History and Philosophy of Science;
Foundations of Physics

Utrecht University: Master thesis

Worldly Patterns

Emergence, Functionalism and Pragmatic Reality in Wallacian quantum mechanics

"While the wave-function realist will deny that three-dimensional objects and spatial structures find a place in the fundamental ontology, this is not to say that the three-dimensional objects surrounding us, with which we constantly interact, and which we perceive, think and talk about, do not exist; that there are no truths about them; it is just to maintain that they are emergent objects, rather than fundamental ones. But an emergent object is no less real for being emergent."

— David Wallace & Christopher Timpson

Author:
Ruward Arthur Mulder

Advisors:
Dr. **Guido** Bacciagaluppi
Dr. **Ronnie** Hermens

15th August, 2018

But I love extravagance
And wanting it has handed down
The glitter and glamour of the sun
As my inheritance

—Sappho of Lesbos (630-570 BCE)

Scientific summary

David Wallace's *The Emergent Multiverse* is the most recent complete presentation of a version of Everett's theory of quantum mechanics that has attracted much scientific activity in the past decade. I present a brief sketch of Wallace's solution to the measurement problem, arguing that the many-worlds interpretation is not as far-fetched as it is often conceived to be. Taking the wavefunction as the fundamental ontology, it claims to solve the measurement problem by recognizing certain (quasi-classical) patterns of the wavefunction in $3N$-dimensional configuration space that functionally behave like the (classical) configuration-space pattern of a N-particle system described by classical mechanics. In this sense, structures within the universal wavefunction are identified with classical 'worlds' at the coarse-grained level. I highlight two elements of this work: the role of emergence and the functionalist framework that Wallace imports through what he calls 'Dennett's criterion'. This criterion appeals to the virtue of usefulness (in the form of predictability and explanatory power) as a criterion for the reality of 'patterns'.

It is shown that, due to decoherence, quasi-classical worlds emerge weakly, similar to that of the emergence of thermodynamic temperature from statistical physics, in the sense that they are autonomous and unexpected with respect to the lower-level domain (as opposed to strong emergence of some *insuper et supra* high-level ontology). However, the use of Dennett's criterion obscures this result, laying bare some philosophical issues, which we address over three axes of distinctions: (i) the objective/subjective-axis, (ii) the quantitative/qualitative-axis within the framework of intertheoretic reduction and (iii) the ontological/epistemological-axis.

First, Daniel Dennett's 'real patterns' are compared to Wallace approach to patterns. Then, I point out (i) an analogy with Bas van Fraassen's idea of causal patterns that become salient due to pragmatic explanation, namely that in the context of a pragmatic goal the quasi-classical pattern is made salient over other objectively existing, non-classical, patterns. I conclude that in the absence of human goals there is no reason to regard the quasi-classical pattern as 'more real' than other patterns. In analogy to Dennett's 'intentional stance', Wallace is committed to a 'classical stance', equivalent to breaking Hilbert space democracy of bases.

Although Wallace's version is a weaker one, his (ii) relations between theories bare resemblance to the reductionist program, and I argue that, next to quantitative deduction, additional conceptual 'bridge principles' à la Ernest Nagel are needed.

The appeal to usefulness as a criterion for reality (iii), I claim, is not necessary to solve the measurement problem itself, but has the further (unwarranted) goal of establishing 'real' worlds. I spell out a solution to the measurement problem, the many-minds theory, which solves the problem along the same lines as *The Emergent Multiverse*, with the exception that the – although the quantum world itself is real – the classical worlds with definite properties are beliefs in the superposed brains of observers.

Persoonlijk voor- en dankwoord

Voor u ligt een scriptie waarmee ik na acht jaar mijn academische opleiding afrond. Na vele jaren te hebben gezworven tussen verschillende velden – natuur- en sterrenkunde, geschiedenis, filosofie, kunstmatige intelligentie en biologie – en het afronden van een opleiding theoretische fysica een jaar geleden, ben ik uiteindelijk trots om de spreekwoordelijke punt te zetten achter de master History and Philosophy of Science (HPS). Deze omzwervingen zijn het resultaat geweest van een gevoel van onbehagen, dat groeide naarmate mijn beeld van de wetenschap afbrokkelde. Steeds meer leek de praktijk niet te bevestigen wat wetenschap mij in eerste instantie toescheen: een manier om onomstotelijke en objectieve feiten te verwerven; om uit te leggen hoe de natuur daadwerkelijk werkt; om menselijke factoren, en vooral politiek en maatschappij, weg te filteren; om universele kennis cumulatief te verwerven vanaf de Griekse Oudheid tot de dag van vandaag. Dat onbehagen is eigenlijk al een *one-way-ticket* naar HPS. Vanuit daar kon ik laten zien wat de wetenschap allemaal níét is; en dat heb ik dan ook met veel overtuiging proberen te doen. Maar het bleek dat ook daar naïviteit loerde en dat ik met enkel kritiek ook niet uit de voeten kon. Ik voelde een behoefte aan een *retour-ticket*, om dat onwaarschijnlijke idee 'wetenschap' weer te gaan verdedigen. Na deze jaren denk ik dat het einde van die tocht voor mij persoonlijk in zicht is; al die dingen die wetenschap niet is – universeel, objectief, cumulatief – maken het voor mij nu juist geloofwaardig en overtuigend. We kunnen feiten wél aannemen van 'autoritaire' wetenschappers, niet omdat zij de waarheid voor eens en altijd in pacht hebben maar omdat zij een leven lang werken om feiten robuust te krijgen. Wetenschap is een culturele onderneming welke wat mij betreft het 'gezonde verstand' idoliseert. Waar het dat voorheen niet deed, lijkt mijn beeld nu veel meer aan te sluiten bij de realiteit. Kortom, ik heb de mogelijkheid gekregen om mijn persoonlijk wereldbeeld te vernietigen en het weer op te bouwen en dat is precies wat we van de universiteiten moeten verwachten—en wat ze nooit als taak mogen vergeten.

Uiteindelijk resulteert bovenstaand proces in twee overtuigingen. Ten eerste dat ik al die tijd (tegen alle initiële pogingen in) tóch aan een maatschappelijk project heb zitten werken; dat is tevens een realisatie die ik met beide handen heb aangegrepen door ook buiten de Ivoren Toren zo veel mogelijk verantwoordelijkheid op me te nemen. En ten tweede is het juist dankzij die herwaardering voor wetenschap dat ik de natuurkunde niet heb losgelaten en alsnog een these heb geschreven over een thema in de grondslagen van de natuurkunde, want naar mijn mening zijn conceptuele analyses van gangbare wetenschappelijk theorieën noodzakelijk. Filosofische kritiek zal altijd nodig zijn, ook voor de komende generaties, zowel binnen de wetenschap als daarbuiten. Het levert handvatten voor beter begrip, leidt soms zelfs tot nieuwe theorieën en houdt bovenal de – altijd verleidelijke – dogmatische interpretaties op afstand.

Deze these, om maar eens op de inhoud over te gaan, is zo'n natuurkundige/filosofische hybride. Voor al die vrienden en familieleden die mij meerdere malen hebben gevraagd wat ik nu eigenlijk al die tijd aan het doen ben, zal ik hier een poging wagen die vragen te beantwoorden. Als uitgangspunt neemt dit werk enkele kernproblemen in onze meest succesvolle theorie van de natuur, de kwantummechanica. Deze problematiek wordt vaak afgekort als 'het meetprobleem' en ligt ten grondslag aan de (letterlijk!) eeuwige anarchie van verschillende theorieën (vaak onterecht verschillende 'interpretaties' genoemd), waarvan hun aanhangers beweren dat ze de problemen oplossen zonder in te leveren op de empirische adequaatheid (de aansluiting met de waarneming). Dit werk is aanvankelijk dan ook een verkenning van één van de mogelijke oplossingen van die problematiek: de relatieve-toestandstheorie van Hugh Everett III, in de volksmond (het volk bestaat uit een aantal verdwaalde natuurkundestudenten) ook wel de 'veelwerelden-interpretatie' genoemd. Everett schreef aan de universiteit van Princeton aan het einde van de jaren vijftig zijn visie op in zijn proefschrift onder John Archibald Wheeler. Zijn werk was gedetailleerd, radicaal vernieuwend op vele vlakken (zoals men dan zegt was hij 'zijn tijd vooruit') en formuleerde buitengewoon scherp de tekortkomingen van de gangbare theorie: Kopenhagen. Everett zelf vervolgde zijn academische carrière niet, misschien mede door de overmatige (en achteraf onterechte) kritiek van leidende natuurkundigen als Niels Bohr en Léon Rosenfeld. Everetts ideeën, echter, aangevuld en aangejuicht door bekende namen (denk onder andere aan Bryce DeWitt, Murray Gell-Mann, John Wheeler zelf), zijn sindsdien eigenlijk voortdurend onderhevig geweest aan een populariteitsconjunctuur; en juist die lijkt in de afgelopen jaren weer te pieken.

Het idee van de veelwerelden-theorie is dat de eigenaardige superposities van de kwantumfysica – superposities zijn een wiskundige beschrijving waarbij het lijkt alsof objecten tegenstrijdige eigenschappen tegelijkertijd kunnen bezitten (zoals een op hetzelfde moment linksdraaiend én rechtsdraaiend voorwerp) – verschillende 'parallelle werelden' beschrijven, waardoor de inwoners van één zo'n wereld geen tegenstrijdigheden opmerken (omdat ze zich dan óf in de linksdraaiende óf in de rechtsdraaiende wereld bevinden, in plaats van in allebei). Dit omzeilt een aantal problemen met andere theorieën—bijvoorbeeld: waar de Schrödingervergelijking die de bewegingen van kwantumsystemen beschrijft volgens sommige kwantumtheorieën in kunstmatige omstandigheden om onduidelijke redenen niet meer geldt, is deze voor Everett universeel en ononderbroken geldig. Daarbij is het belangrijk te realiseren dat die werelden niet 'zomaar verondersteld' worden, maar *gededuceerd* worden als we de wiskundige structuur van de theorie heel serieus nemen—dat die wiskunde niet stiekem alleen maar een handig instrument is om voorspellingen te doen, maar ook op een of andere manier zou moeten corresponderen met de werkelijkheid om ons heen. Het is dus niet zozeer 'doen alsof' er heel veel meer werelden bestaan, maar het 'erachter komen' dat het universum veel groter is dan we dachten.

Deze laatste realisatie wordt vooral duidelijk als we een dynamisch effect beschouwen: decoherentie. Kleine interacties met de omgeving van een kwantumsysteem (de lucht, bijvoorbeeld) zorgen ervoor dat het typisch 'kwantummechanische' uitvloeit over die omgeving en zoveel 'verdund' wordt dat je het niet meer kunt waarnemen. Voor Everetts theorie heeft dat als gevolg dat zijn 'werelden' niet meer met elkaar in contact staan: er onstaan vele afgezondere werelden.

Uiteindelijk resulteerde (onder andere) de bovenstaande ideeën in een boek van David Wallace in 2012, dat (eindelijk) een geïntegreerde en overzichtelijke presentatie van de theorie vormt, aangevuld met een halve eeuw aan uitwerking, kritiek en heroverweging: *The Emergent Multiverse*. Eindelijk kan de veelwerelden-theorie eens goed worden bekritiseerd.

Deze scriptie bestaat eigenlijk uit drie delen: een wiskundige en conceptuele verkenning, een kritische afbrekende beschouwing en een drietal constructieve pogingen tot oplossingen. Allereerst beschrijf ik de veelwerelden-theorie zoals deze wordt gepresenteerd door Wallace, aangevuld met een herformulering van de belangrijke onderdelen op een manier zoals ik het aan mijzelf zou uitleggen. Vervolgens analyseren we het precieze wereldbeeld dat door Wallace naar voren wordt geschoven, wat voor implicaties dit wereldbeeld heeft wanneer we het als oplossing voor het meetprobleem accepteren. In het bijzonder bekijken we zijn gebruik van 'emergentie' (het idee dat in een complex systeem met vele onderdelen er iets verschijnt wat je aanvankelijk niet had verwacht als je de eigenschappen van de afzonderlijke onderdelen had bestudeerd) en het vraagstuk in hoeverre Wallace uitspraken doet over 'de realiteit' vanuit pragmatische of subjectieve of anthropocentrische hoek. Vooral het antwoord op die laaste vraag blijkt nog helemaal niet zo vanzelfsprekend te zijn en leidt tot een uiteenzetting over verschillende theorieën van 'verklaring': wat betekent het als we een verklaring zoeken voor een bepaald fenomeen en wat is er voor nodig om uiteindelijk tevreden te zijn met een verklaring? Dit grenst aan vele kernproblemen in de wetenschapsfilosofie: het realismedebat, de vraag hoe verschillende theorieën samenhangen en hoe we om moeten gaan met verschillende theorieën die over dezelfde entiteiten lijken te spreken maar in een andere taal. In het uiteenzetten van verschillende theorieën van verklaring, zoals die van filosofen Bas van Fraassen en Daniel Dennett, vinden we dat de vraag om verklaring tegelijkertijd een zoektocht naar betekenis is. Daarbij gaat het mij niet zozeer om de vragen 'of er wel betekenis is' of 'wat betekenis is' (al kunnen we dat gesprek ook nog weleens voeren), maar om de vraag *op welk niveau* we die betekenis dan wel niet moeten zoeken. Vinden we betekenis in biologische (of sociale) termen, in kosmische overwegingen, of toch in het gedrag van de fundamentele deeltjes van de natuurkunde?

Als derde deel van de scriptie werk ik een aantal kritiekpunten uit naar aanleiding van Wallace' werk. Hoewel hij claimt niets toe te voegen aan het formalisme van de kwantummechanica, laat ik zien hoe zijn visie afhankelijk is van wat men verstaat onder verklaring en realiteit. In mijn ogen lijkt hij toch specifieke filosofische standpunten binnen te smokkelen die niet altijd even gangbaar zijn: hij lijkt dus een uitgebreidere agenda te hebben dan het meetprobleem oplossen. Ik verhelder dit door een onderscheid te gebruiken tussen het meetprobleem en het 'ontologisch' probleem van de quantumfysica. Waar het meetprobleem onstaat door de toelating van ogenschijnlijk conflicterende regels (unitaire versus projectieve veranderingen) die de voortgang door de tijd bepalen van het centrale object in de wiskunde van de kwantumfysica, de golffunctie, richt het ontologisch probleem zich op de vraag hoe diezelfde wiskundige golffunctie samenhangt met wat er daadwerkelijk in de wereld bestaat. Wallace gebruikt hiervoor een criterium dat hij vernoemt naar Daniel Dennett, waarin hij een beroep doet op het begrip 'patronen', die in een abstracte wiskundige ruimte (de configuratieruimte) zich hetzelfde lijken te gedragen als de 'echte' dingen in de wereld. 'Dennetts criterium' is moeilijk te begrijpen en lijkt vele specifieke filosofische baggage te importeren. Mijn onderzoek gaat veelal uit naar het begrijpen ervan. Mijn kritiek – mits steekhoudend – lijkt strikt genomen geen probleem te vormen voor Wallace' oplossing van het meetprobleem, maar laat zien dat het probleem met de ontologie hardnekkiger is dan Wallace het presenteert en dat er meerdere posities in te nemen zijn dan de zijne. Sommige van die posities zijn moelijk te rijmen met Wallace' uitgangspunt, het bovengenoemde 'heel serieus nemen' van de wiskunde.

Tenslotte is het natuurlijk wel mogelijk om te speculeren over waar Dennetts criterium wel voor gebruikt kan worden, of uit welke overwegingen zo'n soort criterium wenselijk zou kunnen zijn. Het grootste deel van de scriptie is gewijd aan het schijnen van licht op deze zaak aan de hand van drie filosofische distincties, langs de assen 'objectief/subjectief', 'kwantitatief/kwalitatief' en 'werkelijkheid/kennis'. Ten eerste denk ik dat er een specifiek klassiek perspectief moet worden ingenomen voordat Dennetts criterium kan werken, dat ik de 'classical stance' noem. Alleen vanuit een subjectieve, menselijke, invalshoek springt het klassieke patroon in het oog, analoog aan Van Fraassens bewering dat er een menselijke invalshoek nodig is bij het vinden van saillante causale verklaringen. Ten tweede geeft Dennetts criterium volgens mij een manier om (naast het kwantitatief terugvinden van macroscopische structuren) die hogere structuren op kwalitatieve wijze overeen te laten stemmen met concepten die we kennen uit de klassieke wereld. Ik heb dit geïnterpreteerd als een vorm van kwalitatieve 'overbruggingsregels' tussen theorieën zoals geformuleerd door reductionisten als Ernest Nagel. Ten derde denk ik dat de acceptatie van Dennetts criterium een specieke kijk inneemt ten aanzien van de werkelijkheid (het bovengenoemde realisme-debat). Wallace' positie is minder lenig dan hij het laat toeschijnen. Daarentegen zou het een voordeel (voor Wallace) zijn dat dit een andere mogelijke versie van de theorie ondermijnt: de 'many-minds'-theorie, waarin kennis over werelden slechts illusies zijn binnen de breinen van observatoren. Voor Wallace zijn de klassieke werelden *echt* en kunnen daardoor niet worden beschouwd als illusies.

De besten vervallen wel eens in onduidelijkheden (of op z'n minst onvolledigheden). Het is aan de filosoof om vragen ter verheldering te stellen, en waar nodig kritiek uit te oefenen op een redenering die tekort lijkt te schieten—soms op het scherpst van de snede. Na vanaf de wal geroepen te hebben, is het naar mijn mening dan ook de taak van de filosoof om een poging te doen om zelf als stuurman op te treden en het schip (zo goed als het kan) binnen te loodsen. Ik zoek dit uiteindelijk in de behoefte om 'het klassieke' – dat wil zeggen het deterministische, complete en intuïtieve – te hervinden. Dat klassieke patroon is namelijk handiger om voorspellingen te doen en sluit vaak beter aan bij onze intuïties. Daarbij leent de taal van het klassieke zich stukken beter voor het geven van een verklaring dan de – probabilistische, onzekere en tegenintuïtieve – kwantumtaal lijkt te doen. Maar dat het 'handiger' is, is vooral een pragmatisch criterium dat veel met onszelf te maken heeft. Mijn conclusie is dan ook dat deze pragmatische gedachtegang niet uit zichzelf leidt tot een 'realiteit' zoals Wallace dat wil zien. De menselijke behoefte aan verklaring en voorspelling is misschien een criterium om het klassieke patroon in het oog te laten springen en misschien is dit klassieke perspectief wel het beste dat we ons kunnen veroorloven—maar het is niet voldoende om, los van de mens, een realiteit toe te kennen aan die klassieke structuur boven de andere patronen die er bestaan, namelijk de kwantummechanische patronen die ten grondslag liggen aan onze werkelijkheid.

Dan rest mij nog een aantal mensen te bedanken in dit afgelopen hectische jaar waarmee ik het studentenleven achter mij laat. Allereerst mijn begeleider Guido Bacciagaluppi, zelf ook in een hectische periode, voor het vele meedenken, het tijd vrijmaken, en het leveren van talloze scherpe verbeteringen, tips en kritische noten met een altijd vriendelijke knipoog; Ronnie Hermens, voor het toestemmen om mij te begeleiden in een project dat al een tijdje bezig was, en het intensief meedenken over Dennetts criterium en het systematisch zetten van wiskundige puntjes op de *i* van het begrip 'instantiatie'; Prof. Dennis Dieks, die mij voorzichtig aanraadde om op zoek te gaan naar een antropische redenering in Wallace's werk, die ik heb gevonden, maar die ik vooral wil bedanken voor het meedenken over mijn toekomst, het aandurven van het schrijven van een artikel met mij over gesloten tijdachtige krommen (dat is jargon voor tijdreizen), en voor het mij voor het eerst laten inzien dat er daadwerkelijk een evenwicht te vinden is op het snijvlak tussen natuurkunde en filosofie: een balans waar ik toen voor ben gegaan en waar ik mij op mijn plek heb gevoeld; Dr. Maaneli Derakhshani, voor de diepe gesprekken – tot in de te late uurtjes – over objectiviteit (bestaat niet, maar dat maakt niemand jou wijs), sociale ongelijkheid, en de logische gevolgen van Wallace' multiversum voor het menselijk bewustzijn; en Prof. Fred Muller, voor het hameren op de vraag wat het woordje 'quasi' nu eigenlijk doet.

Ik ben de docenten die HPS draaiende houden bijzonder dankbaar – het blijft één van de meest interdisciplinaire en diverse opleidingen die een student kan wensen – en dan vooral David, voor het bieden van de mogelijkheid om de opleiding als 'ambassadeur' te mogen representeren, en Daan, die het aandurfde om mij het werk van zijn studenten te laten nakijken bij zijn vakken 'de wetenschappelijke revolutie' en 'filosofie van de natuurwetenschap' (tevens het vak waar Dennis mijn interesse voor de grondslagen van de natuurkunde deed ontvlammen) en met wie ik tussen het lesgeven door enkele passies kon delen; en ook weer Guido, met wie ik verschillende praatjes heb mogen geven om 'reclame' te maken bij bèta-studenten die iets extra's zoeken. Dank aan Richelle Boone, die mijn stukken kritisch, aandachtig en in onmenselijk detail heeft becommentarieerd; Sam Rijken, met wie ik lief, leed en literatuur gedeeld heb; Noelia Iranzo Ribera, Sam, Daan van den Berg, Sander Kooi en Nick Wiggershaus voor het bekritiseren van mijn argumenten en mijn Engels; maar ook de anderen in de 'Rendierkamer': Roelof, Marte, Anne, Nijs, Remco, met wie ik tijdens de afgelopen twee jaar samen heb gelezen, geschreven en geleden; aan mijn twee voormalig natuurkundegenoten Sander^2, die mijn tirades over de grondslagen van de natuurkunde serieus proberen te nemen; aan Piet, die mij in vorm houdt door me te blijven uitdagen voor squashwedstrijden; aan mijn Tritonese 'roei'-ploegje, Mike, Geert, Bram, Thom en Frits, voor de vele etentjes en gesprekken over identiteit en politiek; aan Roland, Sander, Richelle, Radnev, Maxime, Boyd, Daan, Sam en Lani, die luisterende oren schonken als het even aanpoten was; dank aan de inmiddels vierenvijftig IBB-huisgenoten van de afgelopen zeven jaar voor een altijd inspirerende, volwassen-makende en zeker ook frusterende leefomgeving; dank aan mijn broer Romboud, niet tenminste voor de Nespressomachine die ik op precies het juiste moment kreeg (zonder was het me niet gelukt); en tenslotte wil ik ook mijn vader, Gerard Mulder, bedanken, die misschien niet weet dat ik nog heel goed weet dat hij mij – nog voor ik lange zinnen kon lezen – enthousiast kreeg voor Plato's Grot.

—Ruward,

Utrecht, voorjaar 2018

Contents

1 Introduction

The challenge of quantum mechanics that we find ourselves facing is one where the mathematical formalism, which has proven to be so successful to predict what happens at the sub-atomic level, makes predictions about our macroscopic realm that utterly conflict with the things we see around us. Although the things around us are in some sense supposed to be build out of stuff at that sub-atomic level, somehow macroscopic objects do not behave like that stuff. There is also no reason to believe that the problem disappears when we go to lower-level theories such as quantum field theories, string theory, semi-classical gravity or its fully quantum version, since these are all quantum theories built on similar principles. It is therefore also a problem that will plague the unified theory of physics that many dream of.

This problem is unlike the practical problems of physics, where calculations still need to be done or where novel behaviour needs to be accounted for by the theories at hand (or extensions thereof). This 'measurement problem' is rather problematic at the level of logical consistency; it is a place where physics is in dire need of the conceptual analysis of the kind that is more often found in the humanities department. It requires an interdisciplinary approach that borrows from philosophy to supplement physics. The measurement problem, hence, is one of the central concerns for 'foundational physicists'.

It is easy to lose count of all the ideas that have been proposed to solve the problem. We have come a long way since the early concerns of Einstein, Bohr, de Broglie, Schrödinger and many others, and another long way since the clear objections raised by those like Bohm, Everett and Bell. The resulting anarchy of mutually-competing, measurement-problem-free quantum mechanics has generated theories[1] that are worked out in staggering detail. This thesis is about one of those theories, one where the quantum formalism is not seen as a problem, but where the quantum state is taken as literally as possible: David Wallace's version of Everettian quantum mechanics.

1.1 The 'measurement problem' and the modern Everettians

Let us expand a bit more on the quantum measurement problem. In quantum mechanics, a quantum state is represented by a (normalizable) vector in Hilbert space. Under normal circumstances, it evolves unitarily under Schrödinger's equation: the dynamics is continuous, information-preserving, and deterministic. But Hilbert space is not the world we observe. To extract information about the physics described by the quantum state one performs 'measurements'. However, it is not at all clear what classifies as one. A measuring apparatus itself supposedly consists of particles. But when we describe

[1] I use the word 'theories' instead of the rather vague term 'interpretations' that is usually used. The reason is that sometimes an assumption is made that the bare mathematical formalism is 'the theory' and that the interpretation is somehow secondary. This is mistaken. Mathematical formalism is just that: formalism. And formalism by itself has no relation to the real world. A theory in physics is the formalism plus rules of correspondence that connect (parts of) that formalism to things that we suppose exist in the world. I expand a bit more on this in Section 2.1.

the apparatus with the quantum formalism it does not supply us (in general) with a unique outcome. It supplies us with a superposition of outcomes. This conflicts with the definite outcomes that we experience.[2] How is it that when we look at our world we see objects with definite properties, cleansed of superposition? The initial approach to the problem has been to solve it by fiat: assume a 'projection postulate' that forbids the superposition from boiling upwards towards the macroscopic level. This is a rather *ad hoc* solution: what explains this collapse of the wavefunction when it crosses from the microscopic to the macroscopic realm? That, in a nutshell, is the measurement problem: the difficulty arising from trying to account for a classical world arising from a fundamentally quantum-mechanical picture of the world.

Two strategies to solve the puzzle have usually been distinguished: *(i)* add non-linear terms to the Schrödinger equation that will lead to a dynamical mechanism that explains how collapse occurs, i.e. add extra *dynamics*; or *(ii)* add extra mathematical structure or properties to quantities in the quantum system, such that this extra structure leads to definite values, i.e. add *hidden variables*.

In 1957, Hugh Everett III proposed an unexpected third strategy: denying that the macroscopic world is indeed free of superpositions, that it only *appears* to be definite.[3] Our own perception is then said to be itself contained in one branch of a superposed universal wavefunction[4] that describes everything in the universe; and these branches are effectively separate from each other. Seen from within such a branch one observes a classical world relative to the observer, while in reality there a many such experiencers of classical worlds. From the 1970's onward, this idea gained momentum when combined with another unexpected insight, formulated by Hans-Dieter Zeh: decoherence. As will be extensively discussed in Chapter 3, quantum systems lose coherence (a measure for the entanglement between degrees of freedom in a system) when interacting with an environment.[5] David Wallace's emergence of classical worlds finds itself at the contemporary pinnacle of the school of thought that combines decoherence with Everettian insights.

