion can lead

to an overreliance on technology at the expense of broader socio-political and ethical reforms. Take, for instance, the Paris Agreement, facilitated by the UNFCCC, which heavily emphasizes technological solutions to meet emissions reduction targets, often without sufficient consideration of the underlying socio-economic systems that drive environmental degradation (Tosun and Peters, 2021).

Neoliberalism, as a dominant epistemology in environmental conservation, advocates for market-driven solutions, particularly within the framework of the UNFCCC and other international environmental pacts. This strategy produced the Kyoto Protocol, "the Clean Development Mechanism", its sister market mechanism, and "carbon trading", which all seek to lower greenhouse gas emissions by providing financial incentives for doing so. However, this kind of neoliberalism has put economic efficiency ahead of environmental justice and social equity. Even though market-based solutions frequently help wealthier countries and businesses, impoverished communities, especially those in the global south, bear the burden of environmental damage (Bond, 2012).

This is why Lohmann (2006) maintained that the commodification of nature through mechanisms like carbon credits can lead to the exploitation of natural resources and ecosystems, thereby undermining long-term sustainability.

From the foregoing, there is a growing recognition of the need for epistemological pluralism, according to which there is not just a single way of understanding and tackling environmental crises but multiple and legitimate means of producing and understanding environmental crises/knowledge. This could help in addressing both the current environmental crisis and the need for integrating diverse knowledge systems, ethical considerations, and socio-political perspectives into environmental governance. Such a position is informed by the fact that international agencies like the IPCC and UNFCCC have so far begun to acknowledge that perspective, as shown by the recent reports that emphasize the importance of equity, justice, and inclusion in climate action (IPCC, 2022). Also, the Global Environmental Outlook reports by the UNEP (2023) have highlighted the need for transformative change that goes beyond technical fixes and market mechanisms, advocating for systemic shifts in governance, economic models, and societal values.

3. The impact of dominant epistemologies on environmental crisis

I argue that positivism's focus on scientific objectivity and data-driven approaches can exclude and devalue knowledge systems that are not easily quantified and unable to fit into dominant paradigms. Furthermore, scientism and technocratic bias extend such tendencies. Promoting the beliefs that science and technical expertise alone can solve environmental problems leads to top-down, expert-driven solutions while sidelining the experiential knowledge of local communities that live with and understand their environments intimately. In the same vein, the neoliberal commodification of nature contributes to sidelining alternate knowledge by constantly framing environmental issues through the lens of market efficiency.

The implications of the positions above are numerous. Dominant epistemologies often provide a narrow and incomplete understanding of environmental issues, leading to a lack of awareness of critical ecological relationships and sustainable practices that have been developed over millennia (Kimmerer, 2013). Dominant epistemologies also contribute to the erosion of cultural and biological diversity by undermining the intricacy of non-expert knowledge of specific ecosystems, with cultural practices and languages, as biodiversity coevolves over time (Maffi, 2001). And dominant epistemologies create barriers to achieving sustainability and environmental justice by sidelining alternate knowledge systems, leading to loss of just and more sustainable environmental governance (Martinez-Alier, 2002).

4. Revisiting frontiers for environmental resilience

To improve community resilience, adaptability, sustainability, and disaster preparedness, all of which are primarily context-based, local knowledge and science must work together (Petzold et al. 2020; Reed *et al.* 2023).

However, attempts to combine local knowledge with Western scientific knowledge are frequently ineffective due to researchers'

inadequate understanding of local knowledge, which implies the need for participatory research in the environmental decision-making process (Parsons *et al.* 2017). Leah et al. (2022) contend that indigenous knowledge may include reciprocity in the process of responding to social concerns because they perceive themselves as a part of a genealogical network of interconnected entities and collectives that owe each other reciprocal duties.

A deeper appreciation and respect for local knowledge is often required in order to accurately record and elaborate on their beliefs and ways of life (Chapman and Schott, 2020). Indigenous conceptions of sustainability and well-being, and their incorporation as guiding principles in research and policy, may offer a more inclusive forum for stimulating discussions on goals and outcomes (Parsons *et al.*, 2017). For instance, witness reports can help track and interpret specific changes and effects, as they may not be accessible through scientific technologies (Redvers *et al.*, 2023). This is a situation where native stakeholders' explanations of events, processes, and rates of change could give crucial hypotheses because they take into account contextual aspects that outside researchers are deliberately avoiding or are unaware of (Mustonen *et al*., 2022). To better integrate such views with global assessments of climate change, it is vital to investigate potential complementarities with indigenous and scientific knowledge systems. Accordingly, we must identify patterns of environmental deterioration in regions with limited instrumental data and offer a thorough picture of the consequences of such a crisis. Local knowledge can supply additional data, data that are lacking from many other environmental solutions. (Naess, 2013; Reyes-Garcial et al., 2024).

Failure to understand the accounts of the most vulnerable in the areas impacted may lead to delays in delivering solutions to the environmental crisis (Val, 2002; Kronlid, 2003; Mallory, 2010). Communities should be empowered to speak for themselves. Outside experts can conduct themselves with ethical and epistemological humility, listening to the residents and offering their knowledge to the communities themselves in order to apply and deploy as they fashion their response on their own terms (Kronlid, 2003; Rigby, 2007; Mallory, 2010). Instead, on top of the impact of climate change, the local population often witnesses

contempt for their traditional socioecological resilience system models (Hosen *et al.*, 2020).

It is clear that local knowledge is significant because it has aided in the implementation of eco-decisions and the detection of regular sequences and sociocultural aspects which relate to the ways of life of the local people. Research has in recent times begun to shift towards investigating how community and individual perceptions of climate change provide important information regarding the behavioural dynamics of individuals responding to environmental crises as well as their ability to adjust to new conditions. This is a useful paradigm that, when applied locally, can be more accurate and dependable than scientific knowledge, since communities depend on it to help them deal with the day-to-day difficulties presented by natural processes and aberrations (Whitmarsh and Lorenzoni, 2010). In such cases, knowledge should be sociopolitical, according to Keller et al. (2022) and Reyes-Garcia et al. (2024), respectively. It is crucial to recognize that science shouldn't be made more neutral but rather address issues with the creation, acceptance, and validation of local knowledge in the environmental crisis and how that interacts with power dynamics.

Adopting an indigenous realism which acknowledges the authenticity and depth of indigenous world views and epistemologies could be helpful despite its oversight on historically marginalised practice (Dan, 2020). Indigenous realism is a systematic technique that can pinpoint historical and modern elements that raise the possibility of an environmental crisis by challenging the idea that colonialism is to blame instead. It could concentrate on Indigenous peoples' complex histories of displacement and the impact of invasive practices on their knowledge and way of life, which are strongly tied to the environment.

5. Alternate knowledge for holistic environmental solutions

This section intends to clarify what alternate knowledge means in an environmental crisis. In order to achieve that, concern regarding understanding how traditional ecological knowledge, also regarded as

alternate knowledge is addressed, especially where such knowledge is often grounded in experiential interactions with the environment as opposed to scientific knowledge derived from controlled experiments and empirical observations. How can such a view be used to foster environmental solutions? This argument makes the case that in order to produce more context-specific insights and practical environmental solutions, we must start taking into account the social, cultural, and political aspects of scientific knowledge generation through partnerships and compromises.

One of the questions that such a claim may raise is whether alternative knowledge is value-neutral. In that regard, attempt is made not to fully dig deep into such debate but to show that since value-neutral concern revolves around the imperative to set aside personal values and beliefs to avoid prejudice and to guarantee that rational conclusions take precedence over mere conjecture, it is maintained that concerns about the value-neutrality of knowledge can lead to the adoption of a standpoint epistemology, according to which we acknowledge the situatedness of knowledge and the need to consider diverse perspectives and interests in scientific inquiry and will help us to comprehend the plausibility of reality and the underlying worldviews about knowledge and action in connection to the climate crisis.

These days, ecological degradation and the poverty of hundreds of millions of people are acknowledged as unavoidable outcomes of progress, and the urgent actions required to prevent the eventual destruction of the conditions necessary for humankind to survive seem like a distant goal (Chu and Karr, 2017). In reality, environmental challenges are always viewed as incidental to more pressing problems, which should not be the case (Gare, 1996). Environmental movements need to be reconstructed to address concepts and ways of thinking that genuinely inspire people to take action and foster these kinds of attitudes. Whether such attitudes are best conceived as theoretical, as oriented toward natural kinds or as fundamentally cultural depends on the perspective and conceptual repertoire one brings to bear upon the analysis undertaken (for discussion, cf. Dumitru, 2004). Understanding the concepts and images ingrained in the daily activities of individuals,

along with those in the main societal institutions and their modes of existence, is still essential to comprehending how they interact with their environment. Environmental holism, for instance, emphasizes the interdependence of all components of an ecosystem, including human societies (Behrens, 2010). If this perspective is adopted, indigenous wisdom and local knowledge derived from customary practices, for instance, can be recognized as essential elements of ecosystem management (Brunner and Urenje, 2012; Mazzocchi, 2020).

Nevertheless, as individuals and communities build lasting relationships with profound insights regarding sustainable lifestyle choices, biodiversity preservation, and ecosystem restoration that enhance local knowledge systems and, by implication, prioritize ecological harmony and land preservation, an alternate knowledge claim may be validated and strengthened on such a basis (Dawson *et al.,* 2021). According to proponents of deep ecology, it may be easier to share information and jointly create solutions that respect ecological integrity and cultural diversity if local communities and other stakeholders are encouraged to form partnerships (Akamani, 2020). That will ensure sustainable management of natural resources with indigenous knowledge systems and practices and help to foster close ties to the community and desire to preserve their customs and unique sociocultural and political as well as economic features from those of governing bodies in power (Ens *et al.,* 2021).

Holism is questioned by our seeming knowledge of unchanging meanings because it emphasizes how interrelated all words are (Tony and Sylvia, 2023). One response I share is that explanations that highlight the interdependence of several components routinely prompt an understanding of a system that goes beyond its surface study (List and Spiekermann, 2013; Monika, 2022).

And critics of deep ecology may argue that the world is more threatened by capitalism and class divisions than by the misanthropic biocentric viewpoint that sees people as a threat to non-human existence. I share the response that it may hurt the poor, underprivileged, and Indigenous peoples to foster an idealized depiction of a pristine nature (Chakraborty, 2015).

Indigenous knowledge systems have so far significantly advanced our understanding of biodiversity and its sustainable use and management in a variety of fields, including impact assessment, traditional medicine and health, rural development and agroforestry, natural disaster response and preparation, and customary marine resource management (IPBES 2013). Ellam (2022) asserted that in the past, indigenous peoples have used their own knowledge and science to coexist and adapt to their environment. As their worldviews have recently put the dominant discourse on sustainable development to the test, indigenous peoples demand that their traditional knowledge be acknowledged and respected, granting them the collective right to manage and use the lands and natural resources that they depend on and protect (Mazzocchi, 2020). This is a significant step forward in the development of nature-based conservation and stewardship projects, which will help Indigenous people see preservation of the environment as an obligation (Vogel et al., 2022). According to a recent study, indigenous people legally or customarily own or manage at least 32% of the world's mappable area, and these territories are in outstanding ecological condition because 55% of them have seen little to no human intrusion (Deen, 2023).

There is also evidence that many indigenous people have strong ties to their environmental locations because they have lived there for many generations (Gladun, 2021), suggesting that they view these places as sacred or having spiritual significance (Redvers, 2023). Additionally, according to the World Wide Fund for Nature (2021), 91% of the areas that indigenous people and local communities safeguard are in good or moderate ecological condition. Researchers interested in novel approaches to the current environmental crisis should be concerned about this evidence, which shows that indigenous people's survival depends on how they use natural resources. Knowing this should encourage researchers to respect cultural sensitivities when collecting data and to acknowledge the importance of specific customs and knowledge of a given people or community (Billan, 2020; Mazzocchi, 2020; Estrada, 2022).

Researchers and policymakers also need to critically examine their own biases and presumptions to avoid applying, for instance, Western

frameworks or different interpretations to indigenous knowledge (Simonds and Christopher, 2013; Gonzalez, 2022). In order to obtain consent, establish reliable relationships, and create coalitions that put the needs and opinions of the community first, researchers should, for example, be mindful of how colonial history and power dynamics may affect the research process and be prepared to interact with indigenous populations in a courteous and cooperative manner (Hart et al., 2016). These approaches are unconventional, as they will lead to a fresh way of seeing, presenting, and applying climate conversation to everyday reality. Indigenous people, who until recently did not have a say in policy issues that have to do with them specifically, now get to see an effort from organisations, institutions, researchers, policymakers, and the global society. They are able and ready to appreciate and contribute their quota to fostering productive and sustainable research that can address their needs and priorities.