Wallace presented the world with the culmination of his work on Everettian quantum mechanics in 2012: *The Emergent Multiverse*.[6] Following Everettian insights, he claims that the 'measurement problem' (now in scare quotes) is only apparently problematic. Being an Oxford Everettian, he advocates taking the formalism of quantum theory at

[2]Neither does it provide us with an explanation of the statistics of those outcomes. Hence, one has to account for the Born rule, which states that the probability of a definite outcome is given by the absolute square of the length of the basis vector that is supposed to correspond to that definite outcome.

[3]Hugh Everett, III, "The Theory of the Universal Wavefunction," Princeton PhD dissertation (1957).[34]

[4]The wavefunction is – strictly speaking – a representation of the quantum state in a particular Hilbert space, but often 'wavefunction' is used as synonymously with 'state'.

[5]Hans-Dieter Zeh, "On the Interpretation of Measurement in Quantum Theory," in *Foundations of Physics* **1**, pp. 69–76 (1970).[87]

[6]David Wallace, *The Emergent Multiverse* (Oxford: Oxford University Press, 2012).[82]

face value and interpreting it realistically. My sketch of his strategy for solving the measurement problem is as follows. The fundamental ontology of the theory is given by the quantum state, permanently subject to Schrödinger's unitary evolution. No wavefunction collapse occurs, as each term in a superposed state survives interactions with other physical systems, whether they are electrons, measurement devices, or human eyes—all are equally described by quantum states. Each term of the superposition becomes itself correlated with the systems it is interacting with and becomes a separate 'branch' of its own. The branches are kept from interfering with each other due to decoherence, the dynamical suppression of interference between branches in a particular basis.[7] Furthermore, the patterns of such a branch 'functionally' (a term I will come back to later) resemble the patterns of a classical 'world' with definite outcomes like the one that we experience.

In addition to solving the problem of definite outcomes in Part I, Wallace devotes Part II of the book arguing that the relative branch weights of quantum states are to be interpreted as subjective probabilities that obey the Born measure. This can be done via the decision-theoretic (or Bayesian[8]) approach to probabilities—despite the fundamental determinism of the theory. The idea is that a rational agent who makes bets (and has good knowledge of Everettian quantum mechanics) will always use the Born rule to ensure the best outcome of her bet, because she knows that the relative weights correspond to a measure of the worlds that split off in the future will resemble her own world. This decision-theoretic approach has been given a rigorous gloss through various proofs by Wallace and David Deutsch.[9] Each outcome has a subjective utility measure, which agents use to try and maximize the expected utility (weighted average over utilities) as much as possible, according to the Principle of Maximization of Expected Utility.

To summarize, to interpret the universal wavefunction as fundamental, while observing definite outcomes at the macroscopic level is the first goal of the Everettian agenda. The second goal is to solve a problem that follows from it: to explain that the weights of the decohered branches of the universal wavefunction can have a probabilistic interpretation in accord with quantum experiments, i.e. to reproduce the Born rule. Assume the second goal to be achieved, this research focuses on Wallace's claim to have achieved

[7] Better put: as a quantum system is – in practice – never closed, decoherence is the delocalization of the phase relations of a subsystem over the *environment*, seen from the coarse-grained position or joint position-momentum basis. This plays the role of fulfilling Everett's 'no-interference criterion', but see Section 3.2.

[8] I will stick with the terminology 'decision-theoretic', as Wallace and Deutsch themselves also do. An upshot of that can be that the word 'Bayesian' has connotations with the quantum-Bayesian school, which in fact seems less Bayesian than Wallace's approach.

[9] David Deutsch, "Quantum theory of probability and decisions," *Proceedings of the Royal Society of London* (1999), Vol. **A455**, pp. 3129–3137.[26] David Wallace, "Quantum probability from subjective likelihood: improving on Deutsch's proof of the probability rule," *Studies in the History and Philosophy of Modern Physics* (2007), Vol. **38**, pp. 311–332.[78]

the first.[10]

The recognizing of classical worlds plays a crucial part in Wallace's modern development of the many-worlds interpretation,[11] as it brings about a set of single worlds that by themselves are effectively *autonomous*. These quantum states that describe these worlds share one large Hilbert space, but nonetheless individually behave as if they are in a definite classical state. Such an individual world should be understood as we understand our classical universe: it includes stars and galaxies and all the other macroscopic objects like measurement devices, mountains, cities, the Earth, and ourselves. In such a way, there *emerges* a multitude of non-interacting worlds from the underlying quantum picture; it is indeed an emergent multiverse. Throughout this work, I will follow the above set of claims in detail, and critically evaluate its components; with special emphasis on *(i)* the role of emergence, and *(ii)* the functional view of emergent classicality.

1.2 This Thesis

Although Wallace goes to great lengths in an attempt to underpin his concept of *emergence*, it is rather difficult to see to what extent these 'classical worlds' are in fact emergent properties of an underlying universal wavefunction. The term 'emergence' is used in many fields of science and philosophy (and is often bandied about). Usually it is used as an umbrella-term which covers several underlying concepts, the conflation of which is a great source of confusion. What these applications of the term have in common, though, is a notion that some high-level phenomena (or whole) is brought about by lower-level constituents (or parts), in such a way that the higher-level phenomenon seems greater than the sum of its parts or behaves in a novel way with respect to what one might expect from the constituents beforehand.

To shed some light on this, we will use a distinction (by David Chalmers) between 'weak' emergence and 'strong' emergence. In cases of weak emergence high-level phenomena are unexpected and autonomous. Strongly emergent phenomena occur when high-level truths are not necessitated by low-level truths. We will discuss this more thoroughly in Chapter 4 in order to find out where Wallace's analysis fits in. Hence,

[10] This second goal simply falls beyond the scope of this research. That is not to say that Everett's typicality-approach has not been criticized extensively, *cf.* Neill Graham's, "The measurement of relative frequency," for one of the earliest objections to the feasibility of this second goal (although Graham himself became convinced of Everett's ideas).[37] Likewise, Deutsch and Wallace's decision-theoretic approach has been questioned heavily, *cf.* Adrian Kent, "One world versus many: The inadequacy of Everettian accounts of evolution, probability, and scientific confirmation," *Many Worlds? Everett, Quantum Theory, and Reality* (Oxford: Oxford University Press, 2010), S. Saunders, J. Barrett, D. Wallace, A. Kent (eds.), pp. 307–355;[51] and Richard Dawid & Karim Thébault, "Against the Empirical Viability of the Deutsch-Wallace-Everett Approach to Quantum Mechanics," *Studies in History and Philosophy of Science Part B* (2014), Vol. **47**, pp. 55–61.[24]

[11] Wallace himself is also keen on emphasizing that this is not an *interpretation*, but rather "just quantum mechanics itself, read literally, straightforwardly—naively, if you will—as a direct description of the physical world, just like any other microphysical theory."[82, p. 2]

the first goal of this research will be to answer the question: *in what precise sense of emergence (weak or strong) are Wallace's worlds emergent?*

What Wallace dubs 'classical worlds' are, at least at first sight, not really the classical worlds as described by classical mechanics. This will be clearly outlined in Chapter 3, but what is basically done is the following. On the one hand, one can see in the ($3N$-dimensional) configuration space that (in a suitable basis) a 'pattern' arises from a single branch of the universal wavefunction. This pattern is said to have the property 'quasi-classical', because it approximately resembles the classical trajectories that would arise in the configuration space if one started out with an N-particle system in a (3-dimensional) classical space. The Wallacian turn is to 'functionally identify' these two patterns. This functional identification is done via what Wallace calls 'Dennett's criterion'.

Dennett's criterion specifies the recognition of some higher-level phenomenon as a specific pattern of lower-level elements on the basis of pragmatic virtues as explanatory power and predictive reliability. Wallace formulates this criterion as follows:[12]

> A macro-object is a pattern, and the existence of a pattern as a real thing depends on the usefulness – in particular, the explanatory power and predictive reliability – of theories which admit that pattern in their ontology.

The second part of this thesis (Chapters 5—8)is essentially about finding out precisely what is meant by this criterion and what it entails. Although without this criterion the sense of emergence that I am searching for is almost trivially of the weak kind, the use of Dennett's criterion obscures this conclusion and makes the specific sense of emergence more cumbersome. The work that Dennett's criterion is supposed to do in Wallace's solution of the measurement problem is this: to move from quasi-classical worlds to real classical worlds. It is clear that to match prediction of the quantum theory with predictions of the classical theory, one needs to spell out how higher-level theories and lower-level ones relate. *Prima facie*, however, it is not clear why notions like 'pattern' and 'usefulness' are necessary to interpret macroscopic objects as 'real'. Also, when usefulness is explicated as the pragmatic virtues explanatory power and predictive reliability (which are quite different things), the question rises how this is connected to the concept of 'reality' in the first place. Three themes, then, are distilled from this discussion: pragmatics, intertheoretic relations, and reality.

1.3 Outline of the argument

At the end of section on "Instantiation and the relation between theories" Wallace makes clear that he does not intend to fully work out his appeal to Dennett's criterion:[13]

[12] Wallace, *Emergent Multiverse*, p. 50.
[13] *Ibid.*, p. 58.

> These details [of 'explanatory usefulness'], however, are not crucial for our purposes. This is not a book about the philosophy of emergence: it is a book about the measurement problem.

This is, keeping in mind the scope of his research, a very reasonable decision. But Wallace also found it reasonable to discuss a part of the philosophy of emergence, to serve his purpose of claiming that the quasi-classical patterns are *real* classical worlds. Does he suggest that the solution to the measurement problem is independent of one's specific take on what we mean by reality, by emergence, by usefulness, and the particular way in which we understand the theories in science hang together?

The Emergent Multiverse remains stuck for an answer, but, fortunately, in a paper where Dennett's criterion also plays a central role, Wallace encourages one to formulate ideas:[14]

> Readers will, I suspect, be better served by forming their own criticisms, and seeking them elsewhere, than by any imperfect attempt of mine to pre-empt criticisms.

In this research I have formulated my own criticisms, which I outline as follows. In Chapter 5, I look at Daniel Dennett's seminal paper "Real Patterns" to improve the grasp on the notion of 'pattern' and the specific from of functionalism that Dennett employs. I then contrast Dennett's functionalism with Wallace's worldly patterns. Although the similarities are many, the use of the words 'pattern' and 'theory' are employed with different goals.

Next, I will investigate Wallace's Dennett's criterion to the letter as well as to its intent, conjecturing ideas that might benefit Wallace's program (while also spelling out the costs of these conjectures). I want to emphasize the three topics identified above – pragmatics, unity, reality – which I will use as the three different angles to approach Dennett's criterion, each of which will have a corresponding philosophical distinction. Again, Wallace's presentation of functionalism blurs all these lines, but that is why it is important in the first place to look at Dennett's criterion through the lens of different distinctions. The three distinction to be used are:

- Chapter 6: **Pragmatics ~ the objective/subjective-distinction**. In solving (or dissolving) the measurement problem, it is important to understand the pragmatic factors (the explanatory power and predictive reliability) on which the solution relies. In this way we can specify the context in which the pragmatic arguments hold and in which they might fail. *Through which virtues can we recognize the pattern of a 'world' where, in principle, multiple patterns could be recognized?* To find a satisfying answer to this question, I will draw an analogy

[14] David Wallace, "The Everett Interpretation," *The Oxford Handbook of Philosophy of Physics*, pp. 460–488, Robert Batterman (ed.) (Oxford: Oxford University Press, 2013), p. 2.[83] I confess this a general remark by Wallace, made in the introduction of his book, outside of the specific context of Dennett's criterion.

with van Fraassen's pragmatic theory of explanation, where a certain causal pattern becomes salient in the context of specific questions for explanation.[15] I will argue that in the objective sense the quasi-classical patterns are just as real as any other pattern within the universal wavefunction, but that at the coarse-grained level they can be said to be 'more real' when seen from the classical basis (which approximately coincides with the decoherence basis). This salience is subjective because one needs to stipulate a preferred basis before dynamical considerations apply. Similar to Dennett's 'intentional stance', I claim a specific 'classical stance', a thoroughly *human* view, is needed before Dennett's criterion can find traction.

- Chapter 7: **Unity ~ the quantitative/qualitative-distinction**. In relating higher-level theories to lower-level theories two general features need to match. First, there is the quantitative link, where predictions of the reduced theory should also be predictions of the reducing theory in order to save the phenomena. Second, concepts of the higher-level theory should be translated into the language of the lower-level theory. *How does Wallace relate (the configuration-space trajectories of) the classical theory to quantum mechanics, both at the quantitative and qualitative levels?* Although Wallace's version is much weaker, I draw an analogy with Ernest Nagel's reductionist program, where next to a quantitative deduction also conceptual 'bridge principles' are needed to reduce one theory to the next. Furthermore, following Michael Berry and Robert Batterman's analyses of singular limits of theories, I will investigate if any special emergent phenomena are to be expected for quantitative reasons in the infinite-time limit.

- Chapter 8: **Reality ~ the real/illusion-distinction**. Macro-objects that are said to 'exist' are recognized as useful patterns within the configuration-space representation of the universal wavefunction, but there are many patterns to be found there that should not correspond to 'real' macro-objects. The 'unreal' macro-objects should correspond to patterns of useless theories. *What exactly does Wallace mean by 'reality through usefulness' and what kind of work does this do in his solution to the measurement problem?* I will claim that through Dennett's criterion Wallace imports specific philosophical baggage that takes a specific stance within the scientific realism debate, akin to structural realism or explanationism. Further, I spell out a solution to the measurement problem, the many-minds theory, which solves the problem along the same lines as *The Emergent Multiverse*, with the exception that – although the quantum world itself is real – the classical worlds with definite properties are beliefs in the superposed brains of observers.

[15] The pragmatic theory of explanation serves to underpin van Fraassen's antirealist agenda, but we do not follow that agenda here. Realists often argue that only the realist's position provides an *explanation* of phenomena. Van Fraassen's constructive empiricism disentangles explanatory power from the core of science, and instead builds up a theory of explanation based on contextual factors; explanation becomes a pragmatic issue. The (anti-)realism debate falls outside the scope of this research, but we do not need to be constructive empiricists to *use* van Fraassen's theory of explanation.

2 The quantum puzzle

"Rampant linquistic confusion may contribute to [the undecidedness of interpretation]. It is not uncommon for two physicists who say that they subscribe to the Copenhagen interpretation [...] to find themselves disagreeing about what they mean."

— Max Tegmark and John Archibald Wheeler.[16]

This chapter is meant to highlight the origin of the measurement problem. Also, I will sketch the three strategies that can be followed to solve the problem: spontaneous collapse theories, hidden-variables solutions, and Everett's literal approach to the formalism. Naturally, special emphasis will lie on the third solution, as it is the foundation on which Wallace builds his quantum theory. Before we get to that, a note on the idea of a scientific theory is in order.

2.1 Scientific theories and their relation to the world

Any theory necessarily makes assumptions. It posits some starting points, called the fundamental ontology, from which further statements are supposed to follow (the derivative ontology). These ontological entities are supposed to be connected to things that are there in the world. A good theory needs to specify these connections, since otherwise it would only be a structure of statements or theorems in and of itself, unrelated to actual events. Theories of physics are of a quantitative mathematical nature, which makes the starting point clear: they posit a set of formal axioms which is able to embed the fundamental ontology.[17] A theory is then supposed to give correspondence rules for relating (part of) the formalism to the physical entities that we think should be described by this formalism: the observables.

Not every single part of the formalism needs to be connected to something in the world. It is not at all rare that the formalism contains 'mathematical fictions' that no one (not even the realist) wants to relate to the physical. A good example is the gauge freedom that is present in field theories, which are useful for making calculations, but are not regarded as relating to something that exists in the world.[18]

Quantum mechanics has posed a problem to the above distinctions, not because it cannot be caught in those terms but because it can be caught in those terms in many

[16] Max Tegmark and John Archibald Wheeler, "100 Years of Quantum Mysteries," *Scientific American* 2001: **284**, p. 68-75.[65]

[17] If you are a structuralist, you will say the ontology is provided by the mathematical structures used in the axioms; if you are convinced by the indispensability argument, you will say that the mathematical objects that are indispensable to the formulation of the theory are ontological; if you follow Quine, you will say that the ontology is given by the variables you quantify over; and if you are a primitive ontologist, postulation of the fundamental ontology goes hand in hand with the axioms.

[18] Although it can be made plausible to promote gauge degrees of freedom from mathematical fiction to fundamental ontology. An example is to promote the electromagnetic potential to physically real 'potential fields' (at the expense of the electromagnetic fields, which then becomes derivative ontology) to obtain a local explanation of the Aharonov-Bohm effect in semi-classical electrodynamics.

ways. It is not so clear how to relate the ontology of the theory to the world. One idea would be to regard the quantum state (which appears to be the most straightforward candidate for the fundamental ontology of the theory) as something that directly exists. This is what Schrödinger attempted to do. However, there is a mismatch between the Hilbert space of the quantum state and the real space we live in, as Schrödinger admitted in his "Wave mechanics" about his own theory (which he would very much have loved to interpret as direct physical waves in spacetime):[19]

> Of course this use of the q-space is to be seen only as a mathematical tool, as it is often applied also in the old mechanics; ultimately, in this version also, the process to be described is one in space and time. In truth, however, a complete unification of the two conceptions has not yet been achieved. Anything over and above the motion of a single electron could be treated so far only in the multi-dimensional version; [...].

One can also accept the wavefunction as a mathematical fiction, a tool for calculating the expectation values of Hermitian operators, where these (expectation values of) operators are supposed to represent the physical things in the world. However, this puts question marks around the idea that the electron is in a certain state given by the wavefunction. In the Copenhagen conception of the theory, the idea that the electron is in a certain state is indeed rejected. Other quantum theories (those discussed later in this chapter) are theories that take various positions about what to see as the fundamental ontology, what to see as derivative and what to see as mathematical fiction.

The Copenhagen theory (alhough it is certainly not a homogeneous approach[20]) takes the view that we cannot do without the concepts of the classical world, since these determine how we think (about the physical world). We should just describe things with the mathematical theory we have, and, prior to measurement, we have no words to describe what the state of a system looks like and we will fail in trying to do so. This 'Heisenberg cut' is a conceptual split between the quantum world and the classical world. It is a split between object and instrument. It is fine to treat a measuring device quantum mechanically, at least in principle. But to preserve the ability of observation, we have to *think of* the apparatus as if it is not entangled.[21] It is noteworthy that the classical world is here seen as prior to the quantum world.[22]

[19] As quoted in Guido Bacciagaluppi & Antony Valentini, *Quantum Theory at the Crossroads* (New York: Cambridge University Press, 2009), p. 406.[6]

[20] The Copenhagen interpretation is most commonly attributed to Bohr and Heisenberg. Don Howard raised the issue, though, that it is hard to really trace some of the arguments back to the person who first contrived it or even to trace it back to one period in time and that the 'Copenhagen interpretation' is a vague construct of mostly Bohr's and Heisenberg's arguments and later philosophical developments and characterizations by Popper and Feyerabend, combined with a 'shut-up-and-calculate'-mentality in the phyics community. See "Who Invented the "Copenhagen Interpretation"? A Study in Mythology," *Philosophy of Science* (2002), **71**, No. 5, pp. 669–682.[45]

[21] Based on the Portland lecture "The Decoherence Prism," by Maximillian Schlossauer at the QFQI Conference, Decoherence & Friends, May 2013.

[22] Guido Bacciagaluppi, "The Role of Decoherence in Quantum Mechanics," *The Stanford Encyclo-*

2.2 The measurement problem: formalism versus experience

In the orthodox approach to quantum mechanics (say, the axiomatized versions of John von Neumann and Paul Dirac[23]) there are two rules of evolution. On the one hand, the Hilbert space vectors of quantum mechanics evolve unitarily under the continuous, information-preserving, and deterministic Schrödinger dynamics. On the other, when 'measuring' a quantum system the dynamics is discontinuous, it loses information, and is indeterministic. In the case of a measurement, the non-unitary projection or collapse of the wavefunction takes over from the usual dynamics. It can immediately be seen that these two rules contradict each other, at least without a further exact specification of what is meant by the stipulation 'measurement'.

In particular, though, one of the things that should be explained is how 'measurement' is a different kind of process than any other physical process. Another problem one can anticipate is how it is that the 'collapse' is to be reconciled with special relativity, since it breaks the Lorentz invariance of measurement outcomes between two observers in some situations, e.g. measurements on EPR pairs. One way to understand the measurement problem is to proceed without the collapse postulate and see what kind of conflicts arise.

Consider a quantum system $\mathcal{S}$ in state $|\Psi\rangle = \sum_i c_i |\psi_i\rangle$, superposed in an orthonormal basis set $\{|\psi_i\rangle\}$ with complex coefficients c_i and in Hilbert space $\mathcal{H}_\mathcal{S}$. We wish to measure this state using an apparatus $\mathcal{A}$ which is also described with a quantum state in Hilbert space $\mathcal{H}_\mathcal{A}$, where the basis is chosen such that the basis states $|a_i\rangle$ correspond to different 'pointer states'. This can be imagined as pointing at a certain physical value on a scale; the pointer states $|a_i\rangle$ are then automatically mutually orthogonal, since pointing at one number is clearly distinguishable from pointing at another. The two systems $\mathcal{S}$ and $\mathcal{A}$ interact with each other through an interaction Hamiltonian, and the combined system is described in Hilbert space $\mathcal{H}_S \otimes \mathcal{H}_A$. If the measurement device is properly designed for the observable under consideration, $|\psi_i\rangle \otimes |a_0\rangle$ should evolve to $|\psi_i\rangle \otimes |a_i\rangle$.[24] From linearity, it follows that

$$\sum_i c_i |\psi_i\rangle \otimes |a_0\rangle \to \sum_i c_i |\psi_i\rangle \otimes |a_i\rangle \,, \tag{2.1}$$

which shows that the measuring process is then simply the entangling of the two systems. Clearly, every term in the original superposition is coupled, via the interaction Hamiltonian, to the measurement apparatus' pointer states, which in turn becomes

pedia of Philosophy, ed.: Edward Zalta, `http://plato.stanford.edu/archives/win2012/entries/qm-decoherence/`.[7]

[23] Paul Dirac, *The principles of quantum mechanics* (Oxford: Oxford University Press, 1930).[29] John von Neumann, *Mathematical foundations of quantum mechanics* (Princeton: Princeton University Press, 1955).[75]

[24] This presupposes that the $\{|\psi_i\rangle\}$ have been chosen in such a way that the measurement device measures $|\Psi\rangle$ in this basis. Alternatively, one can safely redefine the $\{|\psi_i\rangle\}$ to meet that end.

superposed in the same basis. Hence, without collapse, we are left with a superposition of (possibly macroscopic) pointer states, utterly in conflict with actual pointers that we observe when measuring, say, the spin of an electron. This is one way to understand the measurement problem: how can we reconcile the macroscopic superposition of Eq. (2.1) with the experience of observing that state of affairs that corresponds to only one of the $|a_i\rangle$?

It does not help to take into account another measurement apparatus $\mathcal{A}^{(1)}$ (with ready pointer state $|a_0^{(1)}\rangle$ and corresponding Hilbert space $\mathcal{H}_{A^{(1)}}$) that measures the state of Eq. (2.1). Including this secondary measuring device just proves that the superposition of system $\mathcal{S}$ is contagious:[25]

$$\sum_i c_i |\psi_i\rangle |a_0^{(0)}\rangle |a_0^{(1)}\rangle \to \sum_i c_i |\psi_i\rangle |a_i^{(0)}\rangle |a_0^{(1)}\rangle \to \sum_i c_i |\psi_i\rangle |a_i^{(0)}\rangle |a_i^{(1)}\rangle , \tag{2.2}$$

where each $\to$ represents a specific step in the temporal evolution.[26] Continuing along this path, a third measuring device $\mathcal{A}^{(2)}$ would couple in the same way, and so on for n measuring devices. This is called the 'von Neumann chain',

$$\sum_i c_i |\psi_i\rangle |a_0^{(0)}\rangle |a_0^{(1)}\rangle |a_0^{(2)}\rangle \dots |a_0^{(n)}\rangle \to \dots \to \sum_i c_i |\psi_i\rangle |a_i^{(0)}\rangle |a_i^{(1)}\rangle |a_i^{(2)}\rangle \dots |a_i^{(n)}\rangle . \tag{2.3}$$

This is where one ends up without invoking the projection postulate after the measurement by $\mathcal{A}^{(0)}$.[27]An additional move is required to make this 'projection-free' non-collapse theory consistent with experience. Solving the measurement problem from here on falls under one of two ways:

1. Collapse solutions: Some mechanism for the projection of the state exists.
   Challenge: how do we justify a suitable termination of the von Neumann chain and at what level?

2. Non-collapse solutions: Projection is not fundamental.
   Challenge: how can Eq. (2.3) account for unique macroscopic outcomes?

While not very original, it is helpful to frame the measurement problem in terms of Schrödinger's cat thought experiment, since this sets a stage for critics to formulate (and agree on the formulation of) a few fundamental points of confusion of the quantum formalism.

Imagine we have a closed system consisting of an initial (macroscopic) state of a cat, $|$  $\rangle$, and a state of a two-level system, say an electron, superposed in the spin-basis,

[25] I drop the direct products $\otimes$ for brevity.

[26] The details of the interaction Hamiltonian depend on the specific kind of measuring devices that are employed, but these details are of no concern to the general treatment of entanglement between the measuring devices and the system.

[27] Note that the quantum statistics is not affected, since according to the Born rule, the probability of finding the particular value $a_p^{(1)}$ when making a measurement with $\mathcal{A}^{(1)}$ is $|c_p|^2$.

where the 'up'-state is $|\uparrow\rangle$, and the 'down'-state is $|\downarrow\rangle$. A mechanism is set in motion that interacts with the electron, i.e. becomes entangled with it, and, via intermediate interactions, becomes entangled with a measuring apparatus. A measurement result of 'spin up' $|\uparrow\rangle$ will leave the system invariant, while a 'spin down' $|\downarrow\rangle$ result will trigger the breaking of a cyanide flask, such that the cat state will quickly evolve to a 'dead cat state' $|$$\rangle$ after a time t_M. Thus, initially, for $t < t_M$, the entire system is in the state

$(c_\uparrow |\uparrow\rangle + c_\downarrow |\downarrow\rangle) \otimes |$  $\rangle\,.$ (2.4)

This is a definite state at the macroscopic level, since there is just one form the cat will have, namely $|$  $\rangle$. But, after t_M, due to the interaction between the two-level system, the cyanide, and the cat, the total system has evolved to

$c_\uparrow |\uparrow\rangle \otimes |$  $\rangle + c_\downarrow |\downarrow\rangle \otimes |$$\rangle\,.$ (2.5)

This state is indeed macroscopically *in*definite, because there are two terms, which each represent a different state of macroscopic affairs. Since this is something we never experience, this is called the problem of definite outcomes, which was Schrödinger's original 1935 objection to the completeness of quantum mechanics—and the intuition is sharply teased out by designing the thought experiment with a living cat.

2.3 The resulting foundational anarchy

For completeness, I want to summarize the three specific strategies that can be followed to solve the measurement problem. As indicated in the above, the problem boils down to the fact that we do not observe macroscopic superpositions, while these should be there according to the formalism. Hence, a natural approach is to see if one can change that formalism in a way as to leave the predictions at the quantum domain intact, while modifying the predictions for the macroscopic level. There are basically two ways to change the formalism: tweak the dynamics or add fundamental ontology. The former strategy relies on building a non-linear collapse mechanism into the Schrödinger equation. The latter approach is somewhat broader and relies on hypothesizing a deeper theory (with or without a collapse mechanism) from which the quantum formalism can be derived.

2.3.1 Solution 1: change the dynamics

The (orthodox) textbook quantum theory, as set out by von Neumann and Dirac, is an explicit collapse theory due to the projection postulate. As argued before, this projection postulate is of a rather *ad hoc* nature. That is not to say that projection itself should be discarded; it is rather that it would be preferable if this non-unitary behaviour can be explained by virtue of some principle or mechanism, which consistently puts it on a par with the Schrödinger equation. It is also important to show at what stage the von Neumann chain terminates. A straightforward way to proceed, then, is to modify the Schödinger equation with a non-unitary term.

The most famous is the GRW-theory, developed by the Italian physicists Ghirardi, Rimini and Weber.[28] They proposed to adjust the Schrödinger equation with a term that produced "spontaneous collapse". Every quantum system spontaneously undergoes collapse every once in a while—something like $10^{15}/N$ seconds for a N-particle system.[29] Since agglomerates of quantum systems are entangled systems, the collapse of one subsystem will drag the whole system with it. Hence, one retains superpositions at the quantum level, but one also has a mechanism that can explain why these superpositions are very unlikely to spread to the macroscopic level. Since everyday objects are entangled systems of a very large number of subsystems, the whole is continuously collapsing and thus never observed in a superposition.

One might be inclined to regard the collapse mechanism as something in need of explanation as well, especially if one is keen on a fundamental unitary behaviour of nature. Nevertheless, spontaneous collapse theories such as GRW do solve the logical conflict between a dual unitary and non-unitary behaviour that is inherent to orthodox quantum theory. A further advantage is that these interpretations are testable. Although the parameter space of the GRW is rather lean (and has been adjusted in response to experiments) it will shrink with technological improvements such that one day (not so far in the future) the theory can be ruled out or found to be empirically adequate.

2.3.2 Solution 2: add variables

Quantum mechanics has proven successful at the experimental level and hence it might seem reasonable to regard the quantum state as a complete description of the state of affairs. Nevertheless, assuming the quantum state is incomplete, one can posit extra features, so-called hidden variables, to complement it. If this route is taken the measurement problem can also be solved, by interpreting the quantum statistics as encoding our ignorance about the underlying level. The wavefunction is then to be regarded as referring to an ensemble of underlying systems from which the quantum formalism emerges. Also, the quantum statistics should then arise naturally from the incompleteness of the description at the quantum level, similar to probabilities in statistical physics.[30]

There are several hidden-variable proposals, of which the most famous was developed by David Bohm in 1952.[31] Note that the hidden-variable strategy will need

[28] Giancarlo Ghirardi, Alberto Rimini and Tullio Weber, "Unified dynamics for microscopic and macroscopic systems," *Physical Review* (1986), Vol. **34**, pp. 470–491.[36]

[29] David Albert, "On the Collapse of the Wave-Function," in *Sixty-Two Years of Uncertainty* (New York: Plenum Press, 1990), Arthur Miller (ed.), pp. 153–165.[3]

[30] This strategy is not countered by von Neumann's impossibility proof. von Neumann only proves that the ensemble interpretation is not possible *on the premise that the quantum description is complete.*

[31] David Bohm, "A Suggested Interpretation of the Quantum Theory in terms of "Hidden" Variables. I" (& II), *Physical Review* (1952), Vol. **85**, No. 2, pp. 166–179 and pp. 180–193.[15]

to circumvent several no-go theorems, notably those by Bell and Kochen and Specker. Bohmian mechanics and several stochastic hidden variable theories do circumvent these theorems; the same goes for the superdeterministic theory of cellular automata by Gerard 't Hooft.[32]

2.3.3 Solution 3: bite the bullet

"To me, therefore, the real usefulness of this picture or theory of quantum mechanics is simply as an alternative which could be acceptable to those who sense the paradoxes in the conventional formulation, and therefore save much time and effort by those who are also disturbed by the apparent inconsistencies of the conventional model. As you know, there have been a large number of attempts to construct different forms of quantum mechanics to overcome these same apparent paradoxes. To me, these other attempts appear highly tortured and unnatural. I believe that my theory is by far the simplest way out of the dilemma, since it results from what is inherently a simplification of the conventional picture, which arises by dropping one of the basic postulates—the postulate of the discontinuous probabilistic jump in state during the process of measurement—from the remaining very simple theory, only to recover again this very same picture as a deduction of what will appear to be the case to observers."

— Hugh Everett III to Max Jammer, 1973.[33]

Everett proposed that the formalism of quantum mechanics is not in need of additions, neither to the dynamics nor the fundamental ontology. On the contrary, he proposed to drop the two measurement postulates, i.e. to do away with Born's rule and projection as fundamental axioms such that one is left only with the state vector in Hilbert space that evolves under unitary evolution. The idea is to then derive the Born rule from that structure. The absence of the projection postulate is justified with the radical idea that it is unnecessary simply by denying that our macroscopic world is cleansed of superpositions. These superpositions are everywhere but cannot be observed by us due to the fact that we are part of only one of the branches in the wavefunction of the entire universe.

Furthermore, to explain the measurement process, Everett used his notion of 'relative states', which he defines as follows.[34] There is a composite system $\mathcal{S} = \mathcal{S}_1 + \mathcal{S}_2$ in the total state $|\Psi^{\mathcal{S}}\rangle$. For every state $|\eta\rangle$ of $\mathcal{S}_2$ we can associate the state of $\mathcal{S}_1$ given by $|\Psi^{\eta}_{rel}\rangle$, which is the state in $\mathcal{S}_1$ relative to the state $|\eta\rangle$ in $\mathcal{S}_2$. Up to a normalization constant, it is given by

$$|\Psi^{\eta}_{rel}\rangle \propto \sum_i \left\langle \eta \otimes \mu_i \,\middle|\, \Psi^{\mathcal{S}} \right\rangle |\mu_i\rangle \tag{2.6}$$

[32] Gerard 't Hooft, *The Cellular Automaton Interpretation of Quantum Mechanics* (Springer Open Access, 2016). *Fundamental Theories of Physics.* [42]

[33] Everett, "Chapter 22: Correspondence: Jammer, Wheeler, and Everett," *Everett interpretation*, J. A. Barrett, & P. Byrne, p. 296.

[34] Everett, "The Theory of the Universal Wavefunction," *Everett interpretation*, Barrett, J. A. & P. Byrne, p. 99.

where the μ_i are basis vectors of an arbitrary (but complete) orthonormal basis set in $\mathcal{S}_1$.[35]

The idea is that in a composite system $\mathcal{S}$, the relative states encode the correlations between the subsystems $\mathcal{S}_1 + \mathcal{S}_2$. The relative state $|\Psi^{\eta}_{rel}\rangle$ is defined in $\mathcal{S}_1$ relative to the state $|\eta\rangle$ in $\mathcal{S}_2$. This can be seen if you obtain $|\Psi^{\eta}_{rel}\rangle$ in the following way: expand the total state $|\Psi^{\mathcal{S}}\rangle$ in such a way that $|\eta\rangle$ forms one branch. Seen from $\mathcal{S}_1$, one is left with $|\Psi^{\eta}_{rel}\rangle$. In Everett's own words:[36]

> There does not, in general, exist anything like a single state for one subsystem of a composite system. Subsystems do not possess states that are independent of the states of the remainder of the system, so that the subsystem states are generally correlated with one another. One can arbitrarily choose a state for one subsystem, and be led to the relative state for the remainder. Thus we are faced with a fundamental relativity of states, which is implied by the formalism of composite systems.

On the one hand, Everett's theory was not taken seriously due the pushback from many prominent Copenhagen theorists, who often seemed to be under the impression that the Copenhagen view worked perfectly well and that there was no immediate reason for replacing it. This would work fine for those who were thoroughly convinced by the Copenhagen language of complementarity, but also for the more pragmatically-minded physicists. Some would even deny the existence of a measurement problem or would even regard the Copenhagen interpretation itself as a consequence (!) of the mathematical formalism:[37] in the absence of a 'phenomenon' (Bohr's terminology for the system of interest and a well-specified experimental context), you do not know how to apply the theory at all, i.e. the quantum state is ill defined in the absence of the measurement device.

Nevertheless, despite this sociological pushback, Everett's theory also has several real problems that must be addressed. The three important problems are: excess ontology, the probability problem and the determinate records problem. Before addressing these, it is intriguing to note that Everett himself did not recognize these problems. For him, his theory was everything that could be expected from a scientific theory and he seems to have regarded his theory as complete. Everett's specific philosophical stance bares a resemblance to constructive empiricism. Jeffrey Barrett has called it 'empirical faithfulness'.[38]

---

[35] The relative state is unique due to its independence on the auxiliary basis $|\mu_i\rangle$, which can be proven by expanding it into another auxiliary orthonormal basis $|\mu_i\rangle = \sum_j c_{ij}|\nu_j\rangle$, plugging it into Eq. 2.6, and using its orthonormality $\sum_i c^*_{ij}c_{ik} = \delta_{jk}$ (unless the projection of $|\Psi^{\mathcal{S}}\rangle$ onto $|\eta\rangle$ is zero, in which case the relative state can be defined arbitrarily).

[36] Everett, "The Theory of the Universal Wavefunction," J. A. Barrett, & P. Byrne, p. 180.

[37] See, for example, Stern's letter exchange with Everett, documented in Barrett & Byrne, *The Everett Interpretation*, pp. 214–224.

[38] Jeffrey A. Barrett, "On the Faithful Interpretation of Pure Wave Mechanics," *The British Journal*

Everett was not a realist about the wavefunction. Instead, he always wrote 'real' in scare quotes and was never very clear about which parts of the formalism should correspond to actual objects in the world. His view was that a logically consistent model that can account for all empirical facts of experience is enough. He considered his version of quantum theory as empirically faithful in this sense. To him, his success was that he solved the logical inconsistency of the dual rules of dynamics and derive the quantum statistics via his relative states. In addition, Everett's empirical faithfulness also provides a lean attitude towards which parts of the formalism can be said to taken seriously and which cannot. In short, as long as 'some part' of the formalism saves the phenomena, the theory is fine.

Although Everett was not a direct realist about the wavefunction, his theory does treat different branches as 'equally real', in the sense that the theory treats them on the same footing (except for their branch weights). *Prima facie*, this seems to be very ontologically extravagant. The other quantum theories seem to be able to actually come away with only one *actual* branch, whereas the other branches that arise from the Schrödinger dynamics are either not there from the start or 'empty' (they play no role in observations nor do they represent probabilities). A debate can be held about how to apply Occam's razor to this situation. This medieval principle is to not assume unnecessary ontology. On the one hand, one can then interpret the Everett interpretation, with its infinity of branching worlds, as ontologically very extravagant. On the other hand, one can argue that GRW theory (even with radically modified weights) and Bohmian mechanics (which add the extra variables as real entities) and hence also have all the branches. If that is accepted, it is even more extravagant to postulate additional dynamical laws or hidden variables (to get rid of all the branches that result directly from Schrödinger dynamics).

The problem with determinate records concerns the question in which way classically behaving measurement records arise at the end of a measurement. If the states are purely relative, one can always represent the records in a different basis in which they cannot be read as records. This would seem to have grave results for how one understands the reality of macroscopic objects.

The probability problem is sometimes also called the incoherence problem, since there is an apparent incoherence if one wants to ascribe probabilities to events, given that every possible event is certain to occur.[39] Everett's view on the matter is that the probabilities are epistemic and based on unitary dynamics. The use of the Born measure on Hilbert space can be justified in the same way as the use of the Lebesgue measure on the classical phase space. The idea is, then, that according to this measure one kind of behaviour is 'typical', i.e. it will occur 'almost everywhere'. This is in the same sense that in statistical mechanics the equilibrium microstates are overwhelmingly represented (if there are many degrees of freedom) in the classical phase space relative to the Lebesgue measure. Hence, it is typical for a system to be in equilibrium. However,

for the Philosophy of Science, Vol. **62**, No. 4 (2011), pp. 693–709.

[39] Hilary Greaves, "Probability in the Everett Interpretation," *Philosophy Compass* (2007), Vol. **2**, No. 1, pp. 109–128.[40]

there are disagreements about how natural the Born measure is if one takes the branches as real. In that case, the theory tells us that atypical branches still exist despite them being atypical. It seems that Everett's approach was to regard those atypical branches as a part of the formalism that need not correspond to an experience. Hence, Everett's specific interpretation in the philosophy of theories, empirical faithfulness, is crucial for the acceptance of his own theory.[40]

As a last remark, notice that the whole Everett approach hinges on the assumption that the Schrödinger equation *exactly* holds. No non-linear term, however small its effect, can be added to it without destroying the entire framework: Everett's work (and that of all versions of the many-world approach) is built around the assumption that unitary evolution is sacred. I do not intend to say that unitarity is an unwarranted assumption. On the contrary, unitarity of the fundamental physics is widely accepted and will not be easily relinquished. However, one need only remember the history of the unification of light and electromagnetism by Maxwell's addition of the (very) small term to Ampère's law to understand that one should not be too dogmatic about the exact validity of fundamental equations. Nonetheless, taking the wavefunction and its unitary evolution as the axioms of a theory is not expensive when it comes to the amount of axioms and deserves to be taken as a serious candidate.

In the next chapter, I will focus on how far the idea of 'taking the quantum formalism literally' goes in the most worked-out version (although it has many differences with Everett's original ideas) of Everett's theory: Wallace's *The Emergent Multiverse*. Wallace's approach differs substantially from Everett in two (major) ways. First, Wallace derives the Born rule probabilities not from typicality-considerations, but from a subjective, decision-theoretic view. This approach (by Deutsch, Wallace, and (to some extent) Hilary Greaves) is fundamentally different from Everett's. For the rest of this research, however, I will leave the probability question for what it is.

The second difference is that where Everett was satisfied with the notion of relative states, Wallace takes into account a multiplicity of 'real' states corresponding to each world, in which determinate records remain stable and robust under decoherence. Therefore, the relative state is not so important for Wallace, although it can also be defined within his framework. For Wallace, the universal state is written in terms of a physically preferred basis at the coarse-grained level in which measurement records become determinate in each branch of this basis. Each branch, in turn, becomes a real physical world (represented by that branch). This process and the interpretation that Wallace attributes to it is the subject of the rest of this thesis.

[40] One of the points of my thesis is that a similar situation occurs in Wallace's *The Emergent Multiverse*: only when one takes a specific interpretation about what a theory is supposed to describe in reality, can one accept his solution to the measurement problem, *cf.* Chapter 8. The difference is that Everett's interpretation bends towards empiricism, while Wallace's has most of the characteristics of realism of scientific theories.

3 Wallace's Emergent Multiverse

"It supposes—as was first proposed by Hugh Everett, fifty years ago—that neither the mathematical formalism of quantum mechanics nor the standard conception of science is in any need at all of modification. Rather: the unmodified quantum theory can be taken as representing the structure of the world just as surely as any other theory of physics. In other words, quantum mechanics can be taken literally."

— David Mark Wallace.[41]

Wallace's view of quantum mechanics starts from adopting Everett's "extremely conservative approach to quantum mechanics," as described by the quote above. This is surely an alluring idea. If the measurement problem can be solved by neither postulating extra dynamics (collapse theories) nor by extra ontology (hidden variables), there is less to explain and no new physics needs to be thought of in the research&development departments. At least, that would be what one expects, since it sounds like the simplest solution: we just need to open our eyes to what the formalism is telling us. Nevertheless, the opening of the eyes requires quite some effort and an occasional blink.

In this chapter, we will look at Wallace's use of the dynamical mechanism of decoherence, through which these 'worlds' are said to emerge, in Sections 3.2, 3.4 and 3.5. Section 3.3 is an interlude that intends to take away the idea that decoherence solves the measurement problem by itself. The following Chapter 4 is devoted to a critical evaluation of the claim of emergence.

Wallace analyses the concept of or emergence in the second chapter of his book, while describing the emergence through decoherence in the third. In the first section, we will look at the idea of 'instantiation' as a route to emergence and how Wallace applies this to branches of the wavefunction. After that we will describe the physical mechanism of decoherence and what it means for the measurement problem. Then, the way in which the (coarse-grained) decoherence basis has a special status (or is 'preferred') is discussed. Finally, a section about the continuous character of branching is included. In the next chapter, Wallace's specific claim of emergence is critically evaluated, which is that[42]

> unitary quantum mechanics, interpreted realistically, is a many-worlds theory—not because the 'worlds' are present in some microphysically fundamental sense but because the quantum state instantiates many different macroscopic systems.

Hence, Wallace undermines the argument against the many-worlds theory that the worlds are added ontology because they are not there in the foundations of the theory. According to Wallace, the worlds are not axiomatic, but emergent entities like all macroscopic objects, such as tables or butterflies.

[41] Wallace, *Emergent Multiverse*, p. 13.

[42] *Ibid.*, p. 63.

3.1 Instantiation

Wallace's claim is that the emerging worlds *behave similarly to classical objects*; this is called 'quasi-classical' behaviour. The central problem he is concerned with is macroscopic definiteness. Consider again Schrödinger's cat, as outlined in Chapter 2, as an example of such an indefinite macroscopic state; repeating Eq. (2.5), both a dead and a live cat are present at the level of the wavefunction:

$|\psi\rangle = c_\uparrow |\uparrow\rangle \otimes |$  $\rangle + c_\downarrow |\downarrow\rangle \otimes |$  $\rangle$. (3.1)

Schrödinger used this to tease out a problem with (the completeness of) the quantum formalism, but Wallace, following quantum state realism, makes the following turn: what we have here are two bits of formalism that 'give rise to' (*instantiate*, see below) both the *structures* of one live cat and one dead cat. So, whereas before there was simply one state of the cat, there are now *two* macroscopic states, distinct from each other.[43] It seems that "[s]uperposition has become multiplicity at the level of structure".[44]

Since the system cannot really be closed in practice, the cat will also interact with the rest of the room where it is in, after which it will interact with the observers, and eventually the entire universe. Following this von Neumann chain leads us to an entirely entangled system. Starting with the universal wavefunction, in some initial state and at an initial time before the time of interaction, looks like

$c_\uparrow |\uparrow\rangle \otimes |$  $\rangle \otimes |\text{rest of universe}\rangle + c_\downarrow |\downarrow\rangle \otimes |$  $\rangle \otimes |\text{rest of universe}\rangle$. (3.2)

After a final time t_F (sufficiently long for the dynamics to take place), I claim, a 'duoverse' arises, because it evolves into a state[45]

$$c_\uparrow |\text{universe with spin up \& live cat}\rangle + c_\downarrow |\text{universe with spin down \& dead cat}\rangle . \tag{3.3}$$

The point is that even though the concept 'duoverse' does not exist in the fundamental ontology of the system, it can be seen as an emergent entity. In this case, the Wallacian turn would be to interpret Eq. (3.2) as two bits of structure that instantiate two macroscopically definite worlds.

An 'instantiation' means finding an 'instance' of something; Wallace uses the term to indicate a relationship between two theories inside a certain domain. The theory of classical mechanics (CM) reigns in the domain of the Solar system; here,

[43] This interpretation should raise a suspicion, namely that in Hilbert space one has the freedom to *rotate the basis*. Why is there a multiplicity in the dead/alive basis and not in another? The reason for this is that a particular basis is picked out by the *dynamics*, see Section 3.4 for this objection, which is called the preferred basis problem.

[44] Wallace, *Emergent Multiverse*, p. 61.

[45] We will see in Section 3.2 that this would be the case for virtually all Hamiltonians that we think should govern the precise dynamics behind this entangling of systems.

molecular physics instantiates a theory of classical point particles subject to Newton's laws. In game theory, extended to a domain of typical behavioural patterns of animals, we find an instantiation of zoology, where evolutionary stable strategies arise from selfish and rational actors. Thus, instantiation stays true to the spirit of reduction, but makes weaker demands than Nagelian reductionism, because an exact derivation is not needed. Wallace admits that a precise definition of instantiation falls short when applied to examples like the ones above, but he does give a working definition:

> Given two theories A and B, and some subset D of the histories of A, we say that A instantiates B over domain D iff there is some (relatively simple) map ρ from the possible histories of A to those of B such that if some history h in D satisfies the constraints of A, then $\rho(h)$ (approximately speaking) satisfies the constraints of B,[46]

where, a 'history' should be understood as a connection between states at different time slices. The idea is clear: it is an attempt to 'weakly reduce' theory B to theory A, within a certain domain D, *by allowing for approximations*. Hence, there exists a map that approximately maps histories constrained by theory A in domain D to histories constrained by theory A in that same domain,

$$\exists \rho : \rho(h^{A_D}) \rightarrow h^{B_D}, \tag{3.4}$$

which I will write in a shorthand notation for instantiation:

$$(A_D \rightsquigarrow B_D), \tag{3.5}$$

where $(x \rightsquigarrow y)$ should be read as 'x instantiates y'.

How are the classical worlds instantiated by quantum mechanics? Armed with a working definition of instantiation, we want to interpret Eq. (3.2) as representing two structures that *instantiate* the structure of a universe containing the dead cat *and* another one containing a live cat.[47] To do this, we need two things. First, we must realize that two branches should not interfere when they interact with other systems, i.e. we do not want cross terms when calculating probabilities. Second, we need to justify the basis in which we are working, as there was a critical assumption in interpreting Eq. (3.1) as multiplicity arising from superposition. We can always rotate our basis such that the superposition looks different or disappears entirely! In one such a rotated basis, Eq. (3.1) reads[48]

$c_{\uparrow+\downarrow}\{|\uparrow\rangle \otimes |$  $\rangle + |\downarrow\rangle \otimes |$ $\rangle\} + c_{\uparrow-\downarrow}\{|\uparrow\rangle \otimes |$  $\rangle - |\downarrow\rangle \otimes |$ $\rangle\}.$ (3.6)

[46] Wallace, *Emergent Multiverse*, p. 54.

[47] And additional live or dead cats, as most universes contain many more cats than the one we are considering.

[48] From here on, $\otimes$ will be dropped for brevity.

According to exactly the same analysis as above, a multiplicity of structures will be instantiated from this. These will not resemble classical cats, but there are parts of the cat structure that before was associated with a (dead or live) classical cat.[49] In other words, if the analysis for Eq. (3.1) was sound, then we should also conclude that the structures associated with the two vectors in Eq. (3.6) should correspond to two distinct real structures. Hence, we must deny that mere superposition is the cause of the multiplicity, even if there is one special basis in which the branches of the superposition are not interfering. We need something extra.

It turns out that the dead cat state and live cat state of Eq. (3.1) will *not* interfere with each other, exactly because it is written down in the decoherence basis. Wallace takes this as sufficient reason to regard the decoherence basis as the preferred basis.[50] Decoherence is a dynamical mechanism within the quantum formalism, independent of interpretation and not picking a 'preferred' basis on its own. Hence, this mechanism is used to solve the determinate records problem (or preferred basis problem) by giving rise to stable patterns that are robust under decoherence.[51]

3.2 The mechanism that makes it all possible: decoherence

Let us start with a general superposition of a quantum system $|\Psi\rangle = \sum_i c_i |\psi_i\rangle$, relative to some set of orthonormal vectors $\{|\psi_i\rangle\}$ and coefficients c_i in the Hilbert space $\mathcal{H}_\mathcal{S}$. Then we perform a measurement using a measurement device $\mathcal{M}$ with eigenstates $\{|m_i\rangle\}$ such that this apparatus will become entangled with the system (subscript 0 will indicate the initial state). Hence,

$$|\Psi'\rangle = \sum_i c_i |\psi_i\rangle |m_0\rangle \xrightarrow{H_1} \sum_i c_i |\psi_i\rangle |m_i\rangle , \tag{3.7}$$

which is governed by an interaction Hamiltonian H_1.[52] The right-hand side again describes a superposition, and since we do not observe superpositions in everyday life,

[49]In Chapter 7, I will make a distinction between qualitative and quantitative features of this analysis. Looking at instantiation at face value reveals that it is quantitative mapping from one theory's histories to the histories of another theory: it is approximate pattern-matching. It says nothing of the qualitative concepts about cats that we know from classical mechanics (or, better put, zoology). Indeed, classical animals are never entangled. However, since decoherence is always approximate, Wallace's classical structures *are* entangled. Here we see a qualitative aspect: classical animals are not entangled *even in principle*. Quantitatively, we can ignore the non-zero support. Qualitatively, we need something more than instantiation to say that in the rotated basis nothing is instantiated.

[50]One of my criticisms in this thesis will be aimed at this identification of the preferred basis with the decoherence (or classical) basis, as discussed in Chapter 6.

[51]We will see in Section 3.3 that it does not, by itself, solve the problem of definite outcomes.

[52]Surely, such a Hamiltonian must exist for the apparatus to qualify as a measurement device. In other words, $\mathcal{M}$ makes a measurement in the $|\psi_i\rangle$ basis.

we are faced with an explanatory gap.[53] The measurement problem revolves around making empirical sense of Eq. (3.7).

Quantum systems are never actually isolated from the environment—neutrinos, the cosmic microwave background, or imperfect vacua, will always interfere. Hence, we cannot naively regard Eq. (3.7) as the state of a closed system.[54] So, we have to take into account all of the environment $\mathcal{E}$, eigenstates $\{|\epsilon_i\rangle\}$, in the gargantuan Hilbert space $\mathcal{H}_{tot} = \mathcal{H}_\mathcal{S} \otimes \mathcal{H}_\mathcal{M} \otimes \mathcal{H}_\mathcal{E}$. The entire state will then be

$$\sum_i c_i |\psi_i\rangle |m_0\rangle |\epsilon_0\rangle \xrightarrow{H_1} \sum_i c_i |\psi_i\rangle |m_i\rangle |\epsilon_0\rangle \xrightarrow{H_2} \sum_i c_i |\psi_i\rangle |m_i\rangle |\epsilon_i\rangle , \tag{3.8}$$

where H_2 is the interaction Hamiltonian between the system+apparatus and the environment. The environment cannot be ignored, but neither can it be controlled. In fact, the properties of the $\{|\epsilon_i\rangle\}$ are practically unattainable, since we are dealing with a large number of environmental states. In such a situation, the (mathematically equivalent) reformulation of quantum mechanics in terms of density operators is advantageous. For our system-apparatus-environment state the total density operator is

$$\hat{\rho}_{\mathcal{SME}} = \sum_{j,k} c_j c_k^* |\psi_j\rangle |m_j\rangle |\epsilon_j\rangle \langle\psi_k| \langle m_k| \langle\epsilon_k| \,. \tag{3.9}$$

Generally the environment is not just unknown but also completely uninteresting; after all, we are interested in the system we are measuring. Thus, we should look at the reduced density matrix – comprising system and apparatus only – by tracing out the environmental degrees of freedom,

$$\hat{\rho}_{\mathcal{SM}} = \mathrm{Tr}_{\mathcal{H}_\mathcal{E}}\{\hat{\rho}_{\mathcal{SME}}\} = \sum_{j,k} c_j c_k^* |\psi_j\rangle |m_j\rangle \langle\psi_k| \langle m_k| \langle\epsilon_j|\epsilon_k\rangle \,. \tag{3.10}$$

We can use this density operator to calculate the expectation value of some observable $\hat{O}$,

$$\langle\hat{O}\rangle := \mathrm{Tr}\{\hat{\rho}_{\mathcal{SME}}\hat{O}\} = \mathrm{Tr}_{\mathcal{H}_\mathcal{E}}\hat{\rho}_{\mathcal{SM}}\hat{O}_{\mathcal{SM}}. \tag{3.11}$$

The crucial assumption is the orthogonality of environmental states, which naturally depends on the form of the Hamiltonian leading to these states. However, usually one appeals to the high dimensionality of $\mathcal{H}_\mathcal{E}$ compared to $\mathcal{H}_\mathcal{S} \otimes \mathcal{H}_\mathcal{M}$, i.e. the huge amount of air molecules, CMB photons, etc.. Many specific models have been worked out that

[53] It will not help to measure this superposition with another measurement device $\mathcal{M}_2$, because the superposition is contagious: $\sum_i c_i |\psi_i\rangle |m_{1,0}\rangle |m_{2,0}\rangle \xrightarrow{H_1} \sum_i c_i |\psi_i\rangle |m_{1,i}\rangle |m_{2,0}\rangle \xrightarrow{H_2} \sum_i c_i |\psi_i\rangle |m_{1,i}\rangle |m_{2,i}\rangle$, *ad infinitum*.