6. Grounding alternate knowledge in contextualist epistemology

This section challenges us to rethink our usage and justification of different kinds of knowledge in environmental conversation. Contextualist epistemology cautions that "single" knowledge is not always a fact but is warranted in relation to the specific circumstances in which one finds it. This view creates space to embrace other forms of knowing, such as Indigenous ways of knowing, experience, and the knowledge of particular ecological local contexts, as rational and viable. But critics may argue that this would lead to epistemic relativism, in which all statements are equally valid, so it would be difficult to cope with disagreement inherent in an environmental crisis or construct cooperative means of action. To prevent this, we need reflective criteria where we will constantly evaluate knowledge on the basis of its relevance, coherence, and usefulness in its cultural and ecological environment. Rather than letting one system dominate, we must be working towards dialogue and collaboration, where knowledge is co-created in mutual respect and democratic engagement, with all voices heard and considered.

According to Fred Dretske's (1981) relevant alternative theory, "knowing a true proposition one believes at a time requires being able to rule out relevant alternatives to that proposition at that time," which calls for the application of contextualism, a collection of philosophical perspectives that emphasize the context of an action, utterance, or expression. This claim is supported by the idea that human words, acts, and expressions can only be completely understood in the context of a particular situation. Contextualist viewpoints hold that theoretically controversial concepts such as "meaning of x or knowing about x", "having a reason for x", "being true about x", or 'being right about x" only have meanings that are relevant to a particular circumstance. This assertion can be seen as supported by situational ethics. Context-sensitive expressions "present distinct assertions in relation to various circumstances a word is used" (Dretske, 2000b). Because of this, contextualist epistemology's central claim is that knowledge attributions are situation-sensitive, which means that the truth values associated with the word "know" vary depending on the scenario.

Contextualism entails that we can reject the dominant argument in contexts like casual conversations, especially when there are different requirements to declare oneself knowledgeable about a given topic. That would be the equivalent of arguing that when we assign knowledge to something, the standards by which "knowledge" is attributed or rejected in that situation will depend on what sense the term is used. It is in that regard that I am of the view that, to solve epistemological problems and conundrums, epistemologists blend contextualism with theories regarding the nature of knowing. An example of contextualism would be an evidentialist explanation of knowledge that maintains that the degree of justification varies depending on the situation. Hence the necessity of maintaining that the range of relevant alternatives is contingent upon the conversational context and that one might be a contextualist by endorsing the relevant alternative's account of knowing.

7. Reframing environmental knowledge through metaphysical coherentism

Local knowledge should be used as a new way of thinking to address climate and environmental issues. I will argue that this requires a correct metaphysical approach that considers the problem from a holistic perspective.

To begin with, there is a need to understand and clarify "*de dicto*" and "*de re*" distinctions and how they could both play out in this discourse on alternate knowledge recommendations and claims. Simply put, "*de dicto*" is a mode of predication where the attribution of a property is made with respect to a description or a proposition rather than directly to the object itself. From its Latin origin, it denotes a predication or reference about the content of a statement or proposition rather than the actual object itself (Nelson, 2023). This propensity manifests itself in an environmental crisis as "M believes that P is important", translating into "The IPCC believes that local knowledge is important in delivering climate solutions." Here, we see that the attribution of importance is made with respect to the content of the proposition "P" rather than to any specific action that can lead to the co-creation or delivery of local knowledge. It poses a concern about whether local knowledge is a given.

On the other hand, a "*de re*" statement attributes a property directly to an object itself, independently of any particular description or proposition. Literally, meaning "about the thing", "*de re*" is simply a mode of predication where the property is attributed directly to the object itself rather than through a description or even an ascription. In this sense, it is fair to conclude that "*de r*e" *stricto sensu* is about the essential nature[2] or intrinsic properties of an object itself, irrespective of its usage, as follows: "The IPCC believes that the adoption of local knowledge can

[2] In general, metaphysical essentialism simply refers to entities that have certain inherent properties that define their identity (Robertson, 2008). Metaphysical essentialism is usually analyzed in this format: If we suppose that an object X has the attribute Y, then X must essentially have Y for it to qualify as the object that it actually is. In any world where X exists, Y must inevitably possess X, provided X contains Y in essence (Mackie, 2006). This form of metaphysical essentialism is an objective concept that is both objective and non-specific with regard to context (Mizrahi, 2014).

diversify climate solutions." Here, the IPCC is referring to the essential nature or properties of local knowledge, which can be considered independent and nonspecific, and it shows that there is a potential for more exploration.

Following the explanation above, though both perspectives can contribute to the delivery of a solution, I will look at how the adoption or recommendation of local knowledge can lead us to a sense of shared responsibility and actions; hence, the emphasis will be on "*de re"* rather than "*de dicto*". This is just a statement about people's beliefs and is primarily the manner in which climate policy decisions are taken, as they concentrate on public perceptions and could be seen where policymakers might consider public opinion polls to gauge support for climate change mitigation efforts, or activists might aim to change public perceptions of climate change through education and advocacy. In exploring "*de re*", it is clear that in such circumstances, concerns will be about the actual impacts of either the crisis or the action taken to address the impacts. Here, debates ought to be centered on direct impacts on specific communities or species.

Let us consider this: "Climate change has a greater impact on the global south," or "Climate change is causing severe weather events in the global south."

Here, attribution to particular effects linked to the climate change phenomenon itself is significant because discourse about national or industry-specific responsibility in the face of climate change frequently reassigns blame to other parties based on their direct involvement in the issue. Such statements are mostly aimed at evaluating the nature of local knowledge, especially with regard to comprehending the nature of the knowledge, the function of various knowledge systems, and the process of acquiring and exchanging knowledge in a changing climate. An inquiry into the ontological status of the environmental crisis thus can inform conversations about different forms of knowledge and investigate whether reality is fundamentally unified or pluralistic, as well as how this relates to the diversity of knowledge systems present in the discourse about climate impacts. An approach such as that can prompt reflections on the ontological status of traditional ecological knowledge and its role

in shaping human-environment interactions. For instance, a conversation about cultural ontology, which examines the nature and structure of cultures, can intersect with considerations of integrating local knowledge into climate policy while also exploring issues related to the existence and nature of cultural entities and their relationship to individuals and communities. More so, a conversation on interconnectedness and interdependence can offer insights into the relational dynamics between different forms of knowledge, their implications for addressing climate impacts, and their role in shaping collective understandings of environmental issues.

Metaphysical theories such as process philosophy or relational ontology, which emphasize the interconnected nature of reality, where entities and phenomena are understood in terms of their dynamic interactions and relationships, may prompt reflections on how different knowledge systems can interact and influence each other within complex socio-ecological systems. Based on such views, it is important that we can apply a similar logic to comprehend how scientific information and local knowledge are interconnected. This understanding can guide tactics for integrating different viewpoints and encouraging teamwork, which are essential for climate action.

I should note that I use the label "metaphysical coherentism" broadly. Its starting point is (1) justificatory holism, Quine and Ullian's (2007) "web of belief", whether such beliefs are considered at an individual or community-wide level. Indigenous people go beyond that because beliefs, methods of gaining knowledge, practices and customs, ways of life in interaction with the environment, throughout history, form a tangled net of mutual influences (2) best labeled as "holism of epistemic practices". In order for these practices to be tethered to the environment and the lived experiences of inhabiting that environment, something metaphysical, worldly, has to be expressed by such practices and their holism, and metaphysics close to local and participatory knowledge might claim (3) mutual grounding, that all things in existence metaphysically depend on each other or mutually ground each other. The seeming triviality that only what exists can ground, considered jointly with mutual grounding, seems to naturally lead to (4) a relational or process-based metaphysics.

(1)–(4) differ starkly, and many more coherentist and relationist views could be considered. My emphasis on contextualism and pluralism, however, is geared to anchor such metaphysical debates in the lived experiences of peoples who face climate change at home. For such practical purposes, and for integrating local knowledge with scientific and policy perspectives on climate change, the conceptual differences between (1)–(4) are not crucial.

8. Towards epistemological pluralism in environmental crisis

By adopting different ways of knowing what matters in an environmental crisis, I argue that we can navigate the challenges of dominant epistemologies. The notion of epistemological pluralism is adopted in order to do justice to diverse knowledge approaches to the environmental crisis. This style of analysis can likely lead us to transdisciplinary considerations: where we can easily see the need to value and incorporate alternate knowledge systems and challenge the limitations inherent in dominant frameworks while promoting more holistic, equitable and effective solutions. Implementing this framework in the context of environmental crises can validate various perspectives and help us to pay attention to the interconnectedness of all living things and the environment, providing a relational understanding of ecosystems that differs from the reductionist methods of dominant epistemologies (Whyte, 2017).

This can at the same help us challenge dominant epistemologies by advocating for the inclusion of diverse perspectives that prioritize social justice, community well-being, and ecological balance as against technical or economic fixes alone (Escobar, 2018). In most cases, dominant epistemologies often prioritize objective or value-neutral knowledge, but such approaches, given the current rate of growth for the environmental crisis, may lead to technically feasible but ethically problematic solutions.

Pluralism can assist in dismantling the hegemony of value-neutral approaches and promote solutions that are not only effective but also morally and culturally appropriate (Coulthard, 2014). This is another way

of encouraging inclusion and participation in such a way that traditional top-down decision-making processes can be well navigated while advocating for a more democratic and participatory approach where diverse voices, particularly the non-experts, will be heard (Jasanoff, 2004).

We can drive systemic change by challenging the underlying assumptions and power structures that sustain dominant epistemologies. Contextualism, pluralism, and holism can jointly enable us to question the primacy of scientific rationality, economic efficiency, and technological progress while opening doors to alternative paradigms that prioritize ecological harmony, social equity, and long-term sustainability (Kallis, 2018). Holding onto both approaches could lead us to support degrowth or post-growth economic models that reject the neoliberal emphasis on unending economic growth while also providing us with a fresh perspective that is essential for tackling the complex and interrelated problems of the global environmental crisis in a way that respects various knowledge systems and values.

9. Conclusion

When it comes to environmental crises, it is crucial that scientists, researchers, and politicians listen to the opinions of those who would be most impacted by their choices, particularly marginalized people whose voices are frequently left out of the mainstream conversation. I made the case that alternative, non-traditional, and community-based knowledge should be included as essential epistemic resources for environmental governance.

Mere recognition for local knowledge in environmental governance is not enough; I proposed metaphysical coherentism as a norm that could permit us to bring together different kinds of climate knowledge – scientific, local, and Indigenous – into a web of mutual intelligibility, which is a basis for reframing environmental emergencies beyond technocratic terms.

This is also a response to epistemic injustice in so far as it acknowledges that non-dominant cultures have valuable knowledge to

offer, derived from lived experience. Although such knowers are, for now, excluded from decision-making, they are frequently among the most susceptible to environmental degradation, and hence ought to be involved in both knowledge co-production and governance of environmental conversation.

This text argued that local knowledge could support inclusive, sustainable, and place-based environmental solutions by deconstructing wrong assumptions about the environment. As a result, environmental justice is only a distributive problem if it is also an epistemological one, necessitating pluralism, participation, and extensive philosophical engagement with other ways of knowing.

References

Abbass, K., Qasim, M. Z., Song, H., Murshed, M., Mahmood, H., & Younis, I. (2022). A review of the global climate change impacts, adaptations, and sustainable mitigation measures. *Environmental Science and Pollution Research*, 29(28), 42539–42559. https://doi.org/10.1007/s11356-022-19718-6

Akamani, K. (2020). Integrating deep ecology and adaptive governance for sustainable development: Implications for protected areas management. *Sustainability*, 12(14), Article 5757. https://doi.org/10.3390/su12145757

Aneesh, P., Berfond, J., Bhargava, A., Bhatt, A., Chen, J., & Gupta, S. (2020). *Racial disparities and climate change*. Princeton University, PSCI. https://psci.princeton.edu/tips/2020/8/15/racial-disparities-and-climate-change.

Annalisa, S. (2016). The Paris Agreement: A new beginning? *Journal of Energy & Natural Resources Law*, 34(1), 16–26. https://doi.org/10.1080/02646811.2016.1133983

Babiker, M., Vinson, T., & Criqui, P. (2000). The Kyoto Protocol and developing countries. *Energy Policy*, 28(8), 525–536. https://doi.org/10.1016/S0301-4215 (00)00033-1

Barnett, J., Lambert, S., & Fry, I. (2008). *Global environmental change and human security*. In H. G. Brauch, J. Grin, C. Mesjasz, P. Dunay, N. C. Behera, B. Chourou, P. Kameri-Mbote, P. H. Liotta, & J. Oswald Spring (Eds.), Globalization and environmental challenges (pp. 315–324). Springer. https://doi.org/10.1007/978-3-540-75977-5_24

Barrett, S. (1998). *Political economy of the Kyoto Protocol*. Oxford Review of Economic Policy, 14(4), 20–39. https://doi.org/10.1093/oxrep/14.4.20

Behrens, K. (2010). *Exploring African holism with respect to the environment. Environmental Values*, 19(4), 465–484. https://doi.org/10.3197/096327110X531552

Billan, J. (2020). Ethical research engagement with Indigenous communities. *Journal of Rehabilitation and Assistive Technologies Engineering*, 7.