[54] The only completely closed system is the entire universe. This is not nit-picking when it comes to entanglement: even when the physical interaction (Hamiltonian) is negligible, the entanglement that arises from that interaction goes through in the same way.

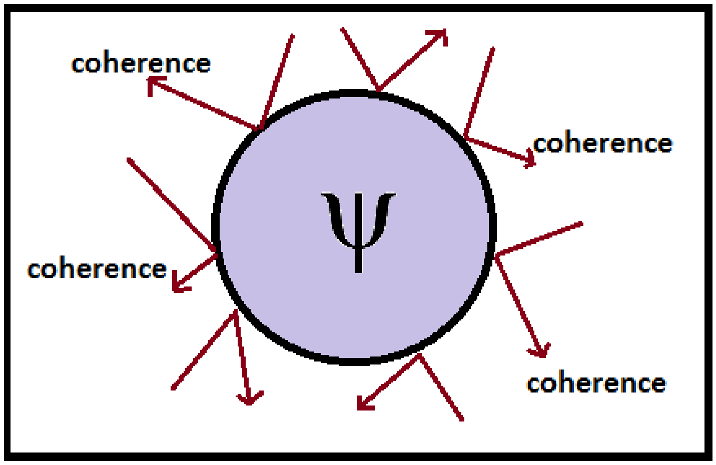


Figure 1: Decoherence, the environmental states (red arrows) continuously bombard the quantum system (blue 'ball'), carrying away the coherence of the system, which in turn loses the off-diagonal terms in its density matrix

show that these states are orthogonal or very rapidly become so:[55]

$$\langle \epsilon_i | \epsilon_j \rangle \approx \delta_{ij}. \tag{3.12}$$

In fact, the 'very rapidly become so' is specifically exponential decay in time. Adler shows this by using a scattering matrix approach, where the environmental states $|\epsilon_i\rangle$ start out as 'ready states' (their inner product is unity) scatter off the system states A and B, changing the state slightly (it is sufficient, but not necessary, that the environmental particles change the state of the system only lightly) the inner product $\langle \epsilon_i | S^{\dagger}_{i(A)} S_{i(B)} | \epsilon_i \rangle$ is slightly smaller than 1.[56] Hence, the number N of sequential interactions of the particles with the system gives sequential multiplications of numbers slightly smaller than 1 (and later terms in the sequence are much smaller than 1), such that the sequence approaches zero exponentially.

$$\prod_i^N \langle \epsilon_i | S^{\dagger}_{i(A)} S_{i(B)} | \epsilon_i \rangle \propto \exp\left[-\lambda_{ch} N\right] \tag{3.13}$$

for some characteristic decoherence rate λ_{ch}. And since the number of interactions grows linearly in time, the environmental states approach orthogonality exponentially in time.

The result of Eq. (3.11) shows the loss of coherence of the system+apparatus,

[55] Erich Joos, Hans-Dieter Zeh, Domenico Giulini, Claus Kiefer, Joachim Kupsch, Ion-Olimpiu Stamatescu, *Decoherence and the Appearance of a Classical World in Quantum Theory* (Heidelberg: Springer, 2003).[50] A very good overview of these studies is given on pp. 64–68.

[56] Stephen Adler, "Why decoherence has not solved the measurement problem: a response to P.W. Anderson," *Studies in History and Philosophy of Modern Physics* (2011), Vol. **34**, pp. 135–142.[1]

as phase relations dissipate into the environment, see Figure 1. Even if the interaction is weak, this is overshadowed by the amount of interactions and the system will still become entangled with the environment; thus coherence is dissipated into the environment. This effect is seen by the disappearance of interference terms (the cross terms) in the state of the reduced system,

$$\hat{\rho}_{\mathcal{SA}} = \mathrm{Tr}_{\mathcal{H}_{\mathcal{E}}}[\hat{\rho}_{\mathcal{SAE}}] \approx \sum_j |c_j|^2 \Big(|\psi_j\rangle \langle\psi_j| \otimes |a_j\rangle \langle a_j| \Big). \tag{3.14}$$

In the right basis, the density operator becomes diagonal very quickly, signalling the loss of coherence between the branches of the superposition.

3.3 Decoherence does not solve the measurement problem: decoherence plus Everett does

Since interference effects at the macroscopic level are suppressed by the process of decoherence, many have claimed that the effect can be taken as solution to the measurement problem, in the sense that it provides a mechanism for effective collapse. This has most famously (because Stephen Adler uses them as examples) been proclaimed by Nobel laureate Philip Anderson[57] and in Tegmark and Wheeler's 2001 article[58], but it is found more often in the (primarily condensed matter) literature at the turn of the century.[59] These are based on a confusion, namely that at the level of the formalism, the results of the projection of pure states are formally similar to the reduced density matrices of improper mixtures. But although the density matrix of the combined system is pure, Eq. (3.9), the reduced density matrix, Eq. (3.9), is improper such that an ignorance interpretation cannot apply.[60] Zeh writes:[61]

> Environment-induced decoherence by itself does not yet solve the measurement problem, since the pointer states [$|a_i\rangle$] may be assumed to *include* the total environment (the 'rest of the world'). Identifying the thus arising global superposition with an *ensemble* of states [...] would beg the question. This argument is nonetheless found wide-spread in the literature.

Although it is enough to state that the ignorance interpretation is not applicable to understand that the problem of unique measurement outcomes is not solved, some further reasoning can provide more insight. We have seen that decoherence is a dynamical

[57] Philip Anderson, "Reply to Cartwright," *Studies in History and Philosophy of Modern Physics* (2001), Vol. **32**, pp. 499–500.[4]

[58] Max Tegmark & John Wheeler, "100 Years of Quantum Mysteries," *Scientific American* (2001), Vol. **284**, pp. 68–75.[65]

[59] It is also still sometimes heard to this day, which compels one to include a section 'Decoherence does not solve the measurement problem'.

[60] Bernard d'Espagnat, *Conceptual Foundations of Quantum Mechanics*, 2nd edition (Reading: Benjamins Publishing, 1976), pp. 86–87.[25]

[61] Zeh in Joos et al., *Appearance of a Classical World*, p. 21.

process resulting from the entanglement of the system $\mathcal{S}$ with the environment $\mathcal{E}$. This process is practically irreversible because the system state is an 'open system' (it is in contact with an environment) and the environment has a large number of degrees of freedom.

However, there is a catch when comparing the ensemble of states and the single pure state. Decoherence does not account for the projection of a single pure state like $|\Psi\rangle = \sum_i c_i |\psi_i\rangle \to |\psi_k\rangle$. What decoherence rather describes is the *suppression of interference* at the level of an improper mixed state, since the reduced density matrix evolves non-unitarily in time (it is an open system). Osvaldo Pessoa shows that what is formally going on is much more like an ensemble version of the projection postulate:[62]

$$\hat{\rho} = \sum_{jk} c_k c_j^* |\psi_k\rangle \langle\psi_j| = \sum_j |c_j|^2 |\psi_j\rangle \langle\psi_j| + \sum_{j \neq k} c_k c_j^* |\psi_k\rangle \langle\psi_j| \xrightarrow{Proj.} \sum_j |c_j|^2 |\psi_j\rangle \langle\psi_j| ,$$

where in the last step the interference terms are suppressed by invoking what Pessoa calls a *statistical* projection. Thus, formally, the same result is obtained by either using this statistical version of projection or by taking the partial trace over the environment. The measurement problem for individual systems is thus not solved for individual systems, which lose information due to the tracing out of the environmental degrees of freedom. Pessoa concludes: "*an explanation for collapse implies an explanation for decoherence, but an explanation for decoherence doesn't imply an explanation for collapse.*"[63] Decoherence does not solve the measurement problem because it does not provide a mechanism for individual collapse.

In conclusion, decoherence is the diagonalization of the density operator, but it does not dissolve the superposition: the diagonal terms are still there in the state of the total system; this could for instance represent a superposition of macroscopic objects at wildly different positions. Thus, although interference is removed, we still have to deal with this problem of indefiniteness: decoherence alone does not solve the measurement problem. In fact, it seems to make the measurement problem worse,[64] whereas before we could have hoped that superpositions could be confined to the microscopic realm, we now realize that all of nature falls prey to the indefiniteness of superposition.

Following Wallace, we can now invoke Everett's insight that we are ourselves part of the universal wavefunction; in other words, we should ourselves be included in the environmental states. As we become entangled to the system in much the same way as the apparatus did in Eq. (3.7), it readily becomes clear that both our observations and registrations of those observations in our brains will be confined to one branch of that superposition.[65] This insight is now backed up by decoherence, as it ensures us

[62] Osvaldo Pessoa, "Can decoherence help to solve the measurement problem?" *Synthese* (1988), Vol. **113**, pp. 323–346. Dordrecht: Kluwer Academic Publishers, p. 326.[66]

[63] Pessoa, "Decoherence," p. 325.

[64] Bacciagaluppi, "The Role of Decoherence."

[65] Our brains and conscious experience will be instantiated by the wavefunction and will emerge along with everything else; this will have far-reaching implications for the philosophy of mind, as we

that the different branches will effectively never interfere with each other. *The branches get an autonomous status because decoherence detaches them from each other.* Thus, at least under the right conditions, a quantum superposition can indeed instantiate a multitude of non-interfering trajectories in phase space that hardly overlap: *quasi-classical trajectories.*

3.4 The dynamically preferred decoherence basis

We have not answered the objection that one can always rotate the basis. If we had rotated the basis, the off-diagonal terms of the density matrix (the interference terms) would have become mixed with the diagonal terms. In fact, because it is Hermitean, any density matrix can be diagonalized in some basis, the so-called Schmidt decomposition.[66] This Schmidt basis will be very different from the decoherence basis, especially when the density matrix has degenerate eigenvalues and starts off near-diagonal.[67] Hence the above analysis is not sufficient to explain the appearance of autonomous 'worlds' in some 'world-basis', by which I mean a basis that respects macroscopic degrees of freedom. This is because we have assumed that the interaction between system and environment is measurement-like, in the sense that the interaction Hamiltonian approximately commutes with some observable of the system, and the eigenstates of this observable couple to different states of the environment.

The reason why this particular basis gets a special status is that Hamiltonians are functions in terms of position operators—two scattering particles, in general, interact through $H_{int} = V(x, x')$. This has been a constant in physical practice ever since Newton's inverse square law of gravitation, depending on the relative position between two masses. This provides a *preferred basis* in the sense that superpositions of positions that are macroscopically far apart will very quickly decohere, while spatially localized states will barely become entangled with the environment. Thus, there is a dynamical reason for the special role of position.[68] But the form of the Hamiltonian specifies a preferred position basis at the fine-grained level. For the coarse-grained level, this has the consequence a coarse-grained position basis is preferred in the sense of decoherence.

will see later.

[66] Keith Hannabuss, "Quantum Theory & Quantum Computers," *Lecture Notes* (2012), Oxford University, section 1.5, URL: `http://users.ox.ac.uk/~kch/demos/quantc7/`.

[67] Matthew J. Donald, "Discontinuity and Continuity of Definite Properties in the Modal Interpretations," in: *The Modal Interpretation of Quantum Mechanics* (Dordrecht: Kluwer Academic Publishers, 1998), pp. 213–222. Editors: D. Dieks, P. E. Vermaas.[32]

[68] This is, however, a bit less straightforward than it is represented here. One can argue that *in addition* to the dynamical preference of decoherence happening in the position basis, this position basis was already preferred before the dynamical analysis, since the Hamiltonian was already written in that basis! Indeed, the decoherence analysis goes through for any basis. If one had written an interaction Hamiltonian in terms of momentum operators, the momentum basis would be dynamically robust. The question then is about the arrow of the argument: does it point from the form of the interaction Hamiltonian to the preferred position basis, or from preferred position basis to the form of the interaction Hamiltonian? This will be addressed in Section 6.3.

Indeed, since the full Hamiltonian is not just a function of position, it turns out that some coarse-grained joint position-momentum basis is preferred. That latter, coarse-grained, position basis is called the 'decoherence basis'.

Zeh and Erich Joos calculated, as a special case of Adler's calculation above, how fast diagonalization in the position basis occurs for the spatial localization of macroscopic objects. They found that after the scattering with environmental particles (in this case polarized photons, but the results turn out to be quite general for different scattering potentials: even when the dynamics is chaotic or far from classical, decoherence nevertheless ensures localization in the position basis.) the reduced density matrix of the system acquires an exponential decay function of the diagonal entries (where $x \neq x'$),[69]

$$\rho(x, x') \to \rho(x, x') \exp\left[-\Lambda t (x - x')^2\right], \tag{3.15}$$

where Λ is the 'localization rate', depending on the scattering cross section and the particle flux. This localization rate is often larger than the rate at which systems reach thermal equilibrium. A dust particle of $10^{-5} cm$, for example, will decohere at a characteristic timescale of 10^{-13} *sec*, due to interaction with air molecules only.[70] This is often used to define the decoherence time, $\tau_D = 1/\Lambda$, as a characteristic time scale for the exponential behaviour of decoherence. Note that this is a gradual behaviour akin to, for example, the penetration depth of the Coulomb force of a charge screened by a fluid or the exponential decay of a radio-active sample.

To cross some t's and dot some i's, Wallace discusses in length the mature version of modeling the environment: the master equation governing the evolution of the density operator. In particular, using the Wigner function $W(x, p)$[71] of the Caldeira-Leggett equation, the evolution under Hamiltonian H is

$$\dot{W} = \{H, W\}_{MB} + \Lambda_{CL} \frac{\partial^2 W}{\partial p^2}. \tag{3.16}$$

This is a master equation for an environment linearly coupled to the system of interest,[72]

[69] Joos et al, *Appearance of a Classical World*, p. 67.

[70] Erich Joos, Hans-Dieter Zeh, "The Emergence of Classical Properties Through Interaction with the Environment," *Zeitschrift für Physik B - Condensed Matter*, Vol. **59**, pp. 223–43, (1985), p. 232.[49]

[71] The Wigner function is a hybrid between the position and momentum representation of a quantum state, defined as $P(x, p) := 1/\pi\hbar \int dy \Psi^*(x + y)\Psi(x - y) \exp[2ipy/\hbar]$, where Ψ is the wavefunction and x and p are are position and momentum (but can be any conjugate pair). $P(x, y)$ lends itself to a 'quasi-probability' interpretation. For a good discussion of the Wigner function and the Moyal bracket $\{,\}_{MB}$ in our context, see Żurek [90].

[72] One can view the Caldeira-Leggett system as a Brownian particle that is attached to particles in its environment through springs, see Amir O. Caldeira & Tony Leggett's "Path Integral Approach to Quantum Brownian Motion" [17]. For everyday (Ohmic) environments, the quantized version of this model is a highly successful way to model quantum Brownian motion and, therefore, to calculate decoherence rates. In a previous Theoretical Physics thesis, I have extensively discussed the Caldeira-Leggett model and its subtleties, see my `https://dspace.library.uu.nl/handle/1874/351831`.[61]

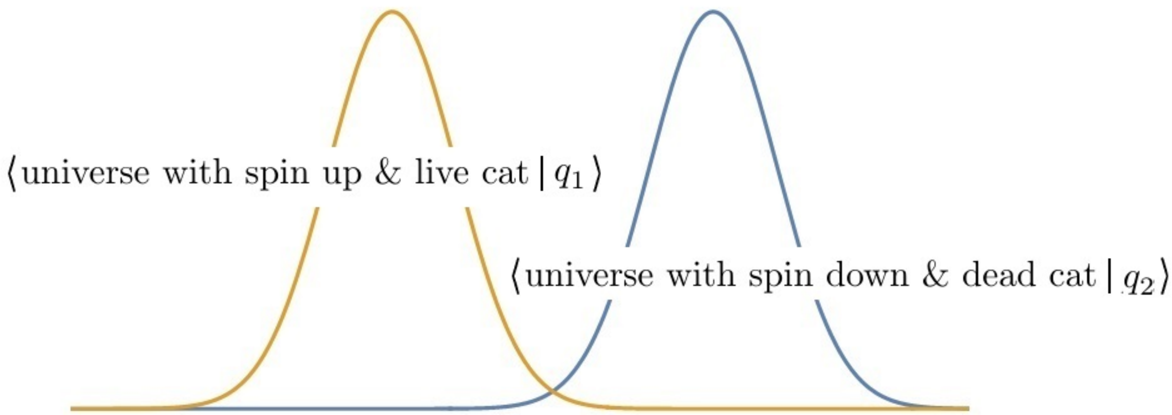


Figure 2: A sketch of the decohering wave packets of the Duoverse, Eq. (3.2). The point of decoherence is to localize the peaks of the individual branches in a basis relative to macroscopic degrees of freedom; this is illustrated by the very small – though non-zero – overlap of the branches. Thus, we observe no interference patterns of live and dead cat branches, or, for that matter, interference between worlds. These wavepackets do not spread out, but remain localized due to the interactions with the environment

showing that states (projected onto phase space) localized in momentum p (delocalized in space) diffuse through decoherence. The first term represents classical dynamics; the second is a quantum correction, which is responsible for the diffusion. Due to the second derivative in momentum, momentum states that start out localized rapidly diffuse. Conversely, it will have little effect on states well-localized in space; even if they spread out a little they will become localized again, such that they stay that way. This is basis preservation in action; such a basis is called 'robust' with respect to the environment.[73] See Figure 2 for a sketch of the two wave packets of the Eq. (3.2) duoverse, which serves as a good example of this basis preservation.

3.5 Continuous splitting: coarse-grained decoherence basis

The decoherence basis is somewhat vaguely defined. Since decoherence is an approximate phenomenon, the moment in time where one 'world' is present where it was not present immediately before (when the wavefunction was still an indefinite quantum soup) cannot be pinpointed. This is certainly the case, and Wallace accepts that this is so: "Put another way, the cat description is only useful when answering questions on timescales far longer than [the decoherence timescale] τ_D, so *whether or not* quantum splitting is occurring, it just doesn't make sense to ask questions about cats that depend on such short timescales."[74]

[73] Wallace, *Emergent Multiverse*, p. 83.

[74] David Wallace, "Everett and Structure," *Studies in History and Philosophy of Modern Physics* (2003), Vol. **34**, pp. 87–105, p. 98.[77]

I would like to illustrate this with the Schrödinger cat example and what I have called the 'emergent duoverse' of Eq. 3.2. More in line with Schrödinger's original formulation, we can replace the spin state with an unstable atom, which, when it decays, triggers the mechanism that kills the cat (and otherwise does nothing). At the long-time level, there is a multiplicity of two worlds: one which contains a live cat and one which contains a dead one. This was identified after 'a sufficiently long time', namely the time ($t \ll \tau_D$) where one can actually identify two decoherent sectors in the configuration space. If we take some steps back and think about how such a state would come about dynamically, taking into account that measurement takes some finite time and our knowledge that coherence decays exponentially, we can write something like[75]

$$|\Psi_{SC}(t)\rangle = \sqrt{\exp\left(-t/\tau\right)}\,|\text{no decay}\rangle\,|\quad\rangle + \sqrt{1-\exp\left(-t/\tau\right)}\,|\text{decay}\rangle\,|\quad\rangle\,, \qquad (3.17)$$

where τ is the decay rate (times $\log 2$) of the atom. Hence, as time progresses, continuously more worlds emerge that contain the dead cat and continuously fewer worlds 'split off' that contain the living cat.

Jeremy Butterfield points out that the approximateness of identifying a 'world' is a difficulty when addressing the preferred basis problem:

> Saunders and Wallace conclude from these difficulties that we should 'be liberal' and accept resolutions of the universal quantum state Ψ in an arbitrary basis—or at least an arbitrary basis that is a fine-graining of 'the' decoherence basis. [...] [I]n Wallace's terminology of 'worlds': they [Wallace and Saunders] consider continuously many bases (even if they restrict themselves to bases that fine-grain 'the' decoherence basis), and so are committed to continuously many worlds.

What is at stake in the above is the dynamical preference for a *precise* basis. Evidently, the decoherence process cannot pick out such a precise basis.

A 'world', emergent due to decoherence, hence, is not sharply defined in Wallace's (and Saunders') approach. This is a result of the dynamics being continuous.[76] But is this really a problem? There are many continuous things in nature where an exact demarcation between present and not-present cannot be drawn. At what exact spot does a mountain stream become a river? At what exact moment in time can an adolescent reindeer be said to have antlers? And where exactly is the boundary between the mesosphere of our world and outer space? Certainly, there is no need to give precise answers to these questions, and if such precision is required in some situation, there are no (I think) objective features in nature to settle them (arbitrary conventional definitions will be used). Can it not just be the same thing with 'worlds', as Wallace suggests?

[75]This state is written down by myself in order to illustrate the point of continuous splitting, and is not written in Wallace's presentation.

[76]In fact, even though it is continuous process, one can still quantify characteristic behaviour, *cf.* Paul Hovda, (2008). "Quantifying Weak Emergence." *Minds & Machines*, Vol. **18**, pp. 461–473.[43]

It is possible to identify (with very high precision) a time when a quasi-classical world has definitely not emerged, namely long before decoherence effects kick in, at times $t \lll \tau_D$. And it is possible to say (with very high precision) when one has emerged ($t \ggg \tau_D$). The decoherence basis can only be seen at a certain course-grained level. The emergence of worlds must simply be taken as a matter of degree.

There is a continuous spectrum of emerging worlds, but these nevertheless fall into two categories: one in which the cat is dead and one in which it is alive. The duoverse, then, is shorthand for these two types of multiverses.

4 Emergence

> *"Today there seem to be no phenomena which contradict quantum theory—perhaps with the sole exception that there are definite ('classical') phenomena at all!"*
>
> — Erich Joos.[77]

Say you decide, in a moment of scientific inspiration, to study a bird (say a starling) hopping and chirping around in your garden, and systematically collect information about it. After some time you will get to know a great deal about it: it eats worms and certain seeds, lays eggs and breeds them, flies around, and even has a favourite song to sing; given enough time, you will be convinced you know everything about this particular bird. Then, ten thousand fellow starlings swarm by and to your amazement your object of investigation decides to fly south for thousands of kilometres. In recognizing this unexpected migration, something originally absent in the behaviour of the bird has *emerged*.

This word 'emergence' derives from Latin 'emergo', which is translated as 'to come forth', 'to appear', or 'to arise'. Emergence is used in philosophy to signal the appearance of phenomena or entities arising at a scale where a description of its separate constituents seems to be inadequate to accurately describe that phenomenon or entity. In other words, the emergent property is, or seems to be, something 'over and above' the sum of the properties of its parts. Emergence is claimed to play a role in a wide range of scientific fields, which signals that there are more concepts contained in it. What is shared by its many uses is the idea that emergent properties or emergent effects at some higher level arise from a collection of constituents at a lower level where these properties or effects were absent. A common example is the emergence of a higher-level consciousness from a lower-level collection of neurons and electrical signals. More concrete situations where emergence is often invoked are the behaviour of colonies of ants or schools of fish, which have properties apparently absent at the level of one ant or one fish;[78] many physiological properties such as the beating of the heart as an emerging result from the interplay of living cells; or, the emergence of thermodynamical quantities like temperature or pressure from interactions between many atoms, whereas such concepts are completely absent at the level of a few atoms.[79]

[77] Joos et al, *Appearance of a Classical World*, p. 1.

[78] Steven Johnson, *Emergence* (New York: Scribner, 2001).[48]

[79] Emergence is in this case appealed to in order to make up for the failure of strict Nagelian reduction of thermodynamics to kinetic theory as given in Ernest Nagel's seminal *The Structure of Science*, pp. 338–345.[62]

4.1 Weak versus strong

Here, it will suffice to distinguish between two notions of emergence: a weak and a strong version.[80] If there truly is some *insuper et supra* high-level ontology which is genuinely new with respect to lower-level constituents, this will raise important metaphysical questions. In that case, we will be dealing with a strong emergence, as was sought after by British emergentists, like Mill, Broad and Alexander.[81] If strongly emergent entities are to be found, this will entail a revision of how we understand Nature, particularly as it will reject the worldview of the physicalist, who claims that everything (logically) supervenes on physical entities.[82]

Weak emergence is often appealed to in the context of complex systems. Mark Bedau states that a weakly emergent entity has to be "generated from underlying processes" and "autonomous" with respect to these underlying processes.[83] A tornado, for example, seems to be a self-organizing entity forming an agglomerate pattern of air and water molecules that behave independently of the laws governing the air and water molecules themselves.

A high-level phenomenon arising from a low-level domain can be called:[84]

- *weakly emergent*, when "truths concerning that phenomenon are in principle deducible, but *unexpected* given the principles governing the low-level domain."
- *strongly emergent* when "truths concerning that phenomenon are *not deducible* even in principle from truths in the low-level domain."

So, on the one hand, if you can explicitly deduce higher-level properties from the lower-level theory, this cannot be regarded as strong emergence. If these properties are "unexpected", they are weakly emergent properties. On the other hand, you might

[80]Different positions on the concept of emergence can be (and are) constructed that underlie the distinction between the weak version and the strong. As a good approximation however, we will stick to the definition of a weak/strong distinction given by Chalmers below.

[81]Timothy O'Connor, Hong Yu Wong, "Emergent Properties," *SEP* (Summer 2015), Edward Zalta (ed.). URL: `http://plato.stanford.edu/archives/sum2015/entries/properties-emergent/`.[64]

[82]In *The Conscious Mind* (pp. 34–42), Chalmers insists on the distinction between 'logical supervenience' and 'natural supervenience'. Reductive physicalism is based on the idea that all existing things logically supervene on physical laws. That is, when it does not happen that two logically possible situation have the same A-properties but are different with respect to their B-properties, then it can be said that B-properties logically supervene on A-properties. The difference with natural supervenienece is that the sense of 'logical' should be interpreted as independent from the factual natural laws of this world. Roughly, if we can 'imagine' a situation with different laws (a strictly Newtonian universe, for example), it is logically possible. An illustrative position is that of the dualists; for them, reductive physicalism is wrong, and consciousness is naturally supervenient on the physical through some natural law that connects matter with mind, thereby explaining the experienced correlations between them.

[83]Mark Bedau, "Weak Emergence," *Philosophical Perspectives, 11: Mind, Causation, and World* (Oxford: Blackwell, 1997), pp. 375–99, p. 376.[13]

[84]Based on David Chalmers, "Strong and Weak Emergence," *The Re-emergence of Emergence*, pp. 244–56 (Oxford: Oxford University Press, 2006), P. Clayton and P. Davies (eds.), p. 244.[19]

not be able to deduce the higher-level phenomenon, due to practical limitations that stand in the way of deduction, e.g., when a set of equations becomes too complex to be solvable. Nevertheless, one might be able to understand the phenomenon in terms of the lower-level parts. Again, this will be weak emergence, since no higher-level ontology needs to be introduced. This is the case for the migration of the starling: being in a swarm enables the bird to fly much further distances than one might expect.

A Laplacian demon, the creature that has ultimate calculational power and the information about all initial conditions of the fundamental ontology of the world, would be able to do away with weak emergence and simply derive weakly emergent phenomena in a rigorous way. However, this all-knowing creature would not derive strongly emergent properties, and therefore, if strongly emergent properties exist, this creature would not be as all-knowing as you would think. Chalmers suggests that consciousness is a strongly emergent phenomenon.[85] Chalmers argues that it is reasonable to doubt that consciousness can be logically derived from physics, since *(1)* a colorblind person could gather all physical facts in the world, including the function of our brains, about the color yellow, but he will nevertheless be unable to deduce what it is like to have a conscious experience of that quale;[86] and *(2)* it seems to be logically imaginable that the world could have existed in precisely the same physical way as it is now (that is, a precise duplication of the initial values of all spacetime fields and the laws of physics) and yet not contain consciousness. If consciousness were strongly emergent, we would still observe systematic correlations between the physical world and the conscious realm; this is something quite different from direct causal effect. In this case, it is often said that consciousness 'supervenes' on physical facts; this supervenience would be the result of laws over and above physical laws, such as 'psychophysical laws', as Chalmers dubs them.[87]

4.2 Wallacian emergence: (quasi-)classicality and functionalism

On a closer look, if we represent the state on $3N$-dimensional configuration space we see how to support the conclusion that branches will become centered around a particular configuration space trajectory. If we look at the *mean* trajectory (the solution of the Ehrenfest equation) of a decohering branch of an N-particle quantum system through $3N$ configuration space and compare it to the classical trajectory of an N-particle system through $3N$ configuration space, the two are very similar; in fact they are practically indistinguishable.[88] Wallace dubs the former (quantum) structures 'quasi-

[85]There are other candidates for strong emergence, too, arising from identifying quantum entangled structures as non-reducible entities. See Gambini, Pullin, Lewowicz[35].

[86]Frank Jackson, "Epiphenomenal Qualia," *The Philosophical Quarterly* (Apr., 1982), Vol. **32**, No. 127, pp. 127–136.[47]

[87]Chalmers, "Weak and Strong," p. 247.

[88]See Joshua Rosaler, ""Formal" vs. "Empirical" Approaches to Quantum-Classical Reduction," *Topoi* (2015), Vol. **34**, No. 2, pp. 325–338.[70]

classical'. It seems that we can indeed have multiple non-interacting structures that resemble independent classical phase space trajectories. It seems that all's well that ends well.

But appearances are deceptive. How can the mean of the trajectories of a quasi-classical world be the same as that of a classical world? After all, the two theories, Everettian quantum mechanics and classical mechanics, are ontologically very different. The former consists of a $3N$-dimensional configuration space wavefunction $\Psi(q)$, while the latter describes a collection of N particles in 3-space. In fact, the Wigner functions are conceptually not the same as classical Newtonian trajectories, not even approximately; they mimic (approximate) classical Liouville flows (under the appropriate initial conditions).[89] Even if they turn out to be *dynamically equivalent*, are the two structures that appear really identical? And are these two separate things or one and the same thing?

Here, I will discuss whether the emergence of quasi-classicality is weak or strong. We need to keep the following distinction in mind. There is a quantitative question about the identification of quasi-classical trajectories with the trajectories of classical mechanics. For this quantitative part, instantiation can be employed in the way Wallace does this as well. The novel part is the qualitative question: how do classical trajectories could emerge from quantum mechanics?

Furthermore, the costs for Wallace to interpret quasi-classical configuration space trajectories as the emergence of a real three-dimensional world are analysed.[90] These costs come along with the use of Dennett's criterion, and they blur the lines between weak and strong emergence.

4.2.1 The emergence of quasi-classical world-like structures

We have seen that the decoherence process forces a localization in a particular basis: the position basis. Also, the configuration space trajectories of the branches of the wavefunction $\langle\psi(t)\,|\,q\rangle$ are approximately the same as the configuration space trajectories of a three-dimensional classical many-body theory. These trajectories are – for all practical purposes – indistinguishable. Here is an example which I use to illustrate the instantiation identification in configuration space: a trajectory of a classical particle $\tilde{e}$ is said to be instantiated in configuration space by the wavefunction. If $(A_D \rightsquigarrow B)$, we can write an instantiation relation for some histories within quantum theory (wave-

[89] I have not been able to find studies on the evolution of an initially delocalized Wigner function in a nonlinear potential well. The question if such a state will localize on physically possible timescales or not will shed much light on the debate in this paper, and if there is a domain where localization does not occur as a result of decoherence. One can then consider experimental set-ups. One suggestion I would like to make is to look into non-ohmic diffusion domains, where decoherence might be tamed—even for low temperatures.

[90] One of the differences between 'quasi-classical' and 'classicaly real' is that it is ontologically simply not true that one robust block in the abstractness of configuration space *is* a Euclidean three-space.

function in theory A) with some histories (classical particle in theory B)

$$\langle \psi \mid q_i \rangle \longrightarrow q_{\tilde{e}}(t), \tag{4.1}$$

and the same goes for all other quasi-particles (the quantitative pattern within quantum mechanics that instantiates the behaviour of a particle in classical mechanics). Nevertheless, although the quantitative patterns of the left-hand and the right-hand side here approximately coincide in the configuration space, we already see a conceptual mismatch here. When the particles are indistinguishable, the indices of the quantum formalism do not correspond to the labels of the emergent quasi-classical particles. The labels in the quantum formalism are merely formal, not corresponding to anything 'particulate' like the particles we know from classical mechanics.[91]

There is only one world, one universal quantum state, but all our perceptions are very nearly contained around one trajectory of this state. That part of the universal state that is centered around a basis vector gives rise to trajectories in the configuration space that (quantitatively) resemble those of many classical particles, in the same way as shown by Eq. (4.1). The evolution of these trajectories behave (approximately) as an agglomerate of classical particles would, shielded from the other parts of the universal wavefunction by the decoherence process, which ensures these branches will be approximately non-interfering, as treated in Section 3.2. What emerges, then, is a certain autonomy of these bundles of trajectories (due to decoherence).

If one is unaware of the process of decoherence, such configuration space trajectories are clearly 'unexpected' in the sense of Bedau. Why, at first sight, would the branches be autonomous with respect to each other when we pick the position basis? This can hardly be foreseen without explicit calculation. But now that these calculations have been made, we have derived that quasi-classical particles do indeed arise from quantum physics. Following either Chalmers or Bedau, we come to the conclusion that these quasi-classical structures are *weakly emergent*, since it is both unexpected and quite autonomous from the behaviour we know from the fundamental Schrödinger equation, as the quasi-particles act as self-organizing entities which evolve quite independently from the rest of the wavefunction. Bedau makes clear that an emergent entity has to be generated from underlying processes. For example, when it comes to emergence, Wallace's quasi-classical worlds do not seem all that different from the self-organization of tornadoes. In this sense, the definiteness of a macroscopic world is an emergent property of quantum physics in the same sense that temperature and pressure are the emergent properties of statistical physics. For the latter it is the underlying ensemble of definite positions and momenta of particles that instantiate thermodynamics; for the former it is the unitary dynamics of quantum states that instantiates configuration space trajectories that behave in the same way as classical trajectories.[92] Thus,

[91] Dennis Dieks & Andrea Lubberdink, "How Classical Particles Emerge From the Quantum World," *Foundations of Physics* (2011), Vol. **41**, No. 6, pp. 1051–1064.[27]

[92] In statistical physics it is also the dynamics that gives rise to thermodynamical properties, here, too, a degree of approximation is permitted (Maxwell's demon) and a case for the instantiation of

at least analogously to properties we usually regard as emergent, quasi-classical worlds are weakly emergent entities. We humans, as Everettian observers, emerge together with such a world, and are effectively confined to observe only what is inside of it.

This conclusion is both incredible and incredibly anti-climactic. It is incredible because it indeed gives us a dynamical result that is the many-worlds theory, without making additional assumptions about the world (apart from quantum state realism). We can justifiably speak of an 'emergent multiverse'. Or, more precisely, we should speak of 'many emergent universes', since we already started from a universal wavefunction, i.e. it is not this multiverse itself that emerges, it is rather the *autonomy* of the branches in the decoherence basis that is emergent.

The anti-climactic part of the conclusion is with respect to our analysis of Wallace's presentation of the many-worlds theory. Why did Wallace take pains to write his second chapter, where he is clearly appealing to pragmatic virtues stated in Dennett's criterion (to be discussed in the next sections), while the pattern of definite macroscopic quasi-classical worlds is an objective feature of the formalism, ready to be interpreted as weak emergence? According to Wallace, the goal of his second chapter is to answer the question "Why does 'quantum mechanics, taken literally' entail that we live in a multiverse?"[93] But a world, or an entire universe, is conceptually not so different from other macroscopic objects. In that case, we have established the answer to the question of Wallace's second chapter, but it was sufficient to use the analysis of the third. Chapter 2 and its appeal to the pragmatic features of Dennett's criterion, I claim, serves a separate purpose, which, as we will discuss in Chapter 6, seems to have origins in a realism that goes deeper than quantum state realism.

4.2.2 The emergence of real classical worlds: what it takes

So far, we have established that when we look at the trajectory of one decohering branch of the universal wavefunction (superposed in the position basis) through $3N$ configuration space and compare it to the classical trajectory of an N-particle system through $3N$-configuration space, the two are structurally very similar, even practically indistinguishable. These approximately classical structures can then be said to weakly emerge from the quantum dynamics, but this is something else entirely than the ontological claim that real classical worlds emerge from the wavefunction. This ontological claim amounts to dropping the prefix 'quasi' from quasi-classical, but, in fact, the ontology of the quasi-classical structure is wildly *unclassical*. Classicality is the property of definiteness, or completeness, which is the idea that physical entities have well-defined definite characteristics, independent of contextual circumstances such as 'being observed' or not.

Wallace nevertheless makes the ontological claim. To achieve this, he uses a functionalist criterion. The argument is that, even if something could be ontologically very different, when it behaves in entirely the same way as something else, such that it

---

thermodynamics by statistical physics can be made, i.e. {statistical physics $\rightsquigarrow$ thermodynamics}.

[93] Wallace, *Emergent Multiverse*, p. 45.

is indistinguishable from it, it is truly the same as that something else. If it walks like a duck, and it quacks like a duck, and is in all respects indistinguishably behaving as a duck, then it definitely is a duck: a *real* duck (or, in Wallace's words, "a tiger is any pattern which behaves as a tiger."[94]) In a similar vein, the structure of a decohering branch of the universal wavefunction in configuration space is a structure which behaves the same as a structure of a classical particle in configuration space would.[95] Thus, since a quasi-classical particle is instantiated by a pattern that acts – for all practical purposes – as if it is classical, a quasi-classical particle *is* a classical particle.

We learn from this that from the definition of instantiation given by Wallace (see Eq. (3.5)), it can be said that {quantum theory$_D$ $\rightsquigarrow$ quasi-classical worlds$_D$} simply from the formalism of quantum mechanics.[96] But to claim that {quantum theory$_D$ $\rightsquigarrow$ classical worlds$_D$} one also need to side with the functionalists and adopt the doctrine 'two structures that are practically indistinguishable are identical'. The quasi-classical structures are quantum-mechanical structures that are responsible for instantiating the classical structures. However, is this instantiation enough to warrant the identification of quasi-classical and classical structures? This is what the emergence of classicality takes.

One can reject that a classical world should in fact be recognized. The pattern that emerges is not what we understand the structure of a classical world to be; after all, the structure represents a quasi-classical structure, practically indistinguishable from a classical structure. The classical world is a definite world, which is composed of classical particles in a three-dimensional space like the one we observe around us. Thus there is room for two positions: *(i)* the functionalist who argues that the emerging structure of a classical world *is* in fact a classical world, since it cannot be distinguished from one; and *(ii)* the anti-functionalist who holds on to the ontological difference between an actual classical world and a structure that merely resembles one, thereby honouring the prefix 'quasi'.[97]

However, I believe this is one step too far in the analysis; we should in fact start by questioning why we are searching to identify what is a 'world' at all. I will argue that it all depends on the context created by the question that is being asked. If you are interested in reducing classical physics to quantum physics through the mechanism outlined in Section 3.2, then you could appeal, as Wallace does, to functionalist criteria.

[94] Wallace, "Structure," p. 93.

[95] It is important to note that the identification between the quantum and the classical is made in the mathematical configuration space, which is *not* the fundamental three-dimensional space of classical physics nor the fundamental Hilbert space of quantum physics. We will come back to this point. See Eddy Keming Chen, "Our Fundamnetal Physical Space: An Essay on the Metaphysics of the Wave Function," *The Journal of Philosophy* (2017), Vol. **114**, No. 7, pp. 333–365.[22]

[96] The domain D captures only those quantum histories in which interaction Hamiltonians are of the right form for the decoherence story to work.

[97] From personal correspondence I have learned that David Wallace does not feel too strongly about the prefix 'quasi-'. However, he is concerned with the functionalist step, while I am using the quasi-prefix to denote that step. So although in Wallace's book the words 'quasi-classical' and 'classical' are nearly interchangeable, in this thesis they are not.

If one asks, though, how it has come to pass that we use the predicate 'classical' the way we do, we can readily see – keeping in mind the empirical vindication of quantum physics over classical physics – that we have assumed the wrong ontology for a pattern that we thought was classical, but is rather a part of a larger quantum-mechanical whole. In this case, no appeal to functionalism is necessary (although questions about the preferred basis might still linger, as we will see in the rest of this thesis).

As a concluding remark, we have seen that quasi-classical worlds emerge weakly from the underlying Schrödinger dynamics in the decoherence basis. However, some extra work has to be done to confirm to Wallace's *desideratum* that these quasi-classical worlds should be identified as real classical worlds. The functionalist criterion that Wallace uses to achieve this is what he calls 'Dennett's criterion'. As we will see in the next section, Dennett's criterion helps Wallace to solve the preferred basis problem by highlighting the classical basis (which coincides with the decoherence basis at the coarse-grained level), but it seemingly also introduces additional metaphysical baggage on top of the analysis of decoherence. Due to this baggage, the demarcation between weak and strong emergence becomes somewhat blurry, as it does not seem to be the case that a Laplacian demon would go along with Dennett's criterion and prefer the classical basis, since there is no objective reason to do so. Hence, Wallace's position might tip the balance towards the strong emergence of real classical worlds.

5 Dennett's criterion: a functionalist's framework

Wallace seeks to establish a deeper reality of the decoherent branches we have called 'worlds', and this is where Dennett's criterion comes in. Daniel Dennett's 1991 paper "Real Patterns"[98] is concerned with explicating the idea of "a mild and intermediate sort of realism"[99] that he proposed in his book *The Intentional Stance* (1987). Dennett's idea is to appeal to the concept of a 'pattern', where Dennett is concerned with the 'reality of beliefs' in the theory of folk psychology. Dennett argues for an intermediate sort of realism, in between the Scylla of the eliminative materialist's claim that beliefs are mere illusions and the Charybdis of a full-blown realism that argues for a direct correspondence between brain states and beliefs. Wallace's goal is to employ a similar sort of intermediate realism in order to establish the existence of the decohered branches of the universal wavefunction, patterns within the mathematical formalism, as actual worlds. Recall Wallace's formulation of Dennett's criterion:[100]

> A macro-object is a pattern, and the existence of a pattern as a real thing depends on the usefulness – in particular, the explanatory power and predictive reliability – of theories which admit that pattern in their ontology.[101]

When confronted with Dennett's criterion, a set of questions imposes itself. What work does Dennett's criterion do in Wallace's approach to quantum mechanics? *Prima facie*, this seems to be the Wallacian turn that I have talked about in the previous sections: moving from something like 'collective trajectories of a classical N-particle system approximately obey dynamical equations of the same mathematical form as those obeyed by quasi-classical branches' to 'classical branching worlds exist as higher-level entities within the universal wavefunction'. This is mostly along the lines of the decoherent-histories approach by Gell-Mann and Hartle, who took this mathematical resemblance as sufficient proof of them being the same.[102] Although the criterion itself still needs to be justified independently, I think this Wallacian turn is indeed achieved by Dennett's criterion. But it also appears to do more work than the promotion of quasi-classical trajectories to classical worlds—it imports extra metaphysical baggage, which will be the subject of remaining chapters.

5.1 Discussion of Wallace's Dennett's criterion

Surely, when you are in a position to choose between several options of patterns to use, you want to use that pattern that is most useful to meet your goals. From that

[98] Daniel Dennett, "Real Patterns," *The Journal of Philosophy*, Vol. **88**, No. **1** (Jan., 1991), pp. 27–51. URL: http://www.jstor.org/stable/2027085.[28]

[99] Dennett, "Patterns," p. 29.

[100] David Wallace, *Emergent Multiverse*, p. 50.

[101] I am not aware of Dennett's endorsement of Wallace's definition.

[102] Murray Gell-Mann & James Hartle, "Classical equations for quantum systems," *Physical Review D* (1993), Vol. 47, No. 8.[39]

perspective it seems to make sense to choose the quasi-classical pattern. Indeed, the quasi-classical pattern *is* useful because it can be used to deterministically predict unique outcomes of experiments. Clearly, classical determinism can make more complete predictions than mere probabilities (disregarding chaotic behaviour). In the words of Żurek, "classical reality can be regarded as nearly synonymous with predictability."[103]

Utility and predictability are key values of the pragmatist, but it is interesting to note that Wallace claims not to be one: "I want to stress that talk of 'explanation' here is not meant to imply that it is just a *pragmatic* matter whether or not, for example, we take seriously tigers and other such macro-objects."[104] Here, the emphasis is on 'explanation', which is used in conjunction with predictability, and classified as something 'useful'—and this is established through Dennett's criterion.

I think it is important to point out that using a conjunction of explanatory power and predictability under the umbrella of usefulness is not as self-evident as it is presented here. While predictive power is universally agreed to be a goal of science, explanation is not. For realists, explanation is provided through universal and fundamental laws, of which the unitary evolution of quantum mechanics is clearly part. Scientific realists often appeal to the most explanatory theory, which in return receives the special status of being real.

Wallace's claim is that his interpretation is "just quantum mechanics itself," taking the quantum formalism "at face value" as his only assumption.[105] But Wallace's claims go further: he wants to make a reality claim about 'actually existing' worlds. That is an aim over and above any mechanism described within quantum mechanics. This is the gap that is supposedly to be filled by emergence, but this emergence relation itself is not completely explicated.

A further question is about the central notion of 'pattern' here. Elsewhere in physics we speak of all kinds of patterns, but do not usually appeal to something as involved as Dennett's criterion to establish the reality of these patterns. We seem not to be in need of this criterion if we want to argue for the reality of mountains or stars. Why do we need this criterion in the particular situation for 'worlds'?[106] Is Wallace's notion of 'real patterns' really the same as Dennett's, as suggested by the name? And is Wallace's notion of real patterns doing what he wants it to do? This last question is related to the central question of this research that remains important in the background: what is achieved by Dennett's criterion in Wallace's approach to Everettian quantum mechanics and what specific role does it play in solving the measurement

[103] Żurek, "Decoherence and the Transition from Quantum to Classical-*Revisited*," *Los Alamos Science* (2002), No. **27**, p. 21.[90]

[104] Wallace, *Emergent Multiverse*, p. 57.

[105] As he mentions in his introduction (see Wallace, *Emergent Multiverse*, p. 3), the only other philosophical assumption Wallace claims to make is that of (methodological) naturalism, that by studying science we must use the tools of science itself.

[106] Of course, Wallace might argue that what we are implicitly doing in science is employing Dennett's criterion all over the place, time and again, but that somehow people have been reluctant to apply it to quasi-classical worlds.

problem? This question will be central in the remaining part of this thesis, where each section highlights the following important notions of Wallace's 'Dennett's criterion':

- Dennett's functionalist's idea of 'patterns' and its relation to Wallace's approach to quasi-classical patterns in the configuration space (Chapter 5)
- the appeal to the pragmatic virtue of usefulness, specified as explanatory power and predictive reliability (Chapter 6)
- the relation between higher-level and more fundamental theories (Chapter 7)
- it is a criterion to establish the reality of macroscopic objects (depending on such pragmatic virtues) (Chapter 8)

In a sense, the above categorization amounts to several variations on the same theme: achieving the reality status of macroscopic objects through pragmatic virtues of higher-level theories by treating these objects as patterns.

5.2 Dennett's "Real Patterns"

Dennett's concern lies with the ontological status of mental states, like the experience of pain and pleasure, or the acts of planning and hoping. Are beliefs and desires real? The short answer: yes and no. Dennett attempts to find a middle way between eliminative materialism and a full-blown realism about mental states. On one side, eliminative materialism (advocated prominently by Paul Churchland, for example) claims that beliefs or mental states are entirely non-existent. On the other side, there is the view 'industrial strength realism' (attributed to Jerry Fodor), which purports that beliefs are things in the head very similar to biological cells or hormones.

Furthermore, Dennett warrants himself against falling into a metaphysical discussion about the ultimate nature of things, but adopts Arthur Fine's natural ontological attitude. The idea is more or less that we should be satisfied with the reality-status of beliefs once we have determined their reality status relative to things that are taken to be real in science throughout. What Dennett intends to argue for is that mental states are as real as centers of gravity or electrons.[107] The sense of reality is claimed to be the scientific route to reality: some abstract objects, like centers of gravity, "are real because they somehow are *good* abstract entities. They deserve to be taken seriously, learned about, used."[108]

To convince philosophers of his "mild realism", Dennett uses the concept of a *pattern*. The basic idea is that patterns are agglomerates of constituents, and that the patterns that we should take seriously are stable under some disturbance or noise. That is, when we distinguish a 'something' that is composed out of smaller things, we recognize a pattern that need not be exact; the pattern can deviate a bit from

[107] Dennett, "Real Patterns," p. 30.
[108] *Ibid.*, p. 29.

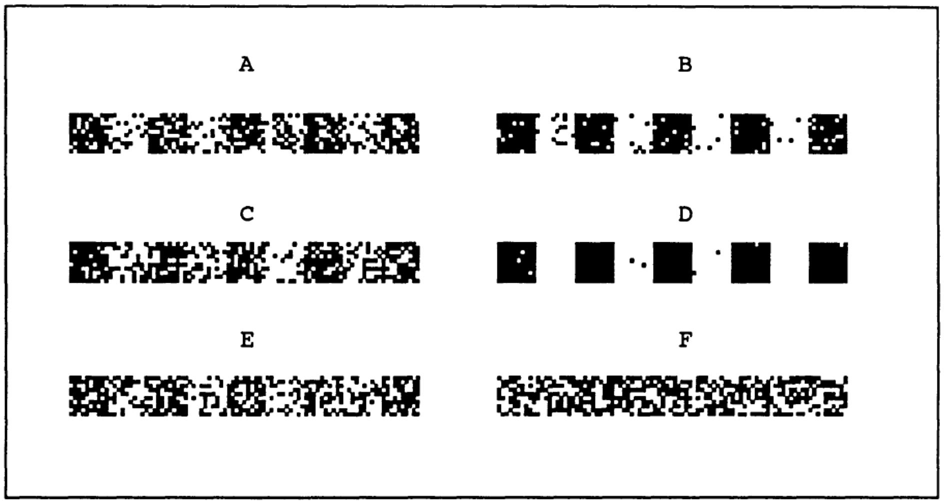


Figure 3: Six barcodes generated by the same algorithm with various degrees of noise; A: 25%, B: 10%, C: 25%, D: 1%, E: 33%, F: 50%. By squinting, one can discern the same pattern in all, except for the fully incompressible noise of barcode F. Figure from Dennett's "Real Patterns," p. 31

the underlying stable thing that it is supposed to represent. In fact, there is a trade-off between accuracy and effort (or computing power). He illustrates this using three examples of patterns: chess programs, Conway's *Game of Life*, and barcodes.

The barcode is the simplest and is sufficient to understand our present concerns. It is a pattern of black and white dots, in an array of ten rows with ninety dots each, where every row is a repetition of ten black and ten white dots. A certain degree of noise is allowed, understood as deviations from the underlying pattern. In the six barcodes of Figure 3, barcode D shows the pattern with only 1% noise, while barcode F shows the pattern with 50% noise. Dennett's argument is that in order to communicate at the level of accuracy of barcode A (for example), it is a lot of work to send information about the exact bitmap (distribution of the dots in the array). Instead, one can simply send the program itself plus the information that the noise has been set to 25%. The reason that we can compress the data is because we are not interested in the exact distribution of dots, but in the underlying pattern. For all practical purposes (!) A and C are the same pattern, which we can then 'functionally identify' with each other. Another functional identification is to say that barcode F, although it is generated by the same algorithm that generates the pattern that it is clear in barcodes A to E, is *indistinguishable* from a random distribution of black and white dots over an array of the same size.

In one sense, functionalism disregards the ontological status of these patterns. That is, what these patterns exactly are, how they are made, or how they are generated, is not important—only the pattern itself matters. In another sense, closer to what Dennett intends, we can *define* the ontological status of the patterns on the basis of their functional purpose. The intentional stance is just that: to regard mental beliefs and desires as patterns or structures in natural systems. Recognizing beliefs coincides with a recognition of patterns that correspond to these beliefs.

During the nineties, Dennett's 'Real Patterns' created much turbulence in the philosophical lake, and it is not my intent to wade through all of it here. Only two elements will be important for our discussion. One of them is that the patterns that correspond to beliefs become salient *only* from a certain perspective, namely the perspective of the agent that is recognizing the pattern. Two individuals may discern quite different patterns from the data in front of them, if those individuals have different goals or standards of accuracy. That is not to say, though, that the pattern is not there when there are no agents in existence; it is only to say that the pattern is not more important in that case. In other words, the pattern is not completely dependent on the observer; Dennett insists that a pattern "exists in some data - is real - if there is a description of the data that is more efficient than the bit map, *whether or not anyone can concoct it*."[109] Nevertheless, human interests exert *some* influence on the salience of a pattern, namely in the choice of highlighting it at the expense of other patterns.