Birkmann, J., Liwenga, E., Pandey, R., Boyd, E., Djalante, R., Gemenne, F., Leal Filho, W., Pinho, P. F., Stringer, L., & Wrathall, D. (2022). "Poverty, livelihoods, and sustainable development." In H.-O. Portner, D. C. Roberts, M. Tignor, E. S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, & B. Rama (Eds.), *Climate change 2022: Impacts, adaptation, and vulnerability* (pp. 1171–1278). Cambridge University Press. https://doi.org/10.1017/9781009325844.010

Bond, P. (2012). *Politics of climate justice: Paralysis above, movement below*. University of KwaZulu-Natal Press.

Burelli, C., & Pala, D. (2021). The environmental crisis and its injustice: An inevitably short introduction. *Rivista di Estetica*, 75, 3–16. https://doi.org/10.4000/estetica.7190

Chakraborty, R. (2015). The deep ecology/ecofeminism debate: An enquiry into environmental ethics. *Journal of the Indian Council of Philosophical Research*, 32(1), 123–134. https://doi.org/10.1007/s40961-015-0005-y

Chapman, J. M., & Schott, S. (2020). Knowledge coevolution: Generating new understanding through bridging and strengthening distinct knowledge systems and empowering local knowledge holders. *Sustainability Science*, 15(3), 931–943. https://doi.org/10.1007/s11625-020-00781-2

Chu, E. W., & Karr, J. R. (2017). Environmental impact: Concept, consequences, and measurement. In *Reference module in life sciences*. Elsevier. https://doi.org/10.1016/B978-0-12-809633-8.02380-3

Coulthard, G. S. (2014). *Red skin, white masks: Rejecting the colonial politics of recognition*. University of Minnesota Press. https://doi.org/10.5749/j.ctt9qh3cv

Dawson, N., Coolsaet, B., Sterling, E., Loveridge, R., Gross-Camp, N., Wongbusarakum, S., Sangha, K. K., Scherl, L. M., Phan, H. P., Zafra-Calvo, N., Lavey, W. G., Byakagaba, P., Idrobo, C. J., Chenet, A., Bennett, N. J., Mansourian, S., & Rosado-May, F. J. (2021). The role of Indigenous peoples and local communities in effective and equitable conservation. *Ecology and Society*, 26(3), Article 19. https://doi.org/10.5751/ES-12625-260319

Deen, S. (2023). Embedding Indigenous knowledge in the conservation and restoration of landscapes (Insight Report). *World Economic Forum*. https://www3.weforum.org/docs/WEF_Embedding_Indigenous_Knowledge_2023.pdf

Dretske, F. I. (1981). *Knowledge and the flow of information*. MIT Press.

Dretske, F. I. (2000b). *The pragmatic dimension of knowledge. In Perception, knowledge, and belief: Selected essays* (pp. 48–63). Cambridge University Press. https://doi.org/10.1017/CBO9780511625299.004

Dumitru, M. (2004). *Explorări logico-filozofice*. Bucureşti: Humanitas.

Ellam, Y., Raymond-Yakoubian, J., Daniel, R., & Behe, C. (2022). A framework for co-production of knowledge in the context of Arctic research. *Ecology and Society*, 27(1), Article 34. https://doi.org/10.5751/ES-12960-270134

Ens, E., Faulks, M., Finlayson, C. M., & Marrfurra McTaggart, P. (2021). Recognition of Indigenous ecological knowledge systems in conservation and their role to narrow the knowledge-implementation gap. In C. C. Ferreira & C. F. C. Klutsch (Eds.). *Closing the knowledge-implementation gap in conservation science* (pp. 109–139). Springer. https://doi.org/10.1007/978-3-030-81085-6_5

Escobar, A. (2018). *Designs for the pluriverse: Radical interdependence, autonomy, and the making of worlds*. Duke University Press. https://doi.org/10.1215/9780822371816

Estrada, A., Garber, P. A., Gouveia, S., Fernandez-Llamazares, Á., Ascensao, F., Fuentes, A., Katti, M., Shaffer, L. J., Spehar, S., Strier, K. B., Vollmer, W., & Wittemyer, G. (2022). Global importance of Indigenous peoples, their lands, and knowledge systems for saving the world's primates from extinction. *Science Advances*, 8(31), Article eabn2927. https://doi.org/10.1126/sciadv.abn2927

European Environment Agency. (2022). *Soil pollution and health*. https://www.eea.europa.eu/publications/zero-pollution/health/soil-pollution

Falkner, R. (2016). The Paris Agreement and the new logic of international climate politics. *International Affairs*, 92(5), 1107–1125. https://doi.org/10.1111/1468-2346.12708

Gare, A. (1996). *Environmental destruction and the metaphysics of sustainability*. Eco-Logical Press.

Gladun, E. (2021). Indigenous economies in the Arctic: To thrive or to survive? *Elementa: Science of the Anthropocene*, 9(1), Article 00088. https://doi.org/10.1525/elementa.2019.00088

Gonzalez, R., Kretz, L., & Martinez, R. (2022). Psychology and Indigenous people. *Annual Review of Psychology*, 73, 431–459. https://doi.org/10.1146/annurev-psych-012921-045304

Hart, M. A., Straka, S., & Rowe, G. (2017). Working across contexts: Practical considerations of doing Indigenist/anti-colonial research. *Qualitative Inquiry*, 23(5), 332–342. https://doi.org/10.1177/1077800416659084

Hosen, N., Nakamura, H., & Hamzah, A. (2020). Adaptation to climate change: Does traditional ecological knowledge hold the key? *Sustainability*, 12(2), Article 676. https://doi.org/10.3390/su12020676

Intergovernmental Panel on Climate Change. (2001). *Climate change 2001: Impacts, adaptation, and vulnerability* (J. J. McCarthy, O. F. Canziani, N. A. Leary, D. J. Dokken, & K. S. White, Eds.). Cambridge University Press.

Intergovernmental Panel on Climate Change. (2022). *Summary for policymakers*. In H.-O. Pörtner, D. C. Roberts, M. Tignor, E. S. Poloczanska, K. Mintenbeck, A. Alegría, M.

Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, & B. Rama (Eds.), *Climate change 2022: Impacts, adaptation, and vulnerability* (pp. 3–33). Cambridge University Press. https://doi.org/10.1017/9781009325844.001

Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. (2019). *Summary for policymakers of the global assessment report on biodiversity and ecosystem services* (S. Díaz, J. Settele, E. S. Brondízio, H. T. Ngo, M. Guèze, J. Agard, A. Arneth, P. Balvanera, K. A. Brauman, S. H. M. Butchart, K. M. A. Chan, L. A. Garibaldi, K. Ichii, J. Liu, S. M. Subramanian, G. F. Midgley, P. Miloslavich, Z. Molnár, D. Obura, A. Pfaff, S. Polasky, A. Purvis, J. Razzaque, B. Reyers, R. Roy Chowdhury, Y. J. Shin, I. J. Visseren-Hamakers, K. J. Willis, & C. N. Zayas, Eds.). IPBES Secretariat. https://doi.org/10.5281/zenodo.355357Jasanoff, S. (Ed.). (2004). States of knowledge: The co-production of science and the social order. Routledge. https://doi.org/10.4324/9780203413845

Kallis, G. (2011). In defence of degrowth. *Ecological Economics*, 70(5), 873–880. https://doi.org/10.1016/j.ecolecon.2010.12.007

Keller, E., Norman, A., & Ebi, K. L. (2022). A systematic review of the psychological distance of climate change: Towards the development of an evidence-based construct. *Journal of Environmental Psychology*, 81, Article 101822. https://doi.org/10.1016/j.jenvp.2022.101822

Kennedy, L. M., & Pauw, P. (2016). *Self-differentiation of countries' responsibilities: Addressing climate change through intended nationally determined contribution*s (Report No. 4/2016). German Development Institute. https://doi.org/10.13140/RG.2.1.3087.0162

Kimmerer, R. W. (2013). *Braiding sweetgrass: Indigenous wisdom, scientific knowledge, and the teachings of plants*. Milkweed Editions.

Kronlid, D. (2003). *Ecofeminism and environmental ethics: An analysis of ecofeminist ethical theory* (Uppsala Studies in Social Ethics, 28). Uppsala University. ISBN 91-554-5516-6

List, C., & Spiekermann, K. (2013). Methodological individualism and holism in political science: A reconciliation. *American Political Science Review*, 107(4), 629–643. https://doi.org/10.1017/S0003055413000373

Lohmann, L. (2006). Carbon trading: A critical conversation on climate change, privatisation and power. *Development Dialogue*, 48, 1–359.

Mackie, P. (2006). *How things might have been: Individuals, kinds, and essential properties*. Oxford University Press. https://doi.org/10.1093/0199272204.001.0001

Maffi, L. (Ed.). (2001). *On biocultural diversity: Linking language, knowledge, and the environment. Smithsonian Institution Press*. ISBN 1-56098-930-0

Mallory, C. (2010). Who knows? Ethics and epistemologies of climate change and environmental justice (Conference paper). *Climate, Ethics, and Energy PanelDiscussion*.https://www.academia.edu/32045105/Who_Knows_Ethics_and_Epistemologies_of_Climate_Change_and_Environmental

Maurya, K. P., Kumar, A., & Kumar, P. (2020). An introduction to environmental degradation: Causes, consequences, and mitigation. In Environmental degradation: Causes and remediation strategies (pp. 1–20). *Agriculture and Environmental Science Academy*. https://doi.org/10.26832/aesa-2020-edcrs-01

Mazzocchi, F. (2020). A deeper meaning of sustainability: Insights from Indigenous knowledge. *The Anthropocene Review*, 7(1), 77–93. https://doi.org/10.1177/2053019619898888

Mizrahi, M. (2014). Essentialism: Metaphysical or psychological? *Croatian Journal of Philosophy*, 14(1), 65–72.

Monika, D. S. (2022). Climate change, holism, evolution, and collective intelligence. *Sustainability and Climate Change*, 15(5), 387–395. https://doi.org/10.1089/scc.2022.0047

Mustonen, T., Feodoroff, P., & Mustonen, K. (Eds.). (2021). *Compendium of Indigenous knowledge and local knowledge: Towards inclusion of Indigenous knowledge and local knowledge in global reports on climate change*. Snowchange Cooperative. https://doi.org/10.13140/RG.2.2.14498.76485

Mustonen, T., Huusari, L., Feodoroff, P., & Mustonen, K. (2022). The role of Indigenous knowledge and local knowledge in understanding and adapting to climate change. In *Climate resilient development pathways* (pp. 2713–2728). Research Gate. https://www.researchgate.net/publication/362432216

Naess, L. O. (2013). The role of local knowledge in adaptation to climate change. *Wiley Interdisciplinary Reviews: Climate Change*, 4(2), 99–106. https://doi.org/10.1002/wcc.204

Nelson, M. (2023). Propositional attitude reports. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Spring 2023 ed.). Stanford University.https://plato.stanford.edu/archives/spr2023/entries/prop-attitude-reports/

Parsons, M., Nalau, J., & Fisher, K. (2017). Alternative perspectives on sustainability: Indigenous knowledge and methodologies. *Challenges in Sustainability*, 5(1), 7–14. https://doi.org/10.12924/cis2017.05010007

Petzold, J., Andrews, N., Ford, J. D., Hedemann, C., & Postigo, J. C. (2020). Indigenous knowledge on climate change adaptation: A global evidence map of academic literature. *Environmental Research Letters*, 15(11), Article 113007. https://doi.org/10.1088/1748-9326/abb330

Pörtner, H.-O., Scholes, R. J., Agard, J., Archer, E., Arneth, A., Bai, X., Barnes, D., Burrows, M., Chan, L., Cheung, W. W. L., Diamond, S., Donatti, C., Duarte, C., Eisenhauer, N., Foden, W., Gasalla, M. A., Handa, C., Hickler, T., Hoegh-Guldberg, O., … Ngo, H. T. (2021). IPBES-IPCC co-sponsored workshop report on biodiversity and climate change. *IPBES and IPCC*. https://doi.org/10.5281/zenodo.5101133

Pratt, G. C., Vadali, M. L., Kvale, D. L., & McKinley, E. L. (2015). Traffic, air pollution, minority, and socio-economic status: Addressing inequities in exposure and risk. *International Journal of Environmental Research and Public Health*, 12(5), 5355–5372. https://doi.org/10.3390/ijerph120505355

Quine, W. V., and J. S. Ullian. 2007. *Ţesătura Opiniilor. (*Romanian edition of *Web of Belief)*. Translated by Mircea Dumitru. Pitești, Romania: Paralela 45.