Hence, Dennett has a perspectival view of intentional psychology, with a rather subtle mix between epistemic and ontological components. Huw Price has identified this with functional perspectivalism. The idea is that one has to start from the perspective of a certain creature, and the function of a certain language or discourse is one which only becomes useful in that perspective. Unless that perspective is in fact occupied, one lacks the ability (the 'function') of making use of that discourse.[110] He defends Dennett against those who claim that Dennett's position detaches psychology from the hierarchy of science and therefore relies on an anti-realism about intentional states. Price indicates that this attack relies on a category mistake: it judges psychological discourse by a standard which ought never to have been applied to it.[111]

A second important part of Dennett's position is that a pattern can correspond to something real *if there are underlying principles to be exploited*. In the patterns of our behaviour, Dennett's intentional stance intends to find the 'real' beliefs. Can certain patterns be 'unreal'? Yes, according to Dennett:[112]

> If we go so far as to distinguish them [the abstract entities] as *real* (contrasting them, perhaps, with those abstract objects which are *bogus*), that

[109] Dennett, "Real Patterns," p. 34. *My emphasis.*

[110] Huw Price, "Psychology in perspective," in: *Philosophy in Mind: The Place of Philosophy in the Study of the Mind*, M. Michael and J. O'Leary-Hawthorne (eds.) (Kluwer, 1994), pp. 83–98.[67]

[111] This part, that patterns can only be recognized from a particular stance, even though the underlying pattern is objective, will be taken up in Chapter 6. It seems far-fetched that Wallace might argue for a detachment of classical physics from the hierarchy of science, though. This is one of the reasons why Wallace's aim is somewhat different from that of Dennett. In particular, an analogy will be drawn with highlighting patterns in van Fraassen's 'web of science' through the lens of demanding classical explanation. This I will call the 'classical stance'.

[112] Dennett, "Real Patterns," p. 29. *Original emphasis.*

> is because we think they serve in perspicuous representations of real forces, "natural" properties, and the like."

Hence, there is underlying objective behaviour that can be appealed to for a specific pattern to be discerned: it is in that sense that the patterns exist.

5.3 Worldly patterns

It is perhaps remarkable that one encounters something called after Daniel Dennett in a work about solving the measurement problem in quantum mechanics. After all, Dennett is a philosopher of mind, active in the fields of artificial intelligence, neuroscience and the question of consciousness. What is the link with quantum mechanics?

Wallace uses his version of 'Dennett's criterion' as a reality condition to establish that quasi-classical worlds, approximately behaving as classical worlds in the configuration space, are *real*, because they are useful patterns that allow us to predict and explain in a classical vein. The idea must be something like this:

> A quasi-classical world is a pattern, and the existence of a worldly pattern as a real classical world depends on the usefulness – in particular, the classical explanations and definite predictions – of classical physics which admits the classical pattern of a world in its ontology.

Clearly, Wallace's concepts bear great resemblance to those of Dennett, but at the same time it is not immediately clear if we should equate the existence of quasi-classical worlds with the same kind of existence of Dennettian beliefs. Is Wallace's notion of 'pattern' identical with Dennett's? The patterns that Wallace is interested in are those patterns that arise in the configuration space with the wavefunction as its fundamental ontology. As mentioned in Section 3.2, Wallace gives an analogy with atoms that move around in space and form a pattern that corresponds to a tiger; when we observe it, we usually identify such a pattern as an actual tiger. Hence: "a tiger is any pattern which behaves as a tiger."[113]

But why should we take seriously the structure of a tiger and not, say, the pattern of the tiger together with the tree it is sleeping in? The answer has a subjective and an objective component. Clearly it is useful to recognize the structure of the tiger, for survival-mode purposes. If we humans were not equipped to recognize quickly the danger that this predator poses, our species would not have gotten very far on the evolutionary time-line. Danger, though, is quite a subjective motivation for pattern-recognition. The atoms by themselves are just atoms, mere jiggling marbles. By themselves they are just themselves and do not explain much. As Wallace remarks, quantum mechanics tells us very little about what will happen to Schrödinger's cat when the atom decays, since we "cannot possibly determine the microphysical state of the cat [...]; even if we could, we could not solve the Schrödinger equation for such a complex system; even if we could, all it would give us is *a bare prediction of the microphysical state* [...]."[114]

[113]Wallace, "Structure," p. 93.

[114]Wallace, *Emergent Multiverse*, p. 59. *My emphasis.*.

On this quantitative matter about the bare prediction, Wallace's ideas seem to be very much in line with Dennett's. Dennett points out that we need something more than the bare prediction in order to *predict* what people will do in the future, namely folk psychology. Interpreting other people's behaviour is possible because we can predict that behaviour using higher-level theories: "If we wouldn't have it, human activity would be just so much Brownian motion."[115]

Patterns of atoms – like the union of the tiger and the tree – can be unreal, or 'bogus', in the Dennettian sense that they are not stable patterns. Such a pattern is, in the Dennettian sense, 'bogus', because it is not a 'perspicuous representations of real forces'. Surely, we have chosen to identify certain objects in particular ways because we want these objects to be useful. We cognitively identify tigers, trees and cups of coffee as particular objects as opposed to raw pixels of sense data. A tiger sleeping in a tree constitutes two objects, because it is useful to explain the autonomous behaviour of these two objects. When the tiger goes hunting, the patterns of the tiger and the tree come apart in a highly autonomous way. However, at the level of the atoms, any agglomerate of atoms containing parts of both the tiger and the tree is real. Surely, if we focus only on the atoms of the leaves of the tree and the tail of the tiger, it forms an agglomerate that is just as real as the agglomerate of atoms that make up the tiger. Then, the tiger is real as an agglomerate, but not (automatically) real as a macro-object, because it is no macro-object: real macro-objects are *useful* agglomerates.

Notice that in the above it is almost impossible to circumvent 'tiger-talk' and 'tree-talk', because these are the concepts that are being used at the higher-level of macroscopic things explained by higher-level theories such as zoology and ecology. We do not have one single concept 'tigertree'[116]; we only have the concept 'tiger' and the concept 'tree'. It would not be fruitful to argue that these classifications are arbitrary. Clearly, there are objective facts of nature that ensure the degree to which what we call 'tiger' and 'tree' are autonomous. Indeed, it is this objective autonomous behaviour that makes it reasonable to construct theories about them and it is the idea that we can identify and study objects in isolation on which the scientific enterprise is founded.

Furthermore, Wallace and Dennett do not seem to fully coincide in their view on how dark the shade of the 'real' in 'real patterns' should be painted. Dennett describes his position as a somewhere between realism and instrumentalism:[117]

> in the root case a pattern is "by definition" a candidate for pattern recognition. (It is this loose but unbreakable link to observers or perspectives, of course, that makes "pattern" an attractive term to someone perched between instrumentalism and industrial-strength realism.)

Wallace seems to make a further claim about the success-rate of theories and the reality that they describe. He does not seems to argue for the interpretation of worlds to be

[115] Dennett, "Real Patterns," p. 29.

[116] Although there appear to be several fashion lines that carry the name.

[117] Dennett, "Real Patterns," p. 32.

Figure 4: Raw sense data. Many patterns can be identified here, but only a few of them are useful. The 'tigertree' pattern is useless for most purposes. But there are much more patterns to be recognized: take an arbitrary (non-random) filter in photoshop that changes the picture beyond recognition. The resulting image identifies an actual pattern that is there in the original, but which is completely useless. (Picture from suwalls.com)

"perched between instrumentalism and [...] realism." He seems to be after real classical worlds in the realist sense of being emergent.[118]

Wallace shares the above sense of 'usefulness', namely that there is some underlying physical process that gives rise to some stability in the patterns. More specifically, the underlying objective physical process is decoherence, which gives rise to stable autonomous patterns at the higher-level (in a specific basis). When Wallace's formulation of Dennett's criterion is read to the letter, the reasoning is a little different. One can cook up some counter-arguments to challenge the wat it is formulated.

In the light of this pessimistic meta-induction, it seems that the appeal to useful theories is, *prima facie*, not the right way to establish the objective existence of macro-objects. Many theories have been useful in the past, but have been discarded as wrong. Flogiston theory, geocentrism, energetism, were all useful once. They predicted the behaviour of certain parts of nature, and they had conceptual framework that explained what was going on. Nevertheless, they are wrong. The pattern that Wallace speaks of is not the substance flogiston itself, but rather some real structure that by approximation can be described by appealing to the substance flogiston. This argument only goes so far, however. Quantum mechanics could of course turn out to be wrong, and hence patterns of quasi-classical worlds would not be real, despite the usefulness of classical

[118] This is as far as I want to go into the relation between usefulness and reality that Wallace seems to be relying on here. I will take up this question again in Chapter 8.

mechanics. Assuming that quantum mechanics is right, though, Dennett's criterion is used as a criterion to distinguish between those patterns that are real and those patterns that are bogus.

Following this literal reading of Dennett's criterion further, we see that where Dennett speaks of useful patterns, Wallace deals with useful theories (that can also admit useless patterns). This brings us to an important question: *can there be unreal patterns in Wallace's view*? An unreal pattern according to Wallace would amount to a pattern that is admitted into the ontology of theories that are not useful, which do not predict well and use concepts that refuse to explain anything to us. Furthermore, if one sticks with a literal reading of Dennett's criterion: take any pattern and take any useful theory T_1. Extend the theory to T_2 to incorporate that pattern. Arguably, T_2 is still useful such that the pattern is real. Again, this argument is not meant as decisive; the defender of Dennett's criterion can argue that the added structure is not useful and one should not be committed to it.

For a literal reading of Dennett's criterion to make more sense, we could turn it around. The assumption that seems to play in the background of this criterion is that usefulness and reality are directly related. Maybe Wallace wants to hold on to some bottom-up approach: that in some (admittedly far-fetched) counterfactual history where quantum theory was developed before the classical theory (and its formalism taken at face-value by wavefunction realists) we would have deduced (loosely speaking) classical mechanics from that quantum formalism. The reason being that there are specific patterns that are useful and deserve to be taken so seriously that we should develop theories about them. When the useful patterns already exist, the theory describing them becomes useful in turn.

At first sight, the idea that we would have deduced classical mechanics from the quantum formalism may not sound very shocking. In the light of the preferred basis problem, however, I believe it nevertheless should be shocking. A top-down approach seems to be taken that is needs explication. Quantum theory by itself allows for many patterns that do not correspond to macro-objects. *It only gives rise to useful macro-objects in the classical basis*.[119] Hence, there is a 'classical stance' that needs to be occupied before one can discern the useful patterns in the universal wavefunction. This stance is not dictated by the quantum theory.

[119] Throughout this research, the classical basis will be identified with the basis that approximately coincides with the decoherence basis. There is some unclarity to what extent Wallace himself has this view. On the one hand, he seems to use the two interchangeably, as if they are the same thing. On the other hand, he mentions that in chaotic systems it is not always so clear that stable quasi-classical patterns will emerge; hence, that could mean that there is a decoherence basis, but no classical basis. It is one of the goals of this thesis to show that the decoherence basis by itself is not enough to establish a preferred basis. In addition to decoherence, the classical basis also needs to be stipulated.

6 Pragmatics

"As in all explanations, the correct answer consists in the exhibition of a single salient factor in the causal net, which is made salient in that context by factors not overtly appearing in the words of the question."

— Bas van Fraassen, 1980. [120]

In Chapter 4, I have argued that the emergence of quasi-classical world-structures, is of the weak kind. That is, a Laplacian demon, would have no trouble deriving these structures. When the universal wavefunction is decomposed in the decoherence basis, Wallace argues that one can see the emergence of real classical worlds, instead of simply their structures. Mathematically, this relies on a functional identification between an N-particle classical system in three-space and a course-grained behaviour of a substructure of the wavefunction projected onto $3N$-dimensional configuration space. In the previous section, I have discussed some specifics of this functional view which are encoded in Wallace's formulation of 'Dennett criterion'.

One of the interesting components of this criterion is its reliance on the pragmatic values of explanatory power and predictive reliability of theories. Incidentally, this relies on the usefulness of theories instead of the usefulness of patterns themselves. That amounts to a subtle difference between Dennett's views and Wallace's. This indicates a further goal by Wallace that is directed at the hierarchy of scientific theories, which I shall elaborate on in Chapter 7. Nonetheless, when it comes to the appeal to pragmatic virtues *per se*, Wallace's view can be construed analogously to that of Dennett. However, Wallace does not discuss the analogies between him and Dennett that much—although he does give a spectacular amount of rich and well-explained examples of functionalist identifications that occur in scientific practice. In particular, the consequences of his functionalist criteria for how we should look at the world are not explicated fully, while there certainly are far-reaching consequences for one's conceptions about ontological and epistemological matters, of which functionalism is a subtle mix. After explaining van Fraassen's pragmatic theory of explanation, I will draw an analogy between pattern-recognition in van Fraassen's web of science and Wallace's configuration space patterns. Concretely, my claim is that if Wallace intends to apply Dennettian functionalism to worlds, a specific philosophical position analogous to Dennett's intentional stance should be taken: the classical stance. On the one hand, when seen from the viewpoint of the wavefunction realist (or even in the weaker sense: from the formalism), the classical pattern is not salient at all and the choice of the classical basis is entirely subjective. On the other hand, I give an anthropic argument why the classical stance can be regarded as more objective than it seems.

[120] Bas van Fraassen, "Chapter 5: The Pragmatic Theory of Explanation," *The Scientific Image* (Oxford: Oxford University Press, 1980, Reprinted 2013) , p. 132.[74]

6.1 Van Fraassen's pragmatic theory of explanation

Lets turn to the question of pragmatics. In the philosophy of language pragmatics is about topics that are context-dependent; it deals with statements that derive their meaning from the context which gives rise to this statement. If I make the statement 'sugar consists of oxygen, carbon, and hydrogen atoms', this statement is true independent of the given context. It is true now and it is true 24 hours ago, whether uttered in Utrecht or uttered in Istanbul, whether *you* say it or *I* say it, and it is even true regardless of whether sugar actually exists in this world or not. Conversely, some claims are context-dependent, e.g., the statement 'I eat sugar', which has quite some different meaning whether you or I say it, when you or I say it, whether eating sugar at the moment or referring to a weekly diet.

According to van Fraassen, scientific explanations are very similar. Scientific explanations, as all explanations, do not stand alone; they require certain facts about the context in which we ask explanation and in which we offer explanation. Without a specification of contextual factors, an explanation is an attempt to clarify what 'yesterday' means without appealing to a time of uttering that word. Thus, according to van Fraassen, explanations are not offered by science in the form of relations between theories and facts, but are rather three-place relations between theories, facts, and contexts. So a theory explains a fact relative to a certain context and an alteration of this context might change whether or not some theory can account for some particular fact.[121]

Instead of the early twentieth-century view that explanations were a straightforward two-valued link between theories and facts,[122] van Fraassen demands that explanation is a three-way relation between theory, fact, and context. In a nutshell, an explanation is always an answer to a 'why-question', which is determined by *(i)* a topic, which is the proposition about the fact which we seek to explain; *(ii)* a contrast class, which is a set of propositions that contrasts alternatives to the topic; and *(iii)* a relevance relation, which signals a requested reason for that question to be worthy of asking.[123] For example, I can ask the question 'Why did Wallace write chapter 2?'. The topic would be 'that Wallace wrote chapter 2'. The contrast-class includes varying alternatives to which the topic can be contrasted, like 'Wallace *composed* chapter 2', 'Wallace wrote *chapter 3*', or '*David Deutsch* wrote chapter 2'. The relevance relation appeals to a reason which is relevant for the topic, where 'because Wallace thinks chapter 2 is important' is irrelevant, since it is naturally understood (that is, such an answer would probably not satisfy the questioner).

Through his theory of explanation, van Fraassen hopes to solve two problems.

[121] Van Fraassen, "Chapter 5". And Bas C. van Fraassen, "VII. The Pragmatics of Explanation," *Americal Philosophical Quarterly* (April 1977), Vol. **14**, No. 2, pp. 143–150.[73]

[122] See, for example, Rudolph Carnap, "Chapter 1: The Value of Laws: Explanation and Prediction," in: *An Introduction to the Philosophy of Science* (New York: Dover Publications, 1995). Martin Gardner (ed.).[18]

[123] Van Fraassen, *Scientific Image*, pp. 141–153.

First, there is the problem of the asymmetry of explanation. Why can some fact explain another, while the converse fails. The fact that it is raining can explain the fact that my hair is wet, but the fact that my hair is wet is not accepted as an explanation of the rain. The second problem is that in everyday use, some explanations, although perfectly valid when seen as a two-way relation, are rejected. According to van Fraassen, the relevance relation is tied to the context and hence it is not an intrinsic property: different contexts can redirect the arrow of explanation.

Van Fraassen intends to divorce explanation from science. For him, science itself contains no explanations, but (merely) consists of a set of causality claims relating certain facts. *If* a certain ratio of oxygen, carbon, and hydrogen atoms form bonds,[124] *then* a certain sugar molecule is formed. *If* I strike this match, *then* a fire will start. An explanation makes use of such empirical laws, of which the most fundamental (and universal) laws are the principal aim of science to construct, forming a causal web between facts in the world. In van Fraassen's own words,

> no account of explanation should imply that we can never give an explanation—and to describe the whole causal net in any connected region, however small, is in almost every case impossible. So the least concession that one would have to make is to say that the explanation need say no more than that *there is* a structure of causal relations of *a certain sort*, which could *in principle* be described in detail: the salient features are what picks out the 'certain sort'.[125]

Thus, there are many possible causal relations leading up to an event. When a bus driver pulls his bus over to pick up a passenger, what is the cause of the bus pulling over? The driver would claim 'because I push this breaking pedal here', while the passenger argues that it is 'because I raised my hand, signaling I wanted to get on the bus', and the enigneer would start explaining the mechanism of the air break systems. The *salient cause*, i.e. the relevant cause to successfully answer why-questions, is therefore relative to whom you ask, due to the background ideas of this person, as well as his entire state of being at the time of asking. Explanation, hence, is context-dependent; and science is used as a tool to provide explanations, in the form of a web of causal relations: the larger the web, the more possible causes there are at your disposal in order to explain. As for determining the salient cause (or *the* cause), this is simply not on the to-do list of science. Connecting this to van Fraassen's antirealist position, he has used this to argue that explanatory power is not a reason to accept a theory as true. The only reason to accept or reject a theory is its empirical adequacy.[126]

[124] *Specific* bonds, namely those with the molecular structure of sucrose (presumably, there are possible bonds of $C_{12}H_{22}O_{11}$ other than sucrose).

[125] Van Fraassen, *Scientic Image*, p. 124. *Original emphasis.*

[126] At a conference in Utrecht, I once asked van Fraassen if he nevertheless can see any small role to play for explanation as an objective component of science. He thinks it has not: explanation is purely anthropomorphic pleasure, as outlined by Alison Gopnik's "Explanation as Orgasm"[38].

We need not follow van Fraassen's empiricism to use his theory of explanation in analogy with Wallace's highlighting of the quasi-classical pattern in configuration space. We are confronted with an important question: is the quasi-classical pattern objectively salient, more real independent of ourselves, or is it subjectively salient, analogous to van Fraassen's explanatory highlighting of causal patterns?

6.2 Salient patterns: 'the classical stance'

In a web of quantum patterns, we pick the quasi-classical pattern over the other possible patterns that we can identify, and we need to justify this pick. There are infinitely many other patterns, corresponding to the 'branches' (if we can still call them that) given by the basis vectors in a rotated basis. Wallace knows that some sort of justification is necessary, because the quasi-classical pattern does not appear to be salient by itself: "[a]fter all, 'structure' and 'pattern' are very broad terms: almost any arrangement of atoms might be regarded as some sort of pattern."[127] To justify his pick of the classical pattern, as we have seen, Wallace uses Dennett's criterion. This does two things: *(1)* it makes the recognizing of the classical depend on the (pragmatic) virtue of usefulness, and *(2)* it claims that (because of this usefulness), this classical pattern is then real. With regard to *(1)*, Wallace is rather silent on what this pragmatic import precisely is, so we will have to fill in the blanks ourselves. He does claim that the classical structure is objectively 'there'. With regard to his favourite macroscopic object, the tiger:

> [...] it is an objective fact that tiger-talk picks out high-level structural properties of the microphysical system under study. [...] 'Explanatory power' is being used as a criterion for the objectivity of structures, but the structures thus identified are objectively real.[128]

Coming from some form of quantum state realism, it is undeniable that the quasi-classical structure is objectively real; it is simply an existing pattern in configuration space in the description of the quantum ontology. But this goes for any pattern in that configuration space. And although it is also "an objective fact" that tiger-talk highlights the pattern of a tiger within an entire web of possible patterns, this is not to say that this tiger-talk signals some feature of the world that is 'more real' than others, as if the other patterns are not 'really' there. Wallace is quite right in asserting that these structures are objectively there, but this is not to say that these patterns are objectively salient, let alone that they are more real than other patterns.

In the quantum formalism all bases are on the same footing in the Hilbert space (so-called Hilbert space democracy), just as all inertial frames in special relativity theory are treated equally. This is encoded in the fundamental laws of these theories. For the latter case, the relativity postulate ensures that all laws are written down in a covariant form. In Hilbert space, the superposition principle ensures that for any arbitrary choice

[127] Wallace, *Emergent Multiverse*, p. 50.
[128] *Ibid.*, p. 57.

of basis the evolution from an initial state to a final state will be the same. The unitary evolution of the Schrödinger equation applies to each part of the wavefunction projected onto the basis vectors individually. This is precisely the property that Everett exploited to define his relative states.

So what do we mean by 'preference' when we speak of the 'preferred basis'? From the above, we see that there is the sense of preference that nature has for evolving in a particular way rather than another. This is directly predicated upon brute facts about the nature of interactions and hence about the specific forms that interaction Hamiltonians take. This is the dynamical preference one often talks about when treating decoherence. The Schrödinger evolution 'prefers' some basis in the sense that it is robust under decoherence, while another choice of basis will generally show wild interference terms. Nevertheless, surely it is not the dynamics that 'prefers' anything in the sense that humans prefer things. A 'dynamically preferred basis' is just another way of saying that there is some basis in which some unique dynamical behaviour occurs which does not occur in the other bases. That is not to say that this unique behaviour is subjective or illusory; the dynamical behaviour itself is objective. What I mean is that there is nothing in the theory that elevates the decoherence basis as something more fundamental than the others, which is in line with what Everett intended with his relative states.

Again, the analogy with the relativity postulate in special relativity applies. In that theory, as long as no gravity is present (the vanishing of the Riemann tensor), the laws of physics are the same in all inertial frames. It is true that, for example, all inertial observers would agree on the intensity of the cosmic microwave background. This introduces a preferred frame when it comes to physical facts, the so-called cosmic frame. Yet, this does not conflict with the relativity postulate at all, which entails that different choices of inertial frames will lead to agreement when calculating Lorentz-invariant quantities.[129] The decoherence basis in quantum mechanics is similar to the cosmic frame in cosmology. There is something objective *as a physical fact* about the decoherence basis, namely that it is the basis in which interference terms vanish. This, however, does not entail that patterns in this particular basis are more objective than patterns in another basis, or more real (as Dennett's criterion implies). The choice to 'prefer' the decoherence basis, is made at the subjective level.

This might seem to be justified by the objective process of decoherence—as Wallace claims. But, the quantum formalism – despite decoherence – does not provide us with an underlying principle for this justification. In that case, one option would be to reason downwards from the classical, higher-level theory, to the quantum level: there are real classical objects around us and that we must find an explanation of there existence. In other words, we use our experience to argue that some patterns in the quantum formalism are 'bogus' (to use Dennett's technical term). Note that this con-

[129] Indeed, the mistake that the cosmic microwave background breaks the relativity of reference frames has been made before, e.g., by Hermann Bondi and Peter G. Bergmann, *cf.* Fred A. Muller, "On the Principle of Relativity," *Foundations of Physics Letters* (1992), Vol. **5**, No. 6, Plenum Publishing Corporation.[60]

flicts with the starting point of the Everrettian, namely that nothing needs to be added to the quantum formalism to explain our experiences of a classical world. There is a circularity here. The emergence of the multiverse should follow from the perspective of the universal wavefunction, but when the usefulness of explanatory power comes in, it seems to follow from the classical world perspective. In other words, patterns are useful for describing classical worlds on the condition that we need explanation in a classical world.

I would therefore claim that the only way out is to regard the classical patterns that are picked out because of their usefulness are *subjectively* salient, because it is up to the observer to say what is useful. Although I think this conclusion is right, it is surely too blunt. From experience, we seem to objectively be part of a classical structure. Thus, the usefulness of this structure to us seems to be objective because of that. The problem is that, due to Hilbert space democracy, quantum mechanics does not tell us why this is the case! In a basis different from the decoherence basis, structures emerge that are just as real to the wavefunction realist, but are (in the basic sense) built out of the parts of the (original) pattern in the decoherence basis. The same information that gives rise to the pattern in one basis is responsible for a different pattern in a different basis. Wallace claims that it is only in the classical basis that classical patterns arise and that it should therefore be preferred, but I think that a different representation of the wavefunction should not change reality. Therefore, the question *why* the world around us is classical – granted that it is – is begging.

Therefore, I am confronted with two possibilities. It could be the case that Wallace does not intend to solve the preferred basis problem with Dennett's criterion. In that case, the preferred basis problem (and therefore the measurement problem) remains unsolved, since decoherence by itself is not enough.

Alternatively, it could be the case that the solution of the preferred basis is merely stipulated by appealing to things we already know from classical mechanics. Indeed, if the latter were the case, there is good heuristic value in choosing the classical basis: it is the basis in which stable macro-objects arise. However, the stipulation of the preferred basis is then also unexplained, as there is nothing in the quantum theory that suggests this is the basis in which real things emerge, while other bases give rise to 'bogus', which is objectively there, but which is just not a 'real' pattern.

It is certainly not possible in the Dennettian sense to pick the classical basis by claiming that it is this basis that deserves "to be taken seriously, learned about, used." Surely we should take the classical basis seriously, learn about it, use it when it suits us, but we should take any basis seriously when we take the quantum formalism seriously. It are the structures that emerge within the universal wavefunction when written down in the classical basis that Wallace want to take as either 'real' or 'bogus' in the Dennettian sense, not that basis itself.

I think that Dennett's criterion should be used to capture the second kind of 'preferred'; the kind of preference that has to do with us humans. I do not want to content the correctness of the choice: it does indeed make sense to pick the classical pattern if we want to predict and explain things at the macroscopic level. It seems to be that it is enough to understand *why we could be under the impression that the world around us is classical*: we are in a superposition along with everything else; we identify ourselves within only one branch; and interference between these branches is suppressed by decoherence. To deny the problem of definiteness, then, is it necessary to explain anything more than that?

Yes. It is necessary to explain that we are confined in such a branch in the first place and not smeared out over a multitude of such branches. Staying true to Dennett's criterion, Wallace goes on to nuance his position by saying that "it is not the properties of the tiger-structure *in isolation*" which justify it as 'real'. Rather, it is the usefulness of the tiger-level description (i.e. of zoology), a description which includes not only a single tiger, but all the other animals in its vicinity as well as salient features of their shared environment.[130] Picking out a particular salient structure, namely that structure that we would call classical, is useful only in the context of us human beings living at the macroscopic level, being contained in one of the branches of the superposition.

It is the fact that we use terms like 'world' in the first place which picks out a specific pattern to identify with a world. That there is such a pattern is an objective fact of the formalism (and thus, according to the wavefunction realist, of reality). That this part of the formalism is identified with a classical world relies on our understanding or preconception of what a classical world is supposed to look like, i.e. it relies on background assumptions. The classical structure becomes the salient structure of explanation, specified by the desire to have a classical explanation. This is 'the classical stance'.

6.3 An anthropic justification for the preferred basis

No matter how we choose to cut the configuration space into subsystems, it it the cutting itself that gives rise to ignoring information that there nonetheless is. This is because there will always be quantum correlations that are ignored during the cutting. But a cut needs to be made. Otherwise, the universal wavefunction would have very little structure at all: a mere ray in Hilbert space. It is the context provided by our pragmatic decision that determines what our system of interest is. The classical basis is objectively there, but the classical stance needs to be taken before that particular pattern is recognized as useful and explanatory. This is the *classical pattern*, which is picked out in a way directly analogous to van Fraassen's contextual picking out the salient cause in the web of causal explanations.

But we do not seem to have much of a choice, we are humans after all. Here, I would like to give an anthropic reason why, in the end, the classical stance can be justified objectively. It is true that the choice of basis precedes recognizing quasi-classical patterns. Indeed, there is nothing in the quantum formalism that tells us this choice is salient. The subjectivity is due to the human choice of deciding which patterns

[130] Wallace, *Emergent Multiverse*, p. 53.

are ‘real’ and which ones are ‘bogus’. However, this might not be so surprising, since the classical pattern seems to be the only pattern that is relevant to us.

What is not meant here is that the *particular* branch that we as individuals happen to inhabit is in need of explanation. It is the choice of the particular *set* of basis vectors, the classical basis or decoherence basis, that is in need of explanation, not any particular basisvector.

It is only from the classical stance that we do not lack the function of making relevant use of a language and view of nature that we have developed. And it can safely be supposed that we have developed the language in this way, because we see a definite, classical world around us. It seems a bit much to regard this development as a purely subjective choice. Ultimately, this relies on objective facts about how nature is, which, in turn, must rely on how the interaction forces of nature look like, not on the choice of basis. In this sense, the classical stance can (at least partly) be justified through objective principle.

Anthropic reasoning is a way to explain some aspects of nature in the light of human existence. It can be used as an objective selection criterion for what *we* can encounter and what not. That it is not a coincidence that the Earth is so fine-tuned to make life possible (a little more towards or away from the sun, etc.), since the universe is so large and there are so many planets that there had to be one that is capable of harbouring life, is often hailed as an anthropic argument. Another example is the search for an explanation of the asymmetry in time in the early conditions of the universe: there must have been a low entropy initial state for us humans to be here in our (relatively) low entropy state and talk about it.[131]

Although there is nothing in the formalism that tells us why, we know that because we exist, in the manner that we exist, nature must be in a certain way. Only in the classical basis can there be a world like ours, where stable, information-processing, memorizing people exist. According to the quantum formalism, if taken seriously and at face value (as Wallace wants), other patterns in other bases are just as real as those in the classical basis. Nonetheless, because the classical patterns are the only patterns that are relevant to us, the classical basis is salient. Why are the classical patterns salient? Because of us. The quasi-classical patterns are the only ones relevant to us and which make our existence possible. It is the way that we are made up that forces us to take the classical stance. Although Wallace never mentions the word ‘anthropic’ in the context of the preferred basis, he might be thinking much along the same lines:[132]

> In fact there will be a subset of history spaces which are much more convenient: we are information-processing systems, and it can be shown that any such system picks out a consistent history space (Saunders 1993). (Reverting to the subsystem description given earlier, the point is (in part)

[131] *Cf.* Paul Davies’ famous “Inflation in the universe and time asymmetry,” *Nature* (1984), Vol. **312**, pp. 524–527.[23]

[132] David Wallace, “Worlds in the Everett interpretation,” *Studies in History and Philosophy of Modern Physics* (2002), Vol. **33**, p. 649.[76]

> that such a system needs to store memories and if it chooses an encoding of memories into states which are not diagonal in the decoherence basis, they will not last long (Halliwell 1993; Zurek 1991).)

Hence, it is because we store memories and because we last for long periods of time that we start out with a preconception of the convenience of classical patterns before one can unravel them.

In this section, I have tried to explicate this preconception a bit further. There is something that precedes recognizing these classical patterns that arise through decoherence, namely the choice of the classical basis. This is a subjective matter, since there is nothing in quantum mechanics that justifies it. Nevertheless, these patterns themselves are objectively there. They are further robust under decoherence, which explains why there can be stable patterns that store memory over time and do not fall apart into the wilderness of quantum interference. Through this specific preconception, the determinate records problem can be solved: one must first take the 'classical stance' before one can see the classical patterns. Hence, when you ask Laplace's demon that lives at the quantum-mechanical fine-grained level, the quasi-classical patterns are just as real as any other patterns. But when you ask a human, living at the decohered coarse-grained level, the quasi-classical patterns are 'more real': they are actually classical patterns for all intends and purposes. This may be subjective, but it is the best we can do.

7 Unity

"Of course, it is widely believed that to detach psychology from the scientific hierarchy is to consign it to the anti-realist wilderness. What I want to stress in closing is that this is simply not so. To read a functional perspectivalism about the mental as an irrealist view is to commit a category mistake, to judge psychological discourse by a standard which ought never to have been applied to it."

— Huw Price. [133]

Similar to the consequences of Dennett's intentional stance that Huw Price indicates for the idea of psychology in the hierarchy of theories, I will argue that what I have called Wallace's 'classical stance' has consequences for classical theories.

In the previous chapter, the role of subjective factors in highlighting classical worlds from the quantum-mechanical patterns has been discussed. There is something objectively 'special' about 'the' decoherence basis, since this basis does indeed give rise to stable patterns that *when identified at the level of three-dimensional blocks in configuration space* approximately resemble classical patterns. However, at the level of the wavefunction, these patterns are not more real than others in the ontological sense of the wavefunction realist. This is because these quasi-classical patterns have the same ontological status as any other pattern, since all such patterns are contained within the universal wavefunction, which is the *only* fundamental ontology of the theory. The highlighting of the quasi-classical pattern, I have argued, is achieved by Dennett's criterion only in the light of subjective goals or an antropic argument that claims that the classical pattern is salient due to the fact that these are the only ones of interest to human beings and, in fact, is the only basis in which the human observer forms a stable pattern.

Here, we turn the part of Dennett's criterion that relates lower-level entities described by lower-level theories to higher-level entities that fit with higher-level theories. Dennett's criterion states that a macro-object is a pattern, which exists when that pattern is admitted by a theory that admits that pattern in its ontology. Hence, we are dealing with higher-level theories, since the pattern of the macro-object is supposed to represent some derivative ontology, some agglomerate of fundamental constituents, of the fundamental ontology of the lower-level theory. That derivative ontology then corresponds to the ontology (fundamental or derivative) of some higher-order theory.

I will draw an analogy between the epistemological and ontological components of Nagelian reduction and corresponding epistemological and ontological ideas in Wallacian emergence. In the former, theories at different levels relate to each other when the higher-level theory is quantitatively deduced from the theory at the lower level, after which bridge principles need to be specified to encode any conceptual changes for objects that look mathematically similar in both theories, but have very different philosophical consequences. For Wallacian emergence, I similarly identify a quantitative

[133] Price, "Psychology in Perspective," p. 19.

part and a conceptual part. The quantitative role is fulfilled by the pattern-matching of instantiation, while the conceptual bridge principles are given by the conceptual part of Dennett's criterion.

7.1 Intertheoretic relations

7.1.1 Nagelian reduction

In accord with the ideal of the logical empiricists, Ernest Nagel defines reduction as the logical deduction of the reduced theory T from the reducing theory T^* via the deductive-nomological model (DN-model, or Hempel-Oppenheim model of covering laws).[134] According to this model, scientific explanation has the structure of deductive reasoning from the general law (the *explanans*) to an observed particular event (the *explanandum*).[135] In the case that T contains terms that are not familiar to T^*, one can define certain 'bridge principles'. These are conceptual laws that connect the different conceptual meanings of terms in both theories. Nagel's prime example is the reduction of thermodynamics (T) to statistical mechanics (T^*), where concepts such as pressure, temperature and entropy of a container of water can be understood in terms of the positions and momenta of H_2O molecules.[136] A 'bridge' would be the identification of the term 'temperature' with 'mean kinetic energy'—as temperature is a thermodynamic concept absent at the level of the atom. This neatly disentangles the qualitative conceptual framework of the laws of a theory from a quantitative and predictive 'fit' of the theory with the data. It is important to keep this distinction between epistemological and ontological parts of theories in mind in order to avoid any category mistakes.

Problems with the DN-reduction of theories are abundant. One of them is the symmetry between explanandum and explanans, discussed in the previous Chapter 6, where a good explanation of my hat getting wet is the rain outside, but my wet hat is not a good explanation of the rain outside. We would demand from a model of explanation that it includes the former and excludes the latter explanation.

A more serious problem for using the DN-model as a model for reducing one theory to a more fundamental one is that it has never been done in practice. Even the usual examples such as the reduction of thermodynamics to statistical physics (attempted by Nagel himself) or the reduction of optics to electromagnetism (attempted by Sommerfeld[137]) run into trouble due to the stringent requirements that are needed to make valid deductions go through, which is on a par with the simplicity implied by the scheme. Such problems arise both at the quantitative and the qualitative level. At the

[134] Ernest Nagel, *The Structure of Science* (Harcourt: Brace & World, 1961).[62]

[135] William Bechtel, "Axiomatic account of theories," *Philosophy of Science* (University of Chicago Press, 1993), pp 27–29, Werner Callebaut (ed.).[11]

[136] Nagel, *Structure*, pp. 338–345.

[137] Arnold Sommerfeld, *Lectures on Theoretical Physics: Optics* (New York: Academic Press, 1950).[72]

quantitative level because it is practically unworkable to deduce a higher level calculated from the constituents of the lower-level theory. The practice of such a derivation often include artificial approximations or boundary conditions that are not in terms of the theory that is to be reduced.

At the qualitative level problems are due to the difficulty of stating higher-level phenomena exactly in terms of lower-level concepts. One often needs additional boundary conditions in terms of concepts corresponding to the higher level to fully specify the explanatory scheme. These problems are illustrated in the next section.

7.1.2 Interlevel explanation

Nagelian reduction becomes problematic when it comes to the reduction of complex phenomena such as those encountered in condensed matter physics and the special sciences. Fundamental physics is concerned with explaining phenomena at one particular (low) scale, say the nanoscale of atoms, and it provides particular theories that are appropriate at this scale (in this case quantum theory). Such a theory is defined in terms specified to that specific scale. The theory at the atomic level can then explain higher-level theories, such as the molecular level of chemistry. The ideal of the logical empiricists would be to work our way upwards in this manner and explain all the phenomena in the world. However, in many fields, scales are often associated with *levels of organization*. For example, in biology a sort of hierarchical structure is involved, that starts from the molecular level, via cells, tissues, and organisms, often all the way up to the biosphere.[138] Biological explanations therefore often involve entities at multiple levels of organization.[139] The crucial problem for Nagelian reduction is that the explanatory power of the terms associated to different levels of organization seem to be connected to each other in such a way that *the explanation of a phenomenon would be lost if we would try to disentangle the levels*.

In studying the human heart, for example, we would like to explain how the heart beats. The explanation might go through in he following way: 'An electrical impulse is generated in the sinusatrial node (SAN) and then travels via conducting cells to the atrioventricular node (AVN) where it is sent into the muscle cells of the ventricle (via the purkinje fibers, which are specialized in conducting the signal) causing the lower chambers to contract by the release of neurotransmitters' (See Figure 7.1.2). Here we already see an interlevel character, because the explanation uses terms like 'cell', 'chamber', 'electrical signal', which all reside on different levels. What would a reduction from one level of organization to another look like in this situation? Say we are trying to

[138] The topic and case studies of this section are to a large extent derivative of a former essay I have written some years ago, in which I addressed similar issues of multiple realizability and interlevel reduction of theories, but with a very different application. It turns out to also fit well with an application to Wallace's many worlds.

[139] Ingo Brigandt and Alan Love, "Reductionism in Biology," *The Stanford Encyclopedia of Philosophy* (Fall 2015 Edition), Edward N. Zalta (ed.), `http://plato.stanford.edu/archives/fall2015/entries/reduction-biology/`.[16]

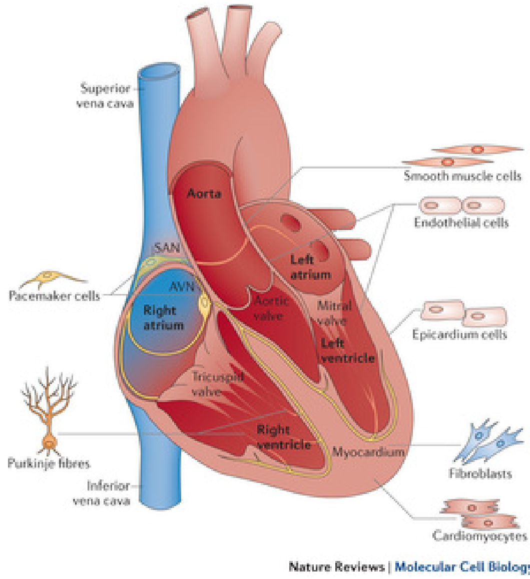


Figure 5: A schematic diagram of the heart. The indications of lower systems such as cells and fibers bare explanatory power when it comes to explaining the workings of the heart, i.e. an interlevel explanation. Picture from: Mei Xin, Eric N. Olson, Rhonda Bassel-Duby [86]

explain the beating of the heart at the level of organs L. The explanatory reductionist's strategy would be to explain this in terms of the parts, i.e. to find constituents C_{ij} at the lower level L_i that constitute level L. The index i labels particular levels of organization and goes from higher levels 1, 2, 3 down to lower levels $n-2$, $n-1$, and finally n, while the index j labels a set of entities at that particular level i. An explanatory reductionist's aim would now be to find an explanation of L in terms of the C_{1j}, which can be understood further in terms of its parts C_{2j}, and so forth. The final goal is to understand everything in terms of C_{nj}—if this works we have found an explanation at a single fundamental level. Furthermore, a theory T has to be formulated to describe L. Then we can construct theories T_i that saves all the phenomena C_{ij} and interlink them to obtain one fundamental theory T_n; this would be Nagelian reduction of theories.

An important basic assumption in the above is that we can provide explanations at a single level L_i *independent from phenomena at other levels* $L_{\neg i}$. In explaining the beating of the heart, the explanation of the contraction of the muscles that make the heart beat is interlevel; it borrows terms from the level of electricity (electrons), cells, fibers, neurotransmitters and entire muscles and organs. The explanation would not go through if one would formulate it in terms of, for example, the cellular level only.[140]

[140] One might, of course, also question why we should be satisfied with an explanation in terms of molecules. Why not go further to nuclei and electrons, or strings, or further? This leads us to another question: is there an ultimate lowest level which is most explanatory? Besides this being a question out of the reach of empirical science, it is problematic for all kinds of explanation. Machamer, Darden and Craver claim that the level where the explanation ends is "relative: Different types of entities

Hence, Nagelian theory reduction fails. If phenomena at level L_i cannot be explained by a theory T_i at that level, i.e. if it needs the help of auxiliary theories from other levels, say T_{i+1} and T_{i+2}, a step by step derivation of T_i by T_{i+1} will necessarily run into vicious circularity.

Weak emergence is often used to fill in the gaps that this absence of direct derivation leaves behind. One can use emergence to argue that one needs not be a reductionist à la Nagel to still hold on to a worldview that is ultimately physicalist. The point is to regard the fundamental constituents as purely physical things described by the most fundamental theory of physics, but that the entities that emerge from this underlying physics can be *explained* by a theory that is not so fundamental.

7.1.3 Singular limits

Another important issue with reduction, raised by Michael Berry[141] and well discussed by Robert Batterman,[142] is the occurrence of singular limits. In many cases of the usual studies of theory reduction only regular limits appear, such the relativistic mass $m_0/\sqrt{1-v^2/c^2}$ in the limit reducing to the Newtonian mass m_0 (at least formally, at the quantitative level) in the smooth limit $c \to \infty$. However, most intertheory relations seem to encounter singular limits, such as the failure of the thermodynamic limit $N \to \infty$ in the attempt to reduce statistical physics to thermodynamics. Berry discusses the reduction of (classical) wave traveling in the x-direction with speed v and wavelength λ, described by

$$\psi(x,t) = \cos\left[\frac{2\pi}{\lambda}(x - vt)\right]. \tag{7.1}$$

For very small wavelengths, we would expect lightwaves to reduce to lightrays, or waterwaves to reduce to smooth currents. But the limit $\lambda \to 0$ is singular, since it is non-analytic at $\lambda = 0$ and oscillates infinitely often between its maxima and minima there. Hence, the wave-description does not smoothly go over into a simpler theory, but breaks down altogether. It is precisely in these disjunctions, according to Berry and Batterman, that new interesting physics is found, such as turbulence in the limit of vanishing viscosity or phase transitions in the limit of infinitely many particles. In those cases where there is no regular limit, Batterman argues, one can speak of novel properties as emergent.

---

and activities are where a given field stops when constructing mechanisms."[56, p. 13].More neutrally out, I think the word 'fundamental' signals only a relational property, 'A is more fundamental than B', which makes ultimate explanation a property relative to the theory; using 'fundamental' as an absolute predicate would be meaningless.

[141] Michael V. Berry, "Asymptotics, singularities and the reduction of theories," *Logic, Methodology and Philosophy of Science IX*, D. Prawitz, B. Skyrms and D. Westerståhl (eds.), pp. 597–607.[14]

[142] Robert W. Batterman, "Chapter 6.2: Singular Limits," *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence* (Oxford: Oxford University Press, 2002), pp. 80–81.[10]

7.2 Wallace's quasi-classical to classical: quantitative versus qualitative

In the above it is discussed that in relating a lower-level theory T^* (like quantum mechanics) to a higher-level theory T (like classical mechanics) there are two parts that should match between T and T^*. They should match quantitatively, one following from the other in an approximated sense or a limiting case; and they should also match qualitatively, in the sense that concepts of T can be explicated in terms of concepts appearing in T^*. As noted, there are many problems with these demands by the logical positivists. Don Howard notices that an overarching explanatory ideal for all the sciences has been abandoned: "[p]hilosophers of physics have overcome the logical empiricist prejudice according to which there is one and only one right method for all scientific domains."[143] Although it is still possible to approximately tie theories together at the epistemic level, explanation does not boil upward from the most fundamental theory to the explanandum.

Nevertheless, one cannot abstain from speaking out about the relations between theories at different levels. This is clearly also the case when confronting the measurement problem of quantum mechanics, since it arises from an apparent conflict between the formalism of the lower-level theory and our experience in terms of the concepts of the higher-level theory. Accepting that the specific form of derivation and explanation depends on the context of the field or science that is under examination, lets look at Wallace's approach to explanation and derivation of higher-level entities like 'worlds'. In his work, the quantitative and qualitative parts often run together. However, it is my intent to disentangle these two in the following sections. I will first address the quantitative part, which works through the pattern-matching of the patterns of the theory in configuration space, where an instantiation-relation connects classical patterns to quasi-classical patterns. After that, I will identify a conceptual identification between classical vocabulary and the vocabulary of the wavefunction, which derives from the reliance of Dennett's criterion on the conceptual framework of the classical theory.

7.2.1 (Approximate) quantitative reduction: Instantiation

The quantitative part of reducing the superposition of quasi-classical worlds to a multiplicity of real classical worlds is given by instantiation relations (see Eq. (3.5) and its definition in the quote above). Wallace does not intend to give a strict derivation of high-level concepts from lower-level constituents, which is in line with the contemporary approach to reduction. That is, today the strict demands of Nagelian reduction are loosened and are often accompanied by emergence claims; we see a similar mix in Wallace's *Emergent Multiverse*. His attempt is to relate the quantitative parts of

[143]Don Howard, "Chapter 6: Reduction and Emergence in the Physical Sciences: Some Lessons from the Particle Physics and Condensed Matter Debate," *Evolution and Emergence: Systems, Organisms, Persons* (Oxford: Oxford University Press, 2007), pp. 141–157, Nancey Murphy and William Stoeger (eds.).[44]

classical mechanics and Everettian quantum mechanics through the elastic concept of instantiation. It is clear that when he speaks of instantiation he has in mind the essence of relating theories to each other :

> Crucially: this 'reduction', on the instantiation model, is a local affair: it is not that one theory is a limiting case of another *per se*, but that, *in a particular situation,* the 'reducing' theory instantiates the 'reduced' one.[144]

He gives a few examples of intertheoretic instantiation. Computability theory is claimed to be instantiated by the solid-state physics of computer chips; zoology is an instantiation of game theory; the Solar System is instantiated by molecular quantum physics; and human emotions by neurological entities.[145] What is more, it seems that all of these high-level entities should be explained *at their own level of organization*. As discussed in Chapter 5, a prediction in terms of lower-level concepts will be just that: a predcition of how the lower-level constituents will behave. Even if we could calculate the exact trajectories of the particles that constitute the tiger, a micro-physical explanation would only give us the solutions to the equations of motion of those particles, and nothing else. In Wallace's words, it would only give us a 'bare' prediction. The bare prediction is clearly meant as the prediction that uses terms only from the more fundamental theory. Hence, the claim that these lower-level constituents are not up to the task of explaining the higher-level phenomenon is against the reductionist aim of finding explanation at the more fundamental level.

But was is a "bare prediction"? Here Dennett's example about the computer running a chess program comes to mind: is the computer playing chess and thinking about its moves or is it simply firing off different electronic circuits according to some algorithm. The latter case seems to be a 'bare prediction', but it would not be particularly useful to use the theory of electrons and circuits to predict which moves the chess computer will represent on the screen. Following Dennett's trade-off between accuracy and effort during prediction-making, it would take much less effort to 'act as if' the computer is actually playing chess and 'thinking about' its moves. The functionalist's claim is that if such 'acting as if' turns out to be indistinguishable from the 'real' thing (that is, humans thinking about chess moves), then both the human and the computer are thinking in the same sense. If the quasi-classical trajectories in the $3N$-dimensional configuration space are 'acting as if' they are the trajectories of a classical N-particle system in three-dimensional space, both patterns *are* the same thing.

Furthermore, the same quasi-classical world can be instantiated by slightly different patterns. Wallace claims that this is comparable to the sense in which the tiger can be instantiated by slightly different patterns of molecules of which it is comprised. The tiger constantly exchanges particles with its environment, and is therefore strictly a different pattern of molecules at every instant in time. However, this never stops one from considering different degrees of 'tigerness'; there is some bandwidth within which

[144] Wallace, *Emergent Multiverse*, p. 55. *Original emphasis.*
[145] *Ibid.*, p. 54.

the different patterns are sufficiently close to each other in order to instantiate the same natural kind 'tiger'. Analogously, there is some tolerance for noise as to in which region in the configuration space the quantum trajectories should be confined in order to be identified with quasi-classical trajectories. This tolerance is the reason why there is so much emphasis in Wallace's presentation on the word 'approximate'.

Lets quickly recap where this approximate sense of individual branches corresponding to quasi-classical trajectories comes from. The emergence of the quasi-classical structure relies on the process of decoherence, as seen in the decoherence basis. We have seen before that decoherence is a gradual, continuous process that makes it impossible to pinpoint an exact moment in time when branching occurs. I have argued, like Wallace, that this continuous process is in principle not a barrier to the idea of splitting at all. Once we start looking at very short timescales, there is no need for clearly distinct worlds in order to solve the apparent conflict between the superpositions of the quantum formalism and the definite world of experience.

Wallace's definition of instantiation is explicitly about a mapping between the structures within some domain of theory A instantiating the structures of theory B. Hence, $(A_D \rightsquigarrow B)$ is a quantitative mapping of structures. But structures by themselves do not have a conceptual interpretation; they are formalism without interpretation. As indicated in the previous chapter, the pattern of flogiston is not the substance flogiston itself, but rather some real structure that by approximation can be described by appealing qualitatively to the substance flogiston. Hence, there is only a quantitative claim here about the matching of quasi-classical world structures described by quantum mechanics and the multitude of classical worlds described by classical mechanics. Any qualitative matching is not achieved by the instantiation relation. Nevertheless, Wallace claims there is no explanation to be found in the "bare" behaviour of the constituents. Hence, some qualitative relation should also be specified. This, I claim, is one of the things achieved by Dennett's criterion.

7.2.2 Qualitative reduction: Dennett's criterion as a functional 'bridge law'

The approximateness of the quasi-classical trajectories means there is a kind of 'tails-problem' when it comes to establishing the autonomy of the branches. In other words, although the branches are autonomous for all practical purposes in the quantitative sense, *in principle* there are still the tails of the other worlds that have non-zero support in the everywhere in the configuration space. Furthermore, Wallace puts some emphasis on the bare prediction; what a bare prediction seems to be is a quantitative approximation at the structural level in the absence of a conceptual explanation *at the appropriate level*. His distinction between the quantitative (bare) prediction and conceptual explanations is one that I agree with. One can then argue about the appropriate level for the explanations to be found. The level at which the explanation is supposed to be found, then, must provide conceptual connections between the theories. A quantitative instantiation-relation cannot account for that conceptual connection. Wallace's position, then, is that only patterns in matter give rise to meaning. Here, I will argue that, although instantiation perfectly deals with the quantitative part of the ideal of theory-reduction, the qualitative jump from quasi-classical trajectories in a $3N$-dimensional configuration space to a multitude of disjoint classical trajectories of 3-dimensional classical spaces needs something more: a higher-level conceptual apparatus.

The careful reader can distinguish a quantitative part and a conceptual part in Dennett's criterion. On the one hand, the pattern – whatever it is – of the macroscopic object that is admitted in the ontology of the higher-order theory should quantitatively match the pattern of the lower-level theory. In the previous section, I have identified this quantitative pattern-matching with the claim that the classical theory and the quantum theory are connected by the instantiation relation, which I accept. On the other hand, Dennett's criterion does not allow for any pattern matching that the ontologies of both theories admit. That is, when it comes to patterns, not just anything goes when it comes to the ontology of the higher-level theory; these higher-level patterns must have explanatory and predictive power as as well. In other words, the patterns admitted in an *un*useful higher-level theory (whichever one we might cook up) do not correspond to real macro-objects. It is in this usefulness relation that we can find bridge principles that are formed by the conceptual part of Dennett's criterion.