Ranniger, G. (2020, July 7). Environmental injustice (News article). *Environmental Health News*.https://www.ehn.org/environmental-injustice-2646393134.html

Redvers, N., Yellow Bird, M., Quinn, D., Yunker, P., & Belcourt, B. (2023). Indigenous peoples: Traditional knowledge, climate change, and health. *PLOS Global Public Health*, 3(10), Article e0002474. https://doi.org/10.1371/journal.pgph.0002474

Reed, M. G., Godmaire, H., Abernethy, P., Guertin, M.-A., Hewitt, N., & Fortier, J.-F. (2023). *Guiding principles for transdisciplinary sustainability research and practice. People and Nature*, 5(4), 1094–1109. https://doi.org/10.1002/pan3.10496

Rentschler, J., & Leonova, N. (2023). *Global air pollution exposure and poverty. Nature Communications*, 14(1), Article 4637. https://doi.org/10.1038/s41467-023-39797-4

Reyes-García, V., Fernández-Llamazares, A., García-del-Amo, D., Matias, D. M., Benyei, P., Junqueira, A. B., & Schunko, C. (2024). Local studies provide a global perspective on the impacts of climate change on Indigenous peoples and local communities. *Sustainable Earth Reviews*, 7(1), Article 1. https://doi.org/10.1186/s42055-023-00063-6

Rigby, K. (2007). Tragedy, modernity, and terra mater: Christa Wolf recounts the fall. *New German Critique*, 34(2), 115–141 bundles. https://doi.org/10.1215/0094033X-2007-005

Robertson, T. (2008). Essential vs. accidental properties. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2008 ed.). Stanford University. https://plato.stanford.edu/archives/fall2008/entries/essential-accidental/

Schlosberg, D., & Collins, L. B. (2014). From environmental to climate justice: Climate change and the discourse of environmental justice. *Wiley Interdisciplinary Reviews: Climate Change*, 5(3), 359–374. https://doi.org/10.1002/wcc.275

Siddiqua, A., Hahladakis, J. N., & Al-Attiya, W. A. K. (2022). An overview of the environmental pollution and health effects associated with waste landfilling and open dumping. *Environmental Science and Pollution Research*, 29(39), 58514–58536. https://doi.org/10.1007/s11356-022-21578-z

Simonds, V. W., & Christopher, S. (2013). Adapting Western research methods to Indigenous ways of knowing. *American Journal of*

Public Health, 103(12), 2185–2192. https://doi.org/10.2105/AJPH.2012.301157

Swiderski, J. (2024). Varieties of metaphysical coherentism. *Erkenntnis*, 89(5), 1861–1886. https://doi.org/10.1007/s10670-022-00607-w

Tony, L., & Vitello, S. (2023). What is a holistic judgment, anyway? Research Papers in Education. *Advance online publication*. https://doi.org/10.1080/02671522.2023.2269960

Tosun, J., & Peters, B. G. (2021). The politics of climate change: Domestic and international responses to a global challenge. *International Political Science Review*, 42(1), 3–15. https://doi.org/10.1177/0192512120975659

Turner, N. J., Boelscher Ignace, M., Marshall, M. B., Bhattacharyya, J., & Cocksedge, W. (2022). Well grounded: Indigenous peoples' knowledge, ethnobiology, and sustainability. *People and Nature*, 4(3), 627–651. https://doi.org/10.1002/pan3.10321

United Nations. (1998). *Kyoto Protocol to the United Nations Framework Convention on Climate Change, United Nations Treaty Series*, 2303, 162. https://treaties.un.org/doc/Treaties/1998/09/19980921%2006-41%20PM/Ch_XXVII_07_ap.pdf

United Nations Development Programme. (2023). *Climate change, environmental degradation, conflict, and displacement in the Arab States region (Policy brief)*. UNDP.https://www.undp.org/sites/g/files/zskgke326/files/202306/climate_environmental_security_rbas_policy_brief_may_2023.pdf

United Nations Environment Programme. (2023). *Emissions gap report 2023: Broken record – Temperatures hit new highs, yet world fails to cut emissions (again)*. UNEP. https://doi.org/10.59117/20.500.11822/43922

UNESCO. (2018). *Indigenous knowledge is needed for climate change assessment and adaptation* (D. Nakashima, I. Krupnik, & J. T. Rubis, Eds.). Cambridge University Press.

Val, P. (2002). *Environmental culture: The ecological crisis of reason*. Routledge. ISBN 9780415178785

Vogel, B., Sangha, K. K., & Low, N. (2022). Indigenous-led nature-based solutions for the climate crisis: Insights from Canada. *Sustainability*, 14(11), Article 6725. https://doi.org/10.3390/su14116725

Whyte, K. P. (2017). Indigenous climate change studies: Indigenizing futures, decolonizing the Anthropocene. *English Language Notes*, 55(1–2), 153–162. https://doi.org/10.1215/00138282-55.1-2.153

World Wide Fund for Nature. (2021). *The state of Indigenous peoples and local communities' lands and territories in Switzerland* [Report]. WWF. https://wwflac.awsassets.panda.org/downloads/report_the_state_of_the_indigenous_peoples_and_local_communities_lands_and_territories_1.pdf

Zvobgo, L., Totin, E., Simpson, N., van der Merwe, L., Vogel, C., & Leal Filho, W. (2022). *The role of Indigenous knowledge and local knowledge in water sector adaptation to climate change in Africa: A structured assessment. Sustainability Science*, 17(5), 2077–2092. https://doi.org/10.1007/s11625-022-01118-x

Revista Română de Filosofie Analitică,
Romanian Journal of Analytic Philosophy
Vol. XVI, Iulie-Decembrie 2022, Nr. 2, p. 171-178
ISSN (ediția electronică): 1843-9969,
ISSN (ediția tipărită): 1844-2218
DOI: 10.62229/rrfaxvi-2/8

MIRCEA FLORIAN ȘI STRUCTURA RECESIVITĂȚII. RECESIVITATEA CA STRUCTURĂ A LIMBAJULUI

MIRCEA TOBOȘARU[1]

Rezumat: Unele cărți conțin, undeva din adâncul lor, fraze cu implicații existențiale care nu doar uimesc orice gândire, ci o obligă să înțeleagă sensul lor. Mircea Florian, în "Recesivitatea ca structură a lumii" (1983; de aici înainte voi folosi prescurtarea RSL) scrie: "Recesivitatea este factorul tulburător, neliniștea lumii, asimetria cosmosului. În această structură complexă, anume recesivitatea, rezidă tragicul lumii." RSL este o carte despre noțiunile prime ale gândirii noastre, despre conceptele-perechi prezente în arhitectura noastră cognitivă. Aceste noțiuni sau categorii sunt noțiuni aflate într-un raport de opoziție, de contrarietate și, mai precis, de recesivitate: noțiuni pozitive opuse, dintre care una are caracter dominant, iar cealaltă, cea care vine după, deci recesivă, are o semnificație superioară. Fiind un studiu despre noțiuni, RSL ar putea fi interpretată ca fiind o carte despre limbaj. Limbajul însă, în viziunea lui Mircea Florian, reflectă structura gândirii, iar gândirea reflectă structura lumii. Ontologia are prioritate în fața mentalului și a semanticii. Am putea spune că este un tratat metafizic clasic despre categorii. Mircea Florian nu vorbește însă de categorii deoarece consideră că acest termen este „contaminat de subiectivism, pe linie kantiană" (RSL, Vol. I, 83). Dar atunci despre ce fel de proiect metafizic este vorba în RSL? Ce tip de discurs filosofic ne este propus și care, în principiu, ne permite să înțelegem tragicul și neliniștea lumii? Scopul principal al analizei mele va fi, în primul rând, să răspund la această întrebare. În al doilea rând, voi încerca să răspund la o altă întrebare, anume dacă nu cumva recesivitatea este o structură nu a lumii (deci ontologică), ci mai degrabă a *unor* limbaje (și, ca atare, conceptuală).

[1] Mircea Toboşaru este cadru didactic în cadrul Universităţii Naţionale de Ştiinţă şi Tehnologie Politehnica București.

Cuvinte cheie: Mircea Florian, recesivitate, pluralism metafizic, incompatibilitate, Robert Brandom.

1. Un proiect metafizic modest și pluralist?

"Recesivitatea ca structură a lumii" este un proiect metafizic sistematic, însă nu în sensul tradițional în care, într-o interpretare lingvistică, se urmărește descrierea fidelă și explicarea realității ultime în termenii unei structuri conceptuale cu o putere expresivă maximă (astfel încât orice alt vocabular să poată fi redus la vocabularul metafizic printr-o explicație ce face apel la anumite principii prime). Ce este important de sesizat în acest context este caracterul procustian al metafizicii tradiționale: ceea ce nu poate fi descris și explicat folosind limbajul metafizic are un caracter deficitar. Metafizica, în acest sens, este un mod de gândire ce poate avea accente diferite. Filosofia antică, de exemplu, dă prioritate ontologiei. Filosofia modernă are valențe epistemologice, iar cea modernă, cel puțin într-o anumită tradiție, este axată pe analiza limbajului. Cu toate acestea, proiectul metafizic conturat mai sus se poate recunoaște în toate aceste diverse perioade, la diverși autori, chiar și la cei care se declară în mod vehement anti-metafizici (de exemplu la autorii asociați empirismului logic). În măsura în care cineva folosește exclusiv limbajul științelor naturii pentru a descrie și explica, se poate spune că adoptă un mod de gândire metafizic în măsura în care are tendințe procustiene: limbajul etic, modal sau estetic vor risca să fie privite ca problematice în măsura în care nu pot fi reduse la limbajul naturalist. Ca proiect sistematic modest, însă, metafizica poate fi văzută ca o încercare de a formula un limbaj cu o putere expresivă universală, care nu are tendințe totalitare. Adoptarea pluralismului conceptual este echivalentă cu respingerea unei presupoziții esențiale pentru metafizica tradițională deoarece implică faptul că pot exista moduri diferite de descriere a realității.[2]

[2] Acest mod de a caracteriza proiectele metafizice clasice aparține lui Robert Brandom (2008, vezi cap. I și Postfața).

În ce categorie se poate integra proiectul metafizic formulat în RSL? Mircea Florian se angajează, aparent, față de un proiect metafizic pluralist. Iată, în primul rând, cum descrie metafizica tradițională: ”Metafizica tradițională a trăit din absolutizarea noțiunilor, în speranța că ea va descoperi noțiunea din care vor decurge, printr-o formulă magică, toate celelalte noțiuni și realități” (RSL, Vol. I, 46). În al doilea rând, viziunea sa despre filozofie este aceea a unui domeniu fundamental, ce ”are misiunea de a dezvălui, de a ierarhiza, de a descrie – pe scurt – de a clarifica noțiunile prime, care sunt cele mai generale sub raportul abstracției și cele mai concrete prin universala lor actualitate” (RSL, Vol. I, 45). Aceste ”noțiuni limită, fiindcă își au sediul la marginea gândirii”, au o structură polară caracterizată printr-un raport de contrarietate specific, anume acela de recesivitate. Câteva dintre aceste perechi conceptuale sunt următoarele: deosebire-asemănare, multiplu-unu, noutate-repetiție, individual-general, materie-formă, corp-suflet, materie-formă, real-posibil, obiect-subiect etc. Ceea ce este important în acest context este faptul că analiza lui Mircea Florian duce, printre altele, la următoarele concluzii specifice:

> 1. Aceste perechi conceptuale sunt perechi de noțiuni contradictorii, iar ceea ce le caracterizează este un raport de recesivitate, astfel încât legătura dintre noțiuni este necesară și unul dintre termeni ”domină pe celălalt, care totuși păstrează o semnificație superioară” (RSL, Vol. I, 56).
> 2. Lista termenilor aflați în raport de recesivitate (minim 40, după Mircea Florian) nu poate fi completă.

Acest ultim punct, precum și susținerea sa că ”[n]u există sistem filosofic definitiv” pledează pentru caracterizarea proiectului său ca un proiect metafizic sistematic pluralist modest. Sistemele filosofice se bazează pe una sau mai multe perechi de concepte care nu sunt isostenice, nu au o putere egală. Aflăm aici posibilitatea pluralismului filosofic: pot exista mai multe sisteme filosofice coerente. Astfel,

> "stă în natura fiecărui sistem să fie necomplet, deschis, nesaturat tocmai fiindcă este un sistem determinat, cu un anumit punct de plecare axiomatic, cu un anumit nivel al problemelor. De aceea un sistem generează unul sau mai mult sisteme sau prin antinomie sau prin antistază (varietate în cadrul aceluiași sistem). Nu există o afirmație filosofică în afară de structura și spiritul unui sistem, toate problemele sunt puse și rezolvate la nivelul unui sistem. De aceea filosofia va continua de a construi sisteme, căci succederea sistemelor nu compromite filosofia, ci dimpotrivă o justifică." (RSL, Vol. I, 63)

În măsura în care avem de-a face în acest pasaj cu judecăți universale, Florian descrie și propriul său proiect filosofic. Ca atare, în paginile *Recesivității* găsim un sistem filosofic cu un punct de plecare axiomatic. Sistemul nu are însă pretenția completitudinii, în sensul că recesivitatea nu este o structură universală. De asemenea, este respinsă și tendința totalitară sau absolutistă: sistemul său nu este decât o etapă dintr-un proces filosofic ce are propria sa dinamică. Această dinamică nu poate fi a priori determinată deoarece "nu există un echilibru obiectiv al adevărurilor eventuale cuprinse în ele" (RSL, Vol. I, 63).

Dar care este exact sistemul filosofic al recesivității? Constă el exclusiv în analiza unor perechi de noțiuni fundamentale și în evidențierea relației de recesivitate dintre ele? Cred că orizontul filosofic al acestui proiect cuprinde, chiar dacă nu mereu explicit, următoarele idei-cheie. În primul rând, analiza unui meta-concept, cel al recesivității, se poate face extensional, însă și intensional, prin caracterizarea principalului pol conceptual opus recesivității care este, conform lui Mircea Florian, isostenicitatea. Putem spune că perechea recesivitate-isostenicitate este una meta-conceptuală, deoarece se aplică la alte perechi de concepte de ordinul întâi (viață-moarte; pesimism-optimism; individ-societate etc.) sau de ordinul al doilea, i.e. concepte care se aplică la alte concepte (individual-general, relativ-absolut etc.). Acestea reprezintă extensiunea conceptului de recesivitate. În plus, recesivitatea poate fi caracterizată, conform lui Mircea Florian, prin alte perechi ce termeni secundari, precum perechea determinat-nedeterminat. Recesivitatea nu poate fi redusă la determinare-nedeterminare deoarece mai are și alte specificații. Un termen determină sau domină, altul este determinat. Însă se insistă și pe semnificația specială a termenului determinat, recesiv.

În repetatele parafrazări referitoare la recesivitate întâlnim, de asemenea, și relația de constituire. Termenul recesiv este constitutiv, adică oferă identitatea primului termen (de exemplu: în perechea individual-general). Semnificația superioară a termenului recesiv poate fi înțeleasă și perspectivist, fapt sugerat de următoarea explicație:

> "Uneori ne-am gândit să exprimăm caracterul specific al recesivității printr-o comparație care acordă fiecărui termen un fel de întâietate alternativă, fără a le considera prin aceasta echivalente. Este o întâietate oarecum pe rând și din puncte de vedere deosebite […]." (RSL, Vol. I, 73)

Mircea Florian evită să ierarhizeze perechile de noțiuni aflate în raport de recesivitate. Toate, într-un anumit sens, sunt pe planul orizontal. Legăturile dintre ele nu sunt asemenea legăturilor dintr-un model piramidal unde unele noțiuni pot fi reduse la alte noțiuni. Cu toate acestea, există raporturi de dependență, iar acestea sunt explorate parțial în RSL, chiar dacă nu (din păcate) sistematic. Închei aceste observații privitoare la structura generală a proiectului filosofic al lui Mircea Florian cu indicarea unui orizont de probleme: i) În ce măsură perechea recesivitate-isostenicitate este reflexivă ? Mai precis: care dintre acești termeni este recesiv și care domină? ii) De ce nu poate fi dedusă o listă completă a noțiunilor recesive ?

2. Recesivitatea, între ontologie și semantică

În continuare voi discuta o posibilă obiecție ce poate fi adusă proiectului teoretizării recesivității. Mircea Florian adoptă o ordine a explicației privind natura recesivității ce pleacă de la ontologie: structura lumii este duală, formată din elemente necesar opuse; limbajul reflectă realitatea și, în consecință, studiul limbajului ne va spune ceva despre structura realității. Teza aceasta deschide următorul orizont de problematizare: poate că există structuri conceptuale diferite ce descriu lumea în moduri diferite. Dacă ceea ce numim realitate este o funcție a cadrului lingvistic pe care îl adoptăm (în sensul lui Carnap (1950)), atunci cum mai putem

vorbi de recesivitate ca structură a lumii, dat fiind faptul că recesivitatea este o structură conceptuală și că ar putea să existe, teoretic, cadre lingvistice ce nu conțin noțiuni contrare (concepte aflate în raport de incompatibilitate)?

Nu este de la sine înțeles că diverse limbaje sau scheme conceptuale descriu în același fel lumea. Uneori în știință apar teorii incompatibile care dau seama de toate datele experienței, postulând însă structuri diferite ale lumii. Această idee a subdeterminării empirice implică o teză metodologică: semantica (teoria) are prioritate, cel puțin în anumite cazuri, în fața ontologiei. Dacă acceptăm această idee, atunci demersul din RSL devine problematic deoarece ne putem închipui un cadru lingvistic în care nu există perechi de noțiuni aflate în raport de recesivitate. În acest caz nu mai putem accepta ideea lui Mircea Florian, anume că recesivitatea este o structură *a lumii*. Limbaje conceptual diferite ne obligă, uneori, să postulăm seturi de fapte și structuri ontologice diferite. Este însă coerentă ideea existenței unei structuri conceptuale care să nu conțină perechi de termeni necesar opuși, aflați în raport de recesivitate, sau este incompatibilitatea conceptuală o structură universală necesară a oricărui limbaj? Dacă recesivitatea este o structură conceptuală universală, prezentă necesar în orice limbaj posibil, atunci recesivitatea poate fi descrisă ca o structură a lumii și nu doar ca o structură a lumii relativă la un anumit limbaj, cel preferat de către Mircea Florian. Dacă incompatibilitatea este necesară unui limbaj ne putem întreba, firește, dacă nu cumva şi cuplul necesitate-(im)posibilitate nu este unul dintre cele vizate de Florian. Pentru discuţia caracterului fundamental al conceptelor modale, cf. Mircea Dumitru (2004, p. 244).

Un mod de a aborda problema în discuție este să ne întrebăm ce face un limbaj să fie limbaj, adică să aibă expresii cu un anumit conținut semantic. Într-un mod plastic: ce dă semnificație enunțului "Pata este roșie"? Nu atât faptul că ne imaginăm ceva, cât faptul că noi știm să navigăm pe o anumită "rețea conceptuală".[3] Știm că dacă ceva este roșu,

[3] Semnificația unei expresii nu poate fi echivalată cu ceea ce ne închipuim atunci când înțelegem expresia. În unele cazuri, precum al expresiei "figură geometrică cu un milion de laturi", nu ne putem reprezenta mental ceva adecvat.

atunci în mod necesar este colorat. Iar dacă ceva este sângeriu, atunci este roșu. Însă trebuie să știm și faptul că, dacă ceva este roșu, atunci nu poate avea în același timp, și sub același raport, și o altă culoare. Ca atare, nu numai condițiile de aplicabilitate și consecințele aplicării conceptului contează, ci și relațiile de incompatibilitate conceptuală dintre un concept și alte concepte. Mircea Florian argumentează în favoarea independenței conceptuale a contrarietății față de contradicție. Se poate merge însă și mai departe și se poate arăta că negația, constitutivă pentru contradicție, poate fi introdusă numai în acele limbaje care conțin deja relații de contrarietate sau de imcompatibilitate. Relația de contradicție poate fi definită ca relația de minimă contrarietate: non-p este negația lui p doar dacă non-p este minimul incompatibil al lui p: ceea ce este implicat de tot ceea ce este incompatibil cu p. Non-roșu este implicat de a fi verde, albastru, etc.; aceste culori sunt incompatibile cu roșu.[4] Ca atare, relația de incompatilititate este necesară pentru structura unui limbaj și, în calitate de contrarietate, are prioritate față de contradicție. Orice limbaj, pentru a avea o dimensiune semantică, trebuie, în consecință, să conțină noțiuni contrare.

Problema este că nu orice noțiuni contrare se găsesc într-un raport de recesivitate. Recesivitatea presupune mai mult decât incompatibilitate, anume noțiuni pozitive opuse, dintre care una are caracter dominant, iar cealaltă, cea care vine după, *are o semnificație superioară*. Deoarece nu cred că se poate formula un argument în favoarea ideii că recesivitatea este un raport semantic universal (prezent în orice limbaj posibil), nu putem decât să o descriem ca fiind o structură conceptuală (nu ontologică) relativă la anumite limbaje. În ciuda unei astfel de concluzii, cred că recesivitatea rămâne totuși o idee originală și fecundă, iar pierderea statului de "structură a lumii" nu știrbește cu nimic valoarea ei pentru noi, cei ce gândim în același limbaj ca și Mircea Florian, asimetria, neliniștea și tragicul lumii.

[4] Acest argument este dezvoltat în Brandom 2008, cap. V.

Bibliografie

Brandom, R. (2008). *Between Saying and Doing: Towards an Analytic pragmatism*. Oxford University Press.

Carnap, R. (1950). Empiricism, Semantics and Ontology, *Revue Internationale de Philosophie* 4(11): 20-40.

Dumitru, M. (2004). *Explorări logico-filozofice*. Bucureşti: Humanitas.

Florian, M. (1983). *Recesivitatea ca structură a lumii*. Bucureşti: Editura Eminescu.

Revista Română de Filosofie Analitică,
Romanian Journal of Analytic Philosophy
Vol. XVI, Iulie-Decembrie 2022, Nr. 2, p. 179-203
ISSN (ediția electronică): 1843-9969,
ISSN (ediția tipărită): 1844-2218
DOI: 10.62229/rrfaxvi-2/9

PRIMARY EXAMPLES

RADU TULAI[1]

***Abstract*:** As characterized here, primary examples are universal judgements which make the connection between theory and natural language, and contribute to the instantiation of a theory or a whole research domain. They are segments of reasoning that are constructed from perceptual states had as part of wide-ranging experiences accessible to those thinkers who encounter primary examples.
There are at least three traits that Primary Examples must have in order to be considered as such. Primary examples must be pertinent to a theory, to the point of being fully inserted and implemented in the theory. Primary examples must exhibit generality, where generality is seen as the template for a general dynamic hidden in space-time and revealed by primary examples. Primary examples must be convincing, to the point that by nature of their perceptual states, their reasoning should prove undeniable, as opposed to abstract principles and notions.
In order to prove the merit of primary examples, I will refer to the literature on objectivity in science and show how primary examples are employed in scientific experiments.

***Keywords*:** primary examples, ordinary examples, pertinence, generality, convincing character, interdisciplinarity, theory-starting perceptions

1. The early stages of a Theory, or of a Whole Research Domain

One may conjure up primary examples by using examples as patterns to be integrated into a nascent theory, and these patterns arise by interpreting observations. A certain interpretation may be accepted and

[1] Radu Tulai is a doctoral student in the Faculty of Philosophy at the University of Bucharest.

turned into a primary example only if adequate consensus is reached, fitting a universal judgement. A primary example is capable of crystallizing into an abstract concept, but the understanding of a primary example must come before the understanding of that actual concept.

Primary examples initiate the construction of a new theory: they are pertinent examples, exhibit generality, and persuade thinkers that the theory before them is valid.

As pertinent examples, primary examples should provide a reliable connection with an ideal world where the concept based on the primary example is already used and fully grasped. The example is pertinent due to the fact that it is within the scope of the theory. Ordinary examples might be used in an analogy, or metaphorically, but primary examples would employ the exact concepts of the theory, via perceptual states, and so would offer a positive example that can be effectively incorporated inside the theory, and it would support the theory, giving it an empirical basis.

A primary example exhibits generality, by employing perceptual states that confirm a dynamic hidden in space-time, in between several entities. This general character has nothing to do with abstraction. The generality of primary examples has to do with the scope of their application, a system level perception, a law-like construct of nature. A counterexample exhibiting generality would be an anecdote. An anecdote could be an ordinary example, a situation where we have an individual account of some incident that is singular and in no way replicable or repeatable, an ordinary happening.

A well-formed primary example can convince thinkers of the validity of a nascent theory. This means that when the primary examples utilize content based on perceptual states, they do so in order to be beyond argument. If you can't trust your perceptual states, or the result of measurement instruments, then how can you trust any other input? Abstract principles can easily be mistrusted, but it is not so when it comes to your own observations. In contrast, ordinary examples are not used for the purpose of persuasion; they are used in an anecdotal way to explain more accessibly a difficult to comprehend theory. The persuasive drive of primary examples is not found in well-crafted arguments or in appeals to

the biases of thinkers; the persuasion of primary examples lies in them being holders of perceived truth.

2. Primary examples as natural kinds

In his book *The Philosophy of Nature. A Guide to the New Essentialism*, Brian Ellis discusses Modern Essentialism, as opposed to Passivism, and a new take on an ancient theory about the sources of power and order in the natural world. Ancient essentialism posits that the laws of nature are immanently to the natural world, and that these laws are not imposed on the world from external sources. There is no outside source of power like God, or even external laws of nature. The dynamic of the natural world progresses the way it does due to intrinsic causal powers of the basic constituents and their arrangement (Ellis, 2002: 1).

According to Passivism, the constituents of nature are essentially passive and obedient to the laws of nature, or God. God and the laws of nature are outside forces that impact passive and inert nature. Inanimate things may only act as commanded, either by God, or by the laws of nature, and so things behave the way they do not by inherent forces in those things themselves, but by the action of forces outside of nature (Ellis, 2002: 2–3).

A major proponent of classical essentialism is Aristotle. Aristotle believed that the things we may come across in the world can exist by nature, or they can exist by art, or even that they can exist by chance. Of the things that exist by nature, Aristotle included animals and their parts, plants, and the elements composing the Earth and the celestial bodies. Natural kinds are distinguished from the other two kinds of things by the fact that each has inside itself a multitude of principles for change and resistance to change; natural kinds are intrinsic causes related to their role in nature, and to their development (Ellis, 2002: 9–10).

Modern essentialism keeps the idea, advanced by Aristotle, that there are natural kinds of substances, but it rejects essentialism about animal and plant species, because the true natural kinds act at a deeper level than living species (Ellis, 2002: 12).

Modern essentialism proposes the concept of metaphysical necessities, propositions that are true in virtue of the essence of things. To understand metaphysical necessities one must consider the distinction between "real essences" and "nominal essences". Real essence is the set of properties that make a thing belong to a certain natural kind. A nominal essence will refer to how we classify things in the world, and the language used in order to mark out those different classes. The distinction made here is that natural kinds, or real essences, are independent of our conceptualization of reality, and independent of the language used to illustrate it. One may discover natural kinds, or real essences, by scientific investigation (Ellis, 2002: 15-16). How, more precisely, kinds and essences relate, and whether there is a form of metaphysical dependence of kinds on essences, or perhaps even a constitutive relation between them, goes beyond the scope of this investigation. The dialectic surrounding these question is carefully mapped in Fine (2020).

As seen above, one may be able to infer that natural kinds are independent of any human's interests, or predispositions, that they can only be discovered, and not invented in any way by any outside agent, as they are the result of intrinsic causal forces.

Steve Woolgar, in his book *Science: The Very Idea*, argues that the objective world is constituted in and through discourse, and rejects realist epistemologies. He states that the objective world is made accessible by means of a multitude of reporting and recording strategies that are given in an assembly of representations. Every scientist should try and collect several independent observations before any statement can be made. The more independent observations corroborate a scientific point of view, the closer we are to the truth. However, Woolgar states this is not enough: no object lies beyond discourse; facts and objects present in the world are necessarily textual constructions (Woolgar, 1988: 72-73). Woolgar's point of view tends to negate the existence of the real essences, and the existence of the intrinsic causal powers of natural kinds. Scientists can only refer to nominal essences. For a contrasting viewpoint that explores the possibility of referring to real essences in the course of scientific inquiry, see Dumitru (2004, p. 50 ff., p. 146 ff.).

Boyd suggests that the philosophical review of scientific methods has concluded with what he calls "Locke's worst nightmare": that there are possible ways to represent the general and systematic knowledge of substance, only that same knowledge is gained by theory-dependent processes: "principles of classification, methods for assessing projectability and for assessing the quality and the evidential import of observations, standards for assessing explanatory power" (Boyd, 1991: 133). Still, proponents of scientific realism argue that inductive inferences which pertain to science and consider observable facts can be seen as reliable in light of methodological principles reflecting prior knowledge of unobservable real essences.

This is where primary examples enter the wider discussion. Primary examples are perceived processes that reveal hidden universal dynamics, and in turn these dynamics ground new theories and scientific paradigms. Primary examples are independent of reigning theories or scientific paradigms, and that is how new theories and scientific paradigms arise. Primary examples close the gap between beliefs, theory and their grounding through the ending of the "experimenter's regress", by way of successful scientific experiments, and such scientific experiments are themselves a type of primary examples. Primary examples are our perceptions of natural kinds. Still, the question arises: if primary examples are perceptions of natural processes, can they also belong to natural kinds?

For Ellis the term 'natural kind' should cover three categories: natural kinds of objects – substantive natural kinds, natural kinds of processes and events – dynamic natural kinds, and natural kinds of properties and relations – natural property kinds (Ellis, 2001: 71-74). Primary examples are covered as dynamic natural kinds. I now turn to how primary examples contribute to objectivity in science.

3. Objectivity in Science

In the book *Patterns of Discovery*, Norwood Russel Hanson discusses the theory-laden character of seeing, where observation is shaped by the

knowledge of the thing we are observing. One running example is the case of Tycho and Simplicius looking at the earth's brilliant satellite, while Kepler and Galileo see the earth spinning back into the light of the local star (Hanson, 1981: 19-20).

He states that observation in physics is not a simple encounter with various instances like unfamiliar and unconnected flashes, sounds and bumps, but rather it is a calculated meeting with these flashes, sounds and bumps of a particular kind. In so doing, an alternative account would seem not only false, but absurd (Hanson, 1981: 24).

Hanson further states that there is a gap between pictures and language, vision being essentially pictorial, and knowledge being fundamentally linguistic. Both vision and knowledge are indispensable items for seeing. Still there are differences between the optical and conceptual features of seeing, and not all elements of statements correspond to the elements of pictures. However, there is a linguistic factor in seeing, without anything linguistic forming in our eyes. It is precisely this linguistic element in seeing that is essential in understanding theories and having knowledge. It is this linguistic factor that makes some observations relevant for knowledge (Hanson, 1981: 25).

It is at this crossroads, of seeing through our eyes and through our concepts, that primary examples lie. Of course, when talking about primary examples, we are talking foremost about perceptual states and sensory perceptions, not just seeing, that inform our observations. While connected with a relevant concept, primary examples can kickstart a theory. However, my take is that primary examples are not just another theory laden entity, but that perceptual states inform one of the competing theories in our mind, and the theory which more closely confirms the primary example is worth keeping. Primary examples are theory-starting entities, entities which enhance our perceptual states.

Suppose, as a matter of definition, that 'heat' means the energy associated with the random motion of the molecules of a substance (Hurley, 2000: 98). When temperature rises, the motion of the molecules accelerates, and when it decreases, the motion of molecules decelerates. Now imagine water boiling. Water molecules will accelerate faster and faster as the water reaches the boiling point. Now imagine an ice cube. All

its water molecules are inert. It is the linguistic understanding of heat that can supply this vision of the dynamic of a substance's molecules.

In Thomas Nagel's book, *The View from Nowhere*, in the chapter called "Knowledge", he provides three types of theories: skeptical theories, reductive theories, and heroic theories (Nagel, 1986: 68–69). He states that skeptical theories go beyond the grounds of our ordinary or scientific beliefs, and they make it impossible to defend them against doubt. Skeptical theories identify an unbridgeable gap between the content of our beliefs and their grounds (Nagel, 1986: 68).

Reductive theories are seen as a projection of skeptical theories. Reductive theories argue that we know certain things, but we can't know as much about the world if the skeptical gap between content and ground is as large as the skeptic claims. And so, the reductionist will employ a reinterpretation on the content of our beliefs that will claim less. For the reductionist, our beliefs are not about the world as it is in itself, they are about the world as it appears to us, following a general claim from the philosophy of Immanuel Kant (Nagel, 1986: 68–69).

The third kind of theories, heroic theories, will take stock of the large gap between the grounds of our beliefs about the world and the content of those beliefs within a realist interpretation, and they will try to bridge the gap without attempting to minimize it. Heroic theories may come in the shape of Plato's theory of Forms, or Descartes' defense of human knowledge resting on the existence of God (Nagel, 1986: 69).

For Nagel, an example of an objective step is the distinction between primary and secondary qualities, which is stated as a precondition for the development of modern physics and chemistry (Nagel, 1986: 75).

> "Things have colors, tastes, and smells in virtue of the way they appear to us: to be red simply is to be the sort of thing that looks or would look red to normal human observers in the perceptual circumstances that normally obtain in the actual world. To be square, on the other hand, is an independent property which can be used to explain many things about an object, including how it looks and feels." (Nagel, 1986: 75).

I believe that primary examples exhibit some kind of similarity to primary qualities, only instead of qualities, we have theory validating

processes involving several observed entities. These theory validating processes I will discuss in the following chapters as scientific experiments, a form of primary examples.

As regards the three types of theories listed above, I argue for a reductive theory, where we can only know as far as primary examples are able to show us and as far as the inductive processes based on primary examples can go.

Speaking of inductive processes, John D. Norton, in an article called „A Material Theory of Induction", speaks about the necessity of a material theory for the process of induction. For Norton, in a material theory, the admissibility of an induction is ultimately traced back to a matter of fact, not a universal schema. He pits a material theory against a formal theory of induction (Norton, 2003: 650). By promoting a material theory of induction, he states that: "All inductions ultimately derive their licenses from facts pertinent to the matter of induction." (Norton, 2003: 650) The licensing facts will be called material postulates of the induction. The inductions will be truth conducive only if the material postulates support them (Norton, 2003: 651).

Another important aspect about inductions revealed by Norton is the tension between universality and successful functioning. If the inductions are general enough so that they can be considered universal and true at the same time, the axioms and principles become vague or circular. There must be a principle of uniformity that limits the extent of the existing uniformity. This is so, for the world is not uniform except for certain instances, and those uniformities are called laws of nature. In so doing, the earlier material postulates are seen to be obtained only in specific domains, i.e. facts obtained locally (Norton, 2003: 651–652).

Norton then continues to consider different induction schemes. In section 6, called "The Problem of Induction Eluded?", Norton exposes the Problem of Induction attributed to David Hume. For Hume, a formal theory separates factual content from formal schemes. Because in a material theory there is no such separation, the problem of induction in Hume's terms does not arise (Norton, 2003: 666). Norton states that: "In order to learn a fact by induction, the material theory says that we must already know a fact, the material postulate that licenses the induction." (Norton, 2003: 666)

In a formal theory, if you justify an inductive inference schema by a deductive argument, it will fail because there is a violation of the inductive character of induction. If there is an effort to justify an inference schema by displaying many successful instances of the schema and applying the same inductive schema, it will fail on account of circularity. If you then try to justify an inductive inference schema by displaying many successful instances of the schema and applying a different inductive schema, it will fail due to infinite regress of meta inductions (Norton, 2003: 667).

In a material theory, if you justify an induction to a fact by a material postulate that is a universal truth known a priori, it will fail because it violates the local character of induction. If you justify an induction to a fact by using the same fact as material postulate, it will fail on the grounds of circularity. But if you justify an induction by using a different fact as the material postulate, a regress is triggered through a sequence of justifying facts, but it will not be infinite regress (Norton, 2003: 667). This is true, as is the case in science, that facts are inductively grounded in other facts, and these grounding facts will be grounded in other facts, and so on (Norton, 2003: 668).

We may understand primary examples as facts. For Norton, in a material theory, the admissibility of an induction is ultimately traced back to a matter of fact, not a universal schema. Primary examples are instances in space time, observed processes involving several entities, that can validate a theory or a series of statements. Primary examples are facts. Due to the pertinence of primary examples, they will only be active in the specific domain where they are postulated.

Another important aspect similar to the material theory of induction presented by Norton lies in avoiding the infinite regress characteristic of inductive processes. Norton suggests that the material postulate of an inductive process is a fact that is also, in turn, based on another fact. The only problem for Norton was that, if the regress will continue, the termination could be problematic or circular (Norton, 2003: 668). Primary examples would be a fitting termination for the cascading regress of facts. The termination will be an objective observation about the process or about several pertinent entities, quite possibly a scientific experiment that must prove the observation.

Julian Reiss, in his article "What are the drivers of induction? Towards a Material Theory+", validates Norton's Material Theory of Induction. He states that material facts are drivers of induction (Reiss, 2020: 10). But Reiss also is quick to point out other drivers that Norton left out. Reiss counts up to six other drivers such as theories, idealizations, purposes, ethical norms, methodological norms and conceptual norms (Reiss, 2020: 11-15). While I agree first and foremost with Norton, I can admit that for a time, even in the hard sciences, a theory, such as the theory of gravitational waves, can be a driver for induction. But ultimately, the proof lies in the primary examples, or belongs to a successful scientific experiment that was able to detect such a gravitational wave. The same goes for the remaining five drivers on Reiss' list.

Circling back to Norton and the danger of infinite regress, we find Harry M. Collins' book *Changing Order*. In this book, Collins talks about the Experimenter's Regress, and he chooses TEA Lasers[2] and gravitational waves as case studies into this matter. In the case of gravitational waves Collins identifies an Experimenter's Regress:

> "What the correct outcome is depends upon whether there are gravity waves hitting the Earth in detectable fluxes. To find this out we must build a good gravity wave detector and have a look. But we won't know if we have built a good detector until we have tried it and obtained the correct outcome! But we don't know what the correct outcome is until … and so on ad infinitum." (Collins, 1985: 84)

Collins further states that experimental work is a valid test only if we can find a way to break the circle of experimenter's regress. In the case of TEA lasers, the experiment was successful because the TEA Laser was able to vaporize concrete, an actual quantifiable result (Collins, 1985: 84).

Gravity waves, or gravitational radiation, are the invisible gravitational equivalent of light or electromagnetic radiation. Moving massive bodies in space will produce gravity waves. These gravity waves are weak, and their detection is difficult. In the case of gravitational

[2] TEA Laser stands for Transversely Excited Atmospheric pressure CO_2 Laser

waves, a device was needed to detect them, and to distinguish between vibrations due to gravitational waves and those induced by any other force. The machine designed by Professor Joseph Weber of the University of Maryland was constructed and, in spite of precautions, the detector was not completely quiet, because there were vibrations induced by the random movements of the detector's atoms. A continual output of thermal noise was registered. It was decided that a gravity wave would be represented by a particularly high peak, and so a decision was made to about the threshold beyond that a gravity wave would be detected. In 1969 Weber declared that he had detected about seven peaks every day, which could not have been noise. His claims were refuted because of the too high frequency of detected gravitational waves, and because the findings were not replicable (Collins, 1985: 79–81).

We can plainly see the need for a suitable outcome of the scientific experiment, an adequate primary example, an appropriate instance where we would have an observable event between several entities[3] that would count as the actual detection of a gravity wave, and that event, on the grounds of generality, should be replicable. We will dig deeper into particular primary examples in the next section about scientific experiments.

As a short summary, we have seen how, according to Hanson's view, our perceptual states and our theories influence each other. At their crossroads, we find primary examples which, by relevant empirical observation, validate a theory. In turn, the theory enhances our perceptual states. Then, in Nagel's book, we have acquired some sense about three types of theories, skeptical, reductive, and heroic, and we have seen that primary examples are seen as reductive theories, because one may build a theory, or build inductive processes, only on the basis of what primary examples would allow. Primary examples relate to primary qualities, in the sense that they are relevant aspects about starting a theory. A primary example will refer only to primary qualities. With the help of Norton, we have seen that a material theory will steer clear an inductive process from infinite regression, and that primary examples or objective observations

[3] In this case, perhaps an interaction between some massive bodies like, for example, black holes, or neutron stars.

about pertinent entities will break the infinite cycle. We then saw that Reiss would add other six drivers for induction, and I agreed that contingently, at different levels of construction, a scientific account may accept those other six drivers, but the ultimate proof will lie in a primary example, or successful scientific experiment. Then, cycling back to infinite regression, we found ourselves facing the experimenter's regress, coined by Collins. The experimenter's regress was broken by an example of the TEA Laser's successful experiment, but when it came to the experiments about gravitational waves, or gravitational radiation, the regress couldn't be broken in the absence of a primary example, or a successful scientific experiment.

We now will turn to examples of successful scientific experiments that can be seen as primary examples.

4. Scientific experiments as cases for primary examples

During 1919, on May 29, a solar eclipse[4] was visible across the continent of South America, the Atlantic Ocean and the continent of Africa. There were two teams of astronomers, based in Sobral, in Brazil, and in the island of Principe, near the African continent, and their mission was to capture and analyze that solar eclipse. It would prove to be the best opportunity to submit to the test Einstein's theory of general relativity. Before that Solar Eclipse in 1919, the general relativity theory, published in 1915, was not at all popular among scientists. What the solar eclipse provided was a test to see if the bending of light by gravity was possible. This issue was one of the main consequences of general relativity. This theory stated that rays of light would bend when traveling by a massive body in space, just like our sun, and this is due to the fact that the rays would have to follow the curving of space-time created by the mass of that immense body. The solar eclipse was important, because, in normal

[4] Information about this topic was taken from the following website: https://www.britannica.com/story/the-solar-eclipse-that-made-albert-einstein-a-science-celebrity

conditions, sunlight would shine together with nearby stars, and so, the effect of gravity was not noticeable. But with the solar eclipse, astronomers would take pictures of that eclipse, when the sunlight would be cancelled by the moon. Only then, the astronomers were able to see if the field of stars would have their light bent by the gravity of our sun. The analysis took a few months, and so the researchers confirmed Einstein's theory.

This successful experiment was a primary example. Primary examples are patterns of data of a theory, and they confirm that theory by their observation. Primary examples are pertinent to their research field due to the fact that they utilize and confirm the fundamental data about a theory. The experiment had an element of generality as it confirmed the theory of general relativity. A primary example is convincing by virtue of its mere observation, and by the way an experiment leading up to that perception was well designed, and if it was able to reach a concrete conclusion. Einstein's primary example, the interaction of light with the gravity of the sun, was so robust, that his preliminary unpopular theory still stands to this day.

Another theory of Albert Einstein's, in the form of the gravitational waves,[5] discussed earlier when we were referring to the experimenter's regress, was finally confirmed in 2015–16, successfully breaking the experimenter's regress by a direct detection. There were some preliminary confirmations on gravitational waves in the form of watching radio flashes emitted by a pair of neutron stars whirling around one another and shifts in the timing of the flashes matched Einstein's predictions.

On September 14, 2015, there was an oscillation detected that began at 35 cycles per second (hertz), and then the oscillation increased to 250 hertz. Afterwards the oscillation became chaotic and abruptly died down. The whole process lasted one-fourth of a second. The gravitational ringing was caused by the collision of two black holes located at about 400 megaparsecs (1.3 billion light years) from Earth.

On February 11, 2016, the Advanced Laser Interferometer Gravitational-Wave Observatory (LIGO) announced the direct detection

[5] Information on this topic was taken from the following website: https://www.nature.com/articles/nature.2016.19361

of gravitational waves. LIGO is made up of twin interferometers that bounce laser beams between mirrors at the opposite ends of 4-kilometre-long vacuum pipes set perpendicularly to each other. A gravitational wave passing through will alter the length of one of the arms, causing the laser beams to shift slightly out of sync.

This successful experiment can be seen as a primary example. The LIGO scientific experiment is pertinent to the issue of Einstein's general relativity theory. Moving massive bodies in space will cause gravitational waves, and the experiment has proven that. The experiment expounds generality. We now have proof that the movement of massive bodies emits gravitational waves in the same manner that a star may radiate light and electromagnetic radiation. The experiment is also replicated by another detection in 2017.

This same fact of direct detection also contributes to the convincing character of primary examples. It was direct detection, announced a century after Einstein's hypothesis, that really convinced the scientific community, and validated the general theory of relativity. As was the case with the 1919 solar eclipse, the direct detection, or the enhanced perceptual states of the scientific experiment were key in proving the primary example of gravitational waves.

We now turn to the equivalence[6] principle, tested by the MICROSCOPE satellite experiment, built by the French National Center for Space Studies (CNES). The equivalence principle states that two objects of unequal mass dropped in a vacuum will reach the ground simultaneously. It is said that Galileo tested this theory from the top of the Leaning Tower of Pisa. In 1971, astronaut David Scott dropped a hammer and falcon feather on the surface of the moon, to test this principle.

Since then, several iterations of testing the equivalence principle have taken place. The MICROSCOPE experiment, published in 2022, verified that two masses of titanium and platinum aboard a satellite orbiting Earth fall exactly in the same way to a precision of 1 part per 10^{15}.

[6] Information on this topic was taken from the following website:
https://physics.aps.org/articles/v15/94

In the experiment there are two coaxial cylinders of titanium and platinum that are placed in free fall in Earth's gravitational field. The experiment confirmed the equivalence principle.

There is the possibility of a MICROSCOPE 2 experiment to test the hypothesis to a precision of 1 part per 10^{17}. Should the experiment uncover violations, that fact will lead to new physical theories to explain dark matter or dark energy.

The MICROSCOPE satellite experiment is another case of a primary example. The interaction of the two cylinders of titanium and platinum, in free fall, aboard a satellite orbiting Earth calculating the interaction to a very high precision, is pertinent to establishing the equivalence principle, which in turn answers Einstein's general theory of relativity. The satellite experiment is general. The experiment targeted the dynamic of bodies with mass in free fall.

The satellite experiment is convincing. The experimenters chose the setting of the experiment in space, in Earth's gravitational field, where there was no interference to Earth-bound measurements like seismic vibrations, or gravitational-field variations caused by nearby mountains. The high precision of the measurements was also a convincing factor.

And of course, as in all scientific experiments presented here, the high precision perceptual states of the experiment were key in providing relevant results. We have seen how the perceptual states of the experiments, starting with the 1919 solar eclipse challenged the scientific orthodoxy of the day, and paved the way for Einstein's theory of general relativity, a very unpopular theory in the absence of what scientists could see with their own eyes. The story repeats itself with gravitational waves and with the equivalence principle. Those theory starting perceptions are primary examples. The shift occurred only because the observations could be interpreted to uphold the right theory, still unchallenged to this day.

All the scientific theories presented here are reductive, because they do not allow inferences outside what is proven via the primary examples encapsulated into scientific experiments.

All the scientific experiments gathered here refer to the primary qualities behind phenomena, and the perceptual states are fine tuned to

detect them, leading to primary examples that confirm the theory via scientific experiments.

All of the scientific experiments listed here break the experimenter's regress, referring to Einstein's general theory of relativity by providing observations based on highly sensitive and highly precise perceptual states that validate the theory and enhance our view of the universe. These perceptual states can be seen as primary examples, as snapshots of space-time that reveal a hidden dynamic.

5. The Connection between Paradigms and Primary Examples

As we have seen in the previous section, the solar eclipse experiment has started a new scientific paradigm: Albert Einstein's theory. Now we have to grapple with the connections between a paradigm and a primary example. In *The Structure of Scientific Revolutions*, Thomas Kuhn gives arguments for the priority of scientific paradigms. He states:

> "Paradigms may be prior to, more binding, and more complete than any set of rules for research that could be unequivocally abstracted from them", and that "scientists do not usually ask or debate what makes a particular problem or solution legitimate tempts us to suppose that, at least intuitively, they know the answer… [P]aradigms could determine normal science without the intervention of discoverable rules" (Kuhn, 1996: 46).

Another claim provided by Kuhn that supports the priority of paradigms when faced with rules and assumptions is the following:

> "Scientists, it should already be clear, never learn concepts, laws, and theories in the abstract and by themselves. Instead, these intellectual tools are from the start encountered in a historically and pedagogically prior unit that displays them with and through their applications. A new theory is always announced together with applications to some concrete range of natural phenomena; without them it would not be even a candidate for acceptance." (Kuhn, 1996: 46)

My claim is that the "applications" referred to by Kuhn are to be equated with primary examples, or with tools that are constructed from primary examples. Scientific paradigms and primary examples are connected. The primary examples initiate a theory, but they cannot account for the totality of the scientific paradigm. Primary examples are patterns of perceptual states that kickstart a paradigm, and they are fully integrated into that particular new paradigm. We can plainly see, in Kuhn own words, the importance of observations and perceptual states for the initiation of a scientific paradigm.

There is one more significant point found in Margaret Masterman's article "The Nature of a Paradigm". There are 21 meanings for paradigm in Kuhn's *The Structure of Scientific Revolutions*, that can be grouped into three categories: "1. a paradigm is a metaphysical notion or entity"; "2. sociological paradigms where paradigms act like a set of political institutions and where paradigms are recognized as a scientific achievement"; and "3. paradigms are artefacts that solve puzzles" (Masterman, 1970: 65).

Primary examples are to be found in the category of artefacts that solve puzzles. A primary example is a pattern, a distinct partition of space-time that is not yet a concept. A concept or an abstract principle can be constructed based on a primary example, and so I claim that primary examples are not metaphysical entities. Primary examples are only revealed by perceptions, and they impact perceptual states, transforming the perceived world in the eyes of a thinker.

6. The similarities and differences between primary examples, prototype theory, exemplar theory, and central cases

The theory of prototypes is a theory about concepts, and it is based on observing similarities. Prototypes were constructed as a response to the classical theory of concepts (Margolis & Laurence, 1999: 27). The source for this theory can be found in psychology testing. Test subjects would store information about categories, and would thus form representations in idealized models. It was discovered that many categories were formed

around sets of typical features. These typical features would be encountered frequently. The models that had the maximum number of features specific to a category were called "prototypes". Of course, it is possible that no member of a specific category will fully identify with the prototype, although a member that will reach as closely as possible to the prototype will be included in the category (Prinz, 2002: 52–53).

Notice the perceptual drive of this prototype theory on concepts, which is similar to the perceptual states necessary for the observation of a primary example. But primary examples do not function on the basis of similarity, and they do not refer to the definition of a single entity. Primary examples can be seen as examples, and they exemplify phenomena, events, or dynamic processes. They are characterized by functional definitions, and traits or features would not be adequate for their description. The experience of primary examples is scientifically formative.

Another theory based on similarities we may find in the exemplar theory. In this case, however, the similarities are tied to an actual member of a category, than with similarities that are connected to an idealized model with specific features. To construct a category, a subject will observe a group of individuals that are members of a category, and these individuals are exemplars. The subject will store information on exemplars, and concepts will result from the exemplar collections (Prinz, 2002: 63–64).

At first glance, it appears that the exemplar theory will contradict the commonly held view that concepts are abstract. Exemplars refer to the concrete example of a concept, or to a subset of the same concept. We may think about the concept "toys", one exemplar may be a concrete favorite toy car, or that toy car may be seen under the subset "toy cars", and that subset is abstract (Smith & Medin, 1999: 207).

The same applies: primary examples are not constructed from similarities, and they do not refer to individuals. They refer to the explication of a dynamic, of a process in between entities. In this sense, primary examples can be seen as a continuation of prototype theory and exemplar theory that refers to the observation of interactions in between entities that can be defined via prototypes and exemplars. But that interaction is scientifically formative, it constructs a new perspective based on

perceptual states that kickstart a theory. Primary examples are theory-starting perceptual states, that carry with them a new scientific point of view that may be incomensurable with other points of view. Similarity might be irrelevant here, whereas conformity seems more accurate.

For Margolis and Laurence (1999), a "central case" refers to a typical or standard example of a concept or phenomenon. It's an instance that most clearly embodies the characteristics or properties that are generally associated with a particular concept or category. Central cases are a reference point for understanding more complex or ambiguous instances related to a certain concept. They are used in arguments to clarify meanings and to draw distinctions.

Primary examples are constructed to establish a new theory or research domain, or even a new scientific paradigm, by way of perceptual states. Perceptions trump all other theoretical constructs. Where central cases can be abstract, primary examples are always empirically derived. While there is some degree of similarity with central cases, primary examples do not refer to ambiguous instances of a concept, and they are not used primarily to clarify meanings, or draw distinctions: that seems like a task for ordinary examples.

Primary examples are pertinent to a theory, as are central cases. Primary examples are general, while central cases don't need to be general. Quite the opposite: central cases are specific cases. Primary examples persuade by conforming to simple perceptual states, while central cases are convincing when they successfully clarify an ambiguous concept. All in all, primary examples can be seen as complementary to central cases, and they can be successfully used in conjunction to promote a new perspective on a scientific problem.

There is a corpus of scientific literature that refers to the term of "primary examples", with some similarity to central cases. Primary examples, in this corpus, refer there to the most typical of examples, the best illustration in a range of other examples about a certain topic; they are robust case studies. As I have shown, *the way I use "primary examples" pertains to an objective dynamic of natural kinds, revealing a hidden structure, or essence, that will provide the start for scientific paradigm* and a theory, within the limits of a perceived truth.

7. Primary examples belong to the theory-theory of concepts

The theory-theory of concepts is a response set against the theory of prototypes and the theory of exemplars. These other two theories, of prototypes and exemplars, are the opposition to my model and they rely on similarities, be it similarities based on a central tendency or similarities with an ideal model. Theory-theory claims that a concept is a part of a more comprehensive body of knowledge, and that concepts are likened to mini-theories of the categories that they represent (Prinz, 2002: 75–76).

The proponents of theory-theory suggest that we explain our environment by theoretical means, and that this process begins at a relatively early age. Children, like scientists, construct primitive theories about their environment, and their first conceptual network will offer an initial theory – the original Sim. This original Sim will be superseded by complex and subtle theories and concepts about their environment, and this is how children, like scientists, overcome their initial theory – the original Sim. This theory was supported by psychological tests (Keil, 1996: 247–265). This is precisely how primary examples function. They reveal a hidden dynamic that eventually leads to theoretical changes.

The role of concepts, according to theory-theory, is to codify information about explanatory connections between components of theories. Concepts codify, with priority, hidden features by resorting to essences (Prinz, 2002: 77).

Other psychological studies suggest that, unlike simple mechanisms that look for visible characteristics and similarities, like prototype and exemplar theories, there are occasions where subjects consider if a certain element possesses "the right hidden quality". This shows that subjects envisage sophisticated structures that entail more than just a simple placeholder for essences. If time is short, subjects will resort to prototype and exemplar theories, but for systemic and complex theoretical judgements the theory-theory is more suitable. Proponents of the theory-theory can easily explain the difference in between children and adults, and I will add that the same difference in between random people and scientists can also be explained, thus illustrating cognitive development. The difference lies in the very distinct theories possessed by children,

random adults, and scientists about their environment (Margolis & Laurence, 1999: 46).

But how are primary examples employed inside a theory-theory framework? Theory-theory proposes structural characteristics, functional characteristics, and dynamic characteristics, all built into a representation. Primary examples are to be found in the dynamic characteristics group. Dynamic characteristics are seen as successive stages. The first stage relates to the accumulation of counterarguments referring to an established theory. Here, primary examples are to be employed as pertinent examples, that are connected to a general process, and that are convincing as they show the perceived truth. Then, the established theory will deny the counterarguments, a necessary stage that takes into account the proofs brought forth into the guise of primary examples contrary to established theory. In the next stage, we have ad-hoc development of auxiliary hypotheses. This is the stage where the dynamic of primary examples are codified into concepts. Then, establishing an alternative model for the initial theory will follow. At this stage, the codified concepts, based on primary examples, are joined into theoretical texts. The last stage, according to theory-theory, is a period of intense experimentation and observation, where the primary examples, in the guise of scientific experiments, are thoroughly tested, and they are challenged en route to becoming new scientific norms (Gopnik & Meltzoff, 1997: 39–41).

There are flaws with theory-theory. Theory-theory suggests that concepts are incorporated into theories, they cannot be separated by theories, and that they incorporate theoretical information. Concepts will "inhabit" theories. This means that the nature of concepts remains unclear and unspecified (Prinz, 2002: 81).

It may seem that prototypes and exemplars have an advantage here, seeing that they are heavily individuated. Another problem sometimes leveled against theory-theory is holism: theory-theory cannot individuate concepts, which exhibit relativity to bodies of knowledge. As a result, a new theory cannot account for an initial theory and the old concepts of that theory. This entails that new concepts will have to be explained from the start, and cannot benefit from an older theory (Fodor, 1998: 114–115).

Prototype and exemplars have no problems of individuation. However, the individuation problem can be overcome by adopting a classical theory of concepts built upon the theory-theory stage of primary examples.

On balance, it seems that primary examples are natural kinds, couched in the theory-theory of concepts, running deeper than similarities pertaining to prototype and exemplar theories. Primary examples are a particular type of examples; they are snapshots of space-time; pertinent, in no way metaphorical and exhibiting generality.

We have seen that primary examples can be accepted into the exact sciences as scientific experiments, at least. But what can we say about mathematics? When they are introduced to the mathematical operations of addition and subtraction in primary school, students are not taught complicated abstract processes, where concepts about numbers and their operations are debated. Instead, they are introduced to mathematics via primary examples. They learn how to add, and subtract, small numbers, in the limit of the sum of 10, via examples. After they master this initial stage, they learn how to multiply and divide, and for this task they need to learn the multiplication table, which is a list of examples, by heart. These examples employed in learning mathematics exhibit generality; you can apply them to any physical magnitude.

8. Conclusion

I conceive of primary examples as based on the perceptual states involving several entities, examples which ground scientific objectivity. I have provided criteria for the identification of primary examples, in the guise of pertinence, generality, and convincing character.

The experimenter's regress can be overcome by appropriate primary examples, most prominently shown in the experiments regarding gravitational radiation, or gravitational waves. As for the 1919 solar eclipse, I have shown how a direct view of the bending of light by the enormous mass of the sun kickstarted a new scientific paradigm attributed to Albert Einstein. In the MICROSCOPE satellite experiment,

we have seen the robustness of the equivalence principle, an important part of general theory of relativity.

Finding primary examples is important because it is the most influential way to overcome the experimenter's regress and stay true to the reductive nature of scientific experiments, where all we can know is directly linked to primary examples in the form of scientific experiments. We may exhibit a healthy dose of distrust when it comes to abstract reasoning, but one may not deny evidential status to perceptual states had as part of scientific experiments.

Primary examples assert the primacy of observations and perceptions in the face of theoretical constructs, abstract principles and theories. Primary examples are more likely to be seen as theory-starting perceptions, rather than theory-laden. It is true that we meet with scientific observations and perceptions in a calculated way, but the calculations employed need to be empirically tested, confirmed, and replicable.

Can the task of finding new primary examples be relegated to artificial intelligence systems? Primary examples are not exactly simple patterns that exist in nature. Artificial intelligence systems could be able to look for generality in certain patterns, but the other two criteria (of pertinence and convincing character) call for a qualitative jump in technology.

References

Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 61(1/2), 127–148.

Brax, P. (2022). Satellite Confirms the Principle of Falling. Physics, 15, 94. DOI: 10.1103/Physics.15.94

Castelvecchi, D., & Witze, A. (2016). Einstein's gravitational waves found at last. *Nature news*, 19361. https://doi.org/10.1038/nature.2016.19361

Collins, H. M. (1985). Detecting Gravitational Radiation: The Experimenter's Regress. *Ders.: Changing order. Replication and induction in scientific practice*. London/Beverly Hills: Sage Publications.

Dumitru, M. (2004). *Explorări logico-filozofice*. Bucureşti: Humanitas.

Ellis, B. (2001). Natural Kinds. *Scientific Essentialism* (pp. 61–105). Cambridge: Cambridge University Press.

Ellis, B. (2002). Introduction. Essentialist Philosophies of Nature. In *The Philosophy of Nature: A guide to the New Essentialism* (pp. 1–20), Chesham: Acumen.

Fine, K., & Dumitru, M. (2020). Comments on Jessica Wilson's "Essence and Dependence. *Metaphysics, Meaning, and Modality: Themes from Kit Fine*, 471–475

Gopnik, A., & Meltzoff, A. (1997). The Theory Theory. In *Words, thoughts, and theories* (pp. 11–48). Cambridge, MA: Massachusetts Institute of Technology Press.

Hanson, N. R. (1981). Observation. In *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science* (pp. 4–30), Cambridge: Cambridge University Press.

Hurley, P. J. (2000). Definitions and Their Purposes. In *A Concise Introduction to Logic* (pp. 96-104), Wadsworth Publishing.

Fodor, J. (1998). Prototypes and Compositionality. In *Concepts*. *Where Cognitive Science Went Wrong* (pp. 88-119), Oxford University Press.

Keil, F. C. (1996). Escaping the Original Sim. In *Concepts, Kinds, and Cognitive Development* (p. 247-266), Massachusetts Institute of Technology Press.

Kuhn, T. S. (1996). The Priority of Paradigms. In *The Structure of Scientific Revolutions* (pp. 43–51), University of Chicago Press, Third Edition.

Margolis E. and Laurence, S. (1999). Concepts and Cognitive Science, in *Concepts. Core Readings, E. Margolis. & S. Laurence* (eds.), Massachusetts Institute of Technology Press.

Masterman, M. (1970) The Nature of a Paradigm, in *Criticism and the Growth of Knowledge*, Cambridge University Press.

Nagel, T. (1986). Knowledge. In *The View From Nowhere* (pp. 67-89), New York, NY: Oxford University Press.

Norton, J. D. (2003), "A Material Theory of Induction", *Philosophy of Science*, 70(4): 647–670.

Prinz, J. J. (2002). Similarity/Based Accounts. In *Furnishing the Mind. Concepts and their perceptual basis* (pp. 51–74), Massachusetts Institute of Technology Press.

Reiss, J. (2020), What Are the Drivers of Induction? Towards a Material Theory+, *Studies in History and Philosophy of Science* Part A 83: 8–16.

Smith E. and Medin, D. (1999). The Exemplar View, in *Concepts. Core Readings*, E. Margolis & S. Laurence (eds.), Massachusetts Institute of Technology Press.

The Editors of Encyclopaedia Britannica (2023, June 23). *The Solar Eclipse That Made Albert Einstein a Science Celebrity*. *Encyclopedia Britannica*. https://www.britannica.com/story/the-solar-eclipse-that-made-albert-einstein-a-science-celebrity

Woolgar, S. (1988). Arguing Science: Discourse and Explanation. In *Science: The Very Idea* (pp. 67–82), London: Tavistzock.

---

tipografia.unibuc@unibuc.ro

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

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