For Wallace himself, this qualitative matching is nothing new, as this is also necessary in the usual cases of specifying intertheory relations, e.g., the reduction of thermodynamics to statistical physics:[146]

> "the temperature of bulk matter is an emergent property, salient because of its explanatory role in the behaviour of that matter. (It is a common error in textbooks to suppose that statistical-mechanical methods are used only because in practice we cannot calculate what each atom is doing separately: even if we could do so, we would be missing important, objective properties of the system in question if we abstained from statistical-mechanical talk.) But it is somewhat unusual because (unlike the case of the tiger) the principles underlying statistical-mechanical claims are (relatively!) straightforwardly derivable from the underlying physics."

Hence, one should not think that the wavefunction realist does not intend to *explain* everything we experience purely in the terms of quantum-mechanical wavefunction behaviour. On the contrary, even if we could derive the evolution of Schrödinger's cat, this does not provide us with an explanation in terms of 'dead' and 'alive', which are predicates we usually reserve for agglomerates of atoms in the form of organisms. This explanation should be given at a higher level, and will have to be formulated in higher-level terms. This goes hand in hand with loss of accuracy: a cat who lost an eye in a fight has different structure, but will still be a cat. The demand for such a higher-level explanation hardly seems able to avoid an import of pragmatic values, precisely because *there is no analytic derivation.* This route leads to the necessity of a criterion

[146] Wallace, "The Everett Interpretation, p. 13.

for highlighting particular higher-level structures over others—Wallace uses Dennett's criterion.

However, as I have argued before, unlike the case of the tiger where we can identify a certain agglomerate of molecules (moving through three-space) that more or less constitute the tiger (moving through three-space) at the macro-object, while in the case of the quantum state (in Hilbert space) that should constitute real macroscopic objects in three-space. The ontologies of the objects do not match. If we would want to explain how the tiger arises from parts of the wavefunction, the often-uttered claim that it is 'build out of' wavefunction does not work very well. What could it mean to say that macro-objects are build out of wavefunctions? After all, is it not the case that the wavefunction cannot be found in the three-dimensional space around us, no matter how far you zoom in towards the so-called fundamental level?

The recovery of a multiplicity of quasi-classical structures in the $3N$-dimensional configuration space that individually resemble the classical structure of an N-particle system in three-dimensional space also needs a conceptual bridge in order to *explain*. The explanation takes place in terms of the higher-level theory: we need classical 'world-talk' to explain what quasi-classical world trajectories in the configuration space (that we perceive as the classical world) are. This is similar to the need for 'tiger-talk' to explain what that stable agglomerate of atoms (that we perceive as a tiger) is.

The conceptual mismatch between the theories is not due to any nit-picking about the tails problem eluded to above, namely the fact that although the branches are autonomous for all practical purposes but, in principle, have non-zero overlap. Even if the support in the configuration space would be strictly zero, and the branches would be completely detached from each other in that space, the mismatch does not go away.

Consider the infinite-time limit, where the autonomy of the branches becomes exact, since the exponential suppression of the tails makes sure they thoroughly vanish after an infinite amount of time.[147] In that limit there is a superposition of exactly detached branches, evolving individually, without interference with other branches. The detachment – going from approximately autonomous worlds to exactly autonomous worlds – is a smooth process; there is no singularity involved such as the singularities found when taking the thermodynamic limit in statistical physics or the limit of vanishing viscosity in fluid mechanics. Hence, following Batterman, there is no interesting new physics to be expected in this limit (like phase transitions or turbulence in the two cases mentioned). The approximate sense in which the quasi-classical and classical patterns coincide is, in this respect, uninteresting.

The conceptual leap is taken by interpreting the superposition of autonomous quasi-worlds as a multitude of classical worlds in classical space. There is no quantitative limit corresponding to the reduction of classical space to the configuration space.

[147]It is of no concern to us that we cannot wait an infinite amount of time in practice, since we are addressing a conceptual issue.

Going from the Hilbert space of the quasi-duoverse of Eq (3.2),

$$c_{\uparrow}\left|\uparrow\right\rangle \otimes \left|\ \right\rangle \otimes \left|\text{rest of universe}\right\rangle + c_{\downarrow}\left|\downarrow\right\rangle \otimes \left|\ \right\rangle \otimes \left|\text{rest of universe}\right\rangle, \tag{7.2}$$

to individual identifications of two universes like[148]

$$c_{\uparrow}\left|\uparrow\right\rangle \otimes \left|\ \right\rangle \otimes \left|\text{rest of universe}\right\rangle \ \ \&\ c_{\downarrow}\left|\downarrow\right\rangle \otimes \left|\ \right\rangle \otimes \left|\text{rest of universe}\right\rangle, \tag{7.3}$$

to the duo of classical spacetimes

$$\textit{`classical universe with dead cat'} \ \& \ \textit{`classical universe with live cat'}, \tag{7.4}$$

is motivated by Dennett's criterion.

The part of Dennett's criterion that establishes this conceptual connection between the higher- and lower-level theories is given by the words 'usefulness of theories', in particular their explanatory power. In Chapter 6, I have argued that the strict reliance on the usefulness of theories is mistaken, because of pessimistic meta-induction: theories in the past have proven to be useful despite them being wrong. However, they have not proven to be wrong within the domain of predictability that they were successful in. Hence, in the quantitative understanding of usefulness, i.e. its predictive power, we can relate the patterns of the higher-level and lower-level theories through instantiation. However, in the qualitative or conceptual understanding of a useful theory, namely the explanatory power it provides, this will not work. Wallace believes the quantum level does not *explain* the classical level of worlds. For him it is enough to establish the instantiation relation between classical and quantum mechanics and then explain the classical level in the conceptual framework of classical physics itself.

There is a further point to be made about the reliance on interlevel explanations from classical mechanics to solve the measurement problem. Niels Bohr considered a classical measurement context as a prerequisite to even talk about quantum-mechanical behaviour, hence the need for a correspondence principle. He claimed that we can only speak of what is going on at the nanoscale in the context of the 'phenomenon', understood as the quantum system entangled to a classical apparatus. Only then do classical concepts such as 'momentum' or 'position' make sense. This is similar to the classical stance, but Wallace does not go quite this far, since he does allow himself to speak of objective behaviour of the universal wavefunction and hence about everything that exists in the multiverse. However, to explain the appearance of a classical world (i.e, solve the measurement problem), he confines himself to only classical terms. According to the quantum theory all Hilbert space bases are on the same footing and patterns will evolve for any choice of bases. Without the classical theories around, there would be no concepts to highlight to classical basis (which happens to coincide with the decoherence basis) and the classical patterns that form within that basis from the universal wave-

[148] Unless the aim is to presuppose *one* classical three-space in which all the quantum patterns in configuration space are instantiated on top of each other.

function. Using classical preconceptions, one needs to stipulate the classical basis by hand in order to see what part of quantum mechanics gives rise to real macro-objects and what part is an excess baggage of patterns. Quantum mechanics itself is silent about this.

To conclude, it should be clear by now that there is something odd about equating predictability with explanatory power under the umbrella of usefulness. A theory can be useful to predict the behaviour of things. This is even more useful when quantitative patterns can be specified. The classical theory has clearly proven useful in this quantitative respect of predictive power. When it comes to explanatory power, though, I think this has nothing to do with that same kind of usefulness. The usefulness that derives from explanations is one that provides an understanding of things by using a particular conceptual framework about those entities that we believe exist in the world. Hence, it boils down to the question of what one believes a successful explanation to be. By equating these two very distinct ideas of usefulness, I think Wallace takes a specific philosophical stance, namely that of the scientific realist, which will be the next topic, Chapter 8.1.

For this chapter, which is about the relations between quantum and classical theories, it is enough to conclude that instantiation provides a good link between the quantitative predictions of the theories, while Wallace's emphasis on explanation provides explanatory links between the classical and quantum theory. The instantiation relation is an effective *mapping* between the predictions of the theories, such that we are ensured that the phenomena are being saved. When it comes to explaining those phenomena, though, Wallace borrows the concepts from the classical theory in a very particular way: by appealing to useful higher-level theories and claiming the concepts in those theories to be real because of that usefulness on the one hand, and because there is that quantitative structural connection with parts of the universal wavefunction on the other. In the previous sections, I have called this the 'classical stance'.

8 Reality

> *"But I insist, yet again, that even this decoherence basis does not admit of being made wholly precise, and is certainly not to be thought of as preferred in any* deep *sense, such as would justify our regarding the formation of superpositions with respect to it as amounting to a genuine fissioning of reality."*
>
> —Michael Lockwood[149]

In the previous sections I have primarily identified and analyzed two elements in Dennett's criterion. First, that the reliance on usefulness has a pragmatic flavour, akin to van Fraassen's context-dependent explanations making a particular pattern in the causal web salient. And second, that the pattern-matching between classical and quantum-mechanical patterns needs a conceptual component next to the approximately quantitative component – analogous to Nagelian reduction through DN-derivation plus conceptual bridge principles – which are respectively given by the conceptual part of Dennett's criterion and instantiation.

In the previous sections, I have often said that usefulness does not automatically entail that the classical pattern obtains a special status above the others or that it is the salient pattern. Some philosophers, however, would be inclined to say this. This step is at the very heart of the realism debate. We turn now to the question of what could be meant by the concept of 'reality' in Dennett's criterion. It says that "a macro-object *is* a pattern, and *the existence* of a pattern as a *real thing* depends on the usefulness [...] of theories which admit that pattern in their ontology."[150] The emphasized words signal the reality of some things over other things. The idea seems to be that quasi-classical worlds are patterns in the configuration space of the quantum theory (and hence strictly part of a superposition of quasi-classical worldly patterns), but that those quasi-classical patterns, which form a subset of the entire quantum structure, find a home in the higher-order ontology given by the theory of classical mechanics. If we follow Dennett's criterion, we can promote these quasi-classical worlds to independent *real* classical worlds, because classical mechanics is useful (it provides deterministically unique predictions) and explanatory (humans tend to think and explain using classical concepts).

At the quantum or Hilbert space level, reality does not depend on the choice of basis. I believe Wallace would also hold this belief, that at the fine-grained level the patterns in any basis are as real as in another basis (and for the same reason). It is at the coarse-grained level, at the level where a decoherence basis can be found in which 'large' structures arise (in the configuration space), where Wallace wants to promote the patterns that are useful to us, to real emergent objects.

[149] Michael Lockwood, "'Many Minds' Interpretations of Quantum Mechanics," *British Journal of Philosophy of Science* (1996), Vol. **47**, p. 186.[55]

[150] Wallace, *Emergent Multiverse*, p. 19. *My emphasis.*

In the following, I point out some of the common issues in the realism debate that arise with Wallace's mixed use of explanatory power and predictability as criteria for reality. Next, I discuss a rival theory of quantum mechanics that does not regard classical worlds as real—quite to the contrary, one could say that in this theory the classical universe is an illusion of the brain.

8.1 Scientific realism through usefulness

In the previous section, I have asserted that by equating two distinct ideas of usefulness, predictability and explanatory power, I think Wallace takes a specific philosophical stance in the scientific realism debate. The reason for this assertion lies in the interplay between the concepts in Wallace's Dennett's criterion, which I repeat one final time:[151]

> A macro-object is a pattern, and the existence of a pattern as a real thing depends on the usefulness – in particular, the explanatory power and predictive reliability – of theories which admit that pattern in their ontology.

At the end of the day, this criterion establishes the *existence of a pattern as a real thing*. In Chapter 6, I have argued that the appeal to the usefulness of theories to establish an objective reality has an anthropic flavour (which is not to say that there are no objective stable patterns due to the process of decoherence). In Chapter 7, I have claimed that specifying usefulness in terms of explanatory power and predictive reliability relies on two quite different interpretations of the word 'useful'. Here, I conjecture further what could be the reason why such language is employed in Wallace's attempt to solve the measurement problem.

Note that in the formulation of the criterion, there is a residue of the realism debate in the philosophy of science. Scientific realists of the twentieth century used to argue for the truth of scientific theories on the basis of explanatory usefulness and predictive success. Lets briefly turn to the two most famous arguments.

First, the idea that the macro-object is a real pattern because the theory that admits it in its ontology has predictive power is similar to Putnam's 'no-miracles' argument. According to this argument, theories are decribing real things, because otherwise their predictive success would be utterly mysterious.[152] Clearly, our best theories are very successful, because they accurately predict and retrodict empirical data and give us tools to manipulate the natural world around us to such a degree that even Francis Bacon would not have imagined it. How is this success to be explained? The intuitive answer, at least for the realist, would be that those theories are (approximately) true, since if they were not true their empirical success would be miraculous. It seems that only a true theory can explain its success.

[151] Wallace, *Emergent Multiverse*, p. 50.

[152] Hilary Putnam, *Mathematics, Matter and Method* (Cambridge: Cambridge University Press, 1975), p. 73.[68]

Similarly, in Wallace's Dennett's criterion a pattern exists as a real thing because the theory that admits it in its ontology is useful in making predictions and retrodictions. Similar antirealist objections as those employed against Putnam's argument can be made here. For example, some have claimed that the argument is mistaken due to committing the base rate fallacy. The simple reason is that there is no way of ascertaining the base rate probability of successful predictions independently of ascertaining the probabilities of the false negative and false positive rates of the theory.[153] Also, a selectionist will claim that the useful theory survives not because it is true and therefore describes real things, but because theory-construction is a trial-and-error process during which only the useful theories survive by adaptation, analogous to an evolutionary survival of the fittest.[154] Hence, the useful higher-level theories that Wallace is speaking of could just be useful because they have been constructed through many trials (and many errors) of attempting to find stable patterns within the universal wavefunction, not because they describe real macro-objects.

Second, the idea that the macro-object is a real pattern because the theory that admits it in its ontology has explanatory power is similar to the 'inference to the best explanation' account of scientific realism. Truth claims made by the theory are justified by their explanatory value.[155] All over the fields of science, the goodness of explanations is used to decide whether we believe in a theory or not. Inference to the best explanation takes a realist leap when one claims that because the theory provides the best explanation it must therefore also be true.

A problem that immediately imposes itself is the question of how to judge which explanation is better than another. This brings us back to the human goal that is involved in explanation, as argued for in Chapter 6. There, I claimed that the classical basis is preferred because we wish to explain in classical terms that describe stable patterns over time. In an attempt to make this more objective, one can employ an anthropic argument by claiming that the classical basis is the only basis that is relevant to us. But this argument has been extensively treated before.

To put it in a nutshell, one should acknowledge that scientific realism is a broader position than that of quantum state realism. Quantum state realism asserts that the quantum state is all there is. All the other things that exist, should ultimately supervene on the behaviour of the unitarily evolving universal wavefunction. Wallace holds that emergent things are no less real for being emergent, but this is not a position that derives from state realism alone. It is a functionalist claim, which is independent of wavefunction realism or quantum state realism.[156]

[153] P. D. Magnus & Craig Callender, "Realist Ennui and the Base Rate Fallacy," Philosophy of Science (1994), Vol. **71**, No. 3, pp. 320–338.[57]

[154] Van Fraassen, *Scientific Image*, p. 40.

[155] Peter Lipton, *Inference to the Best Explanation* (London: Routledge, 2004), 2nd edition (original 1991).[54]

[156] Clearly, the argument here ties in to a general debate between Dennett's specific functionalist stance and the views of other realists (whether milder of more committed than Dennett). In that

It might also be the case that for Wallace only the quantitative part, the pattern-matching, matters. The real pattern explains better *because it is predictive*. I believe this would lead to a disregard of the conceptual side of explanation, but even if this would be Wallace's position, it would be redundant to use the word 'explanatory' in Dennett's criterion. One could then simply speak of the useful predictions of theories and leave explanation out of the equation. In that case, I think it would be wiser to commit fully to a form of structural realism (a view that Wallace accidentally holds, but which he claims not to be relying on in *The Emergent Multiverse*), where only patterns are real, objects are not, and concepts about any of those illusory objects are just useful short-hands for patterns.[157]

Alternatively, Wallace's approach also fits with explanationist realism.[158] The explanationist realist identifies specific parts of the theory that are worthy of being taken seriously on the merit of their explanatory value. Those parts that explain better the empirical success of the theory are better candidates to be 'real parts'. For example, Philip Kitcher claims there are "presuppositional posits" of the theory, which do no work in the explanatory scheme, and "working posits", which do the explanatory work and help to indicate novel predictions of the theory. In this framework, the explanationist should commit to the working posits.[159] This seems to be part of what Wallace is doing. Wallace commits to the patterns in the classical basis that effectively behave classically due to decoherence. The patterns that arise in the other bases can be regarded as completely idle, they are like Kitcher's presuppositional posits. For the wavefunction realist, all patterns in any basis are equally real. But the explanationist realist could argue that the decoherence basis should receive a higher ontological status than the other possible bases in Hilbert space.

My claim is not that Wallace's realist position is mistaken. Indeed, I do not expect to find many wavefunction realists who are also full-blown empiricists (because of the cognitive dissonance). But my point is that it is not a necessity to import ideas from the realism debate like the 'no-miracles' argument or inference to the best explanation in order to solve the measurement problem. Wallace does not claim to do this

sense, the critique at Wallace in this section is more generally aimed at functionalism than at Wallace's employment of it. However, Wallace does use a specific form of functionalist (rather differently than Dennett), and I have included this critique to highlight the point that functionalism need not play a role in the solution of the measurement problem.

[157] Note that even if structural realism is the background view that motivates Dennett's criterion, the talk of the "macro-object" in the formulation of that criterion is not in major conflict structural realism, since it defines the objects as some emergent (or conventional) property of the underlying structure. Nevertheless, different shades of 'real' should then be attributed to the emergent properties which are "no less real for being emergent" and the underlying structures, which are the actual fundamental ontology.

[158] See Anjan Chakravartty, "Scientific Realism," *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), E. N. Zalta (ed.), URL: `https://plato.stanford.edu/archives/sum2017/entries/scientific-realism/`.[21]

[159] Philip Kitcher, *The Advancement of Science: Science Without Legend, Objectivity without Illusions* (Oxford: Oxford University Press, 1993), pp. 140–149.[53]

explicitly, but he does rely on specific relations between explanation and reality without formalizing the ideas behind these relations. It would be better to explicitly take a stance on how the use of Dennett's criterion relates to explanationism or structuralism, or scientific realism in general. From the above analysis, I believe it would be difficult to show why Dennett's criterion would not entail a specific position beyond quantum state realism.

The above point is important, because it leaves room for positions that want to be less metaphysically committed, while retaining a realistic attitude towards the wavefunction. Indeed, one can be a wavefunction realist, follow Wallace's analysis of decoherence, emergence and the recovery of the Born rule and still deny the objective multiplicity of macroscopic classical worlds. Nevertheless, this position does make it possible to undermine less realist positions, like the many-minds interpretation.

8.2 A possible foil: real worlds versus illusory worlds

Everett's idea was that the universal wavefunction bears within itself a multiplicity of branches when decomposed in a relative basis, where each of these branches corresponds to a definite description of a world that is much like our classical world of experience. But the level at which this multiplicity comes about, which is the level where the von Neumann entanglement chain practically ends, can be challenged. We have seen that Wallace seeks for a multiplicity at the physical level of real macroscopic three-dimensional worlds. However, Everett's ideas can also be applied to different levels of description, as long as this is done consistently.

Some have sought for this multiplicity at the level of the observer itself, be it in terms of the perception apparatus, the beliefs that correspond to certain states of the brain, or in the 'mental realm'. This search is at the foundation of several versions of what has become known as the 'many-minds' theory, first conceived by Zeh,[160] but most prominently developed by Albert and Loewer,[161] Matthew Donald,[162] and Michael Lockwood.[163]

8.2.1 Many minds

In the many-minds theory, the multiplicity that results from identifying separate branches occurs at the mental level instead of the physical. In the physical realm, there are indeterminate quantum states, as given directly by quantum mechanics. This should be understood as one physical state of affairs: the state specified by the universal

[160] Zeh, "Measurement."

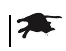
[161] David Albert & Barry Loewer, "Interpreting the Many-Worlds Interpretation," *Synthese* (1988), Vol. **77**, pp. 195–213.[2]

[162] Matthew J. Donald, "Quantum Theory and the Brain." *Proceedings of the Royal Society* A (1990), Vol. **427**, pp. 43–93.[30]

[163] Michael Lockwood, "'Many Minds' Interpretations of Quantum Mechanics," *British Journal of Philosophy of Science* (1996), Vol. **47**, pp. 159–188.[55]

wavefunction—and that state naturally includes observers. This entails that there will also occur improper mixtures of brain states in the basis where different definite brain states can be distinguished from each other. Each of the brain states taken *by itself* corresponds to a different mental state. Hence, there is a unique physical state, but a multiplicity of mental states. One can describe such a situation as if the (physically unique) observer has 'many minds', each 'mind' corresponding to a definite perception. If one adopts the above view, the measurement problem reduces to recovering how we experience a classical world, not how there are actual, ontological, classical worlds. The measurement problem revolves around making sense of the apparent conflict between the indefinite quantum-mechanical description and the definite world of experience. Where the many-worlds theory solves this conflict by accounting for our definite perceptions as a feature of nature itself, the many-minds theory accounts for definiteness as a limitation to the way our perceptions work.[164]

Lets illustrate this by reiterating the analysis that led to the version of the duoverse, Eq. (3.2), as it was discussed in Section 3.2. We start out with the superposition that includes Schrödinger's cat as in Eq. (3.1) and consider its further entanglement with its immediate environment. Say one intends to 'measure' the state of the cat using a measurement apparatus like a thermal scanner $A^{(0)}$ (eigenstates $a^{(0)}$), which is displayed on some screen. The screen itself is observed by the physical state of an observer $A^{(1)}$ (eigenstates $a^{(1)}$). At this level of analysis of the von Neumann chain (Eq. (2.3)) the measurement problem is solved. This is done by interpreting the superposition at this level as a mixture for the brain of the observer, which can be interpreted as a multiplicity of mental states. Let $A^{(1)}$ be the observer herself. The total physical state is given by

$|\Psi\rangle = c_\uparrow |\uparrow\rangle \, |$  $\rangle \, |a^{(0)}_{warm}\rangle \, |a^{(1)}_{alive}\rangle + c_\downarrow |\downarrow\rangle \, |$  $\rangle \, |a^{(0)}_{cold}\rangle \, |a^{(1)}_{dead}\rangle \, .$ (8.1)

This state is physically complete and a good description of the entangled situation, but at the mental level of $A^{(1)}$ this can be seen as two mental states, one of which contains the belief that the cat is definitely dead, and one which contains the belief that it is definitely alive. Hence, there is a superposition of physical brain states, while there is a multiplicity of mental states.[165] At higher levels of analysis, i.e. further links

[164] The question is then, of course, 'how do our perceptions work?'. The above-mentioned authors disagree thoroughly about how to answer this question. These differences are discussed in the next section.

[165] Note that this should not be confused with solutions to the measurement problem that suggest that collapse of the wavefunction occurs at the mental level due to interaction with 'consciousness', as suggested in various ways by, most prominently, London & Bauer, von Neumann, or Wigner. Although the many-minds theories have a similar ring to them as these conscious collapse interpretations, they are completely different. In particular, the many-minds theories are versions of the non-collapse solutions to the measurement problem, while the von Neumann-Wigner interpretation suggests consciousness as a *cause* of collapse. Furthermore, the many-minds theories do not use a concept of 'consciousness' at all and can simply deal with mental states corresponding to brain states. Even Albert & Loewer, who introduce some dualism into their theory, deal with the measurement problem by postulating 'minds',

in the von Neumann chain, the observer (presumably) interacts with its environment and becomes entangled with more and more physical systems in the universe, but the state of which it is now part remains indeterminate. This is of no concern to the measurement problem, since it is the experience of definite outcomes that conflicted with the indefinite superposition, and these definite perceptions have been accounted for.

The entire argument of the many-worlds theorist goes through in an analogous way for the many-minder. The incoherence problem is solved by understanding probabilities at the level of an individual mind. For each outcome, the Born rule gives a probability $|c_i|^2$ for some individual mind to evolve into the branch $|\psi_i\rangle$. The observer's physical state evolves deterministically and continuously, while the mental state experiences quantum jumps described probabilistically. Furthermore, the process of decoherence applies at the level of minds in much the same way. Zeh himself proposed to use decoherence as a mechanism to interpret a multitude of autonomous *experiences*:

> Unavoidable "continuous measurement" of all macroscopic systems by their environments (inducing strong entanglement) was indeed initially discussed precisely in order to support the concept of a universal wave function, in which "branching components" can only be separately experienced.[166]

Technically, 'many minds' differs from 'many worlds' through different solutions to the preferred basis problem. The preferred basis problem is solved locally for each individual observer, not at a global level.[167] But there are multiple positions that can be taken within the many-minds idea and these have remarkable differences. In particular, Albert and Loewer go as far as to suggest that one has to postulate a separate ontology: the 'minds' in their theory do not obey the Schrödinger equation. It is only the set of all mental states that supervenes on the physical state. Individual minds lack such a supervenience. Indeed, it makes no sense for one to think of the mind of an observer to be in an indefinite state.[168] An observer's belief about something is never in a superposition and hence it cannot really be physical. This commits them to a form of dualism.

This is a strong claim. Note that when this is indeed the case, when there is some 'mind-stuff' that, although it supervenes on, is irreducible to physical stuff (understood as the fundamental ontology of the fundamental theory of physics), one can speak of the *strong emergence* of minds from quantum mechanics. As layed out in Section 4.1, this strong emergence boils down to the fact that not even a Laplacian demon would be able to derive these minds.

for which the question of conscious experience is independent of the measurement problem.

[166] Hans-Dieter Zeh, "The Problem of conscious observation in quantum mechanical description," *Foundations of Physics Letters* (1999), Vol. **13**, No. 3, p. 233.[88]

[167] Guido Bacciagaluppi. "Remarks on Space-time and Locality in Everett's Interpretation." In: T. Placek & J. Butterfield (eds.), *Non-Locality and Modality* (2002), Kluwer Academic Publishers, pp. 105–122.[5]

[168] Albert & Loewer, "Interpreting Many Worlds," pp. 110–111.

But one need not go this far to be a many-minder. What the many-minders say is that the multiplicity is at the mental level, and one can try to flesh out what is understood by the 'mental level' in different ways. Clearly, the mental has to be correlated to the physical, as mental beliefs are not independent of physical perception: sense data influence what we think. The physical environment interacts with the observer's brain through physical and chemical links that give rise to our states of awareness. Zeh suggests that these links can make several many-minds theories falsifiable.[169] Donald, in fact, works out in detail what such a (falsifiable) physical model should look like.[170] In his view, the part of the brain that is relevant for mental beliefs can be modelled as a series of 'switches' that can capture what we understand as an observed phenomenon. And we are aware of the firing patterns of neurons in our brain, which allows us to mentally reflect on those firings: "we are only aware of those connections through our awareness of the firing patterns."[171] "[I]t will be assumed that mind exists as awareness of brain, but it has no direct physical effect."[172] I do not intend to present Donald's view as conclusive (in particular, his treatment of memory leaves some explanatory gaps), but at least it has the crucial ingredients to solve the measurement problem in place and forms a detailed, falsifiable, theory without requiring dualism, but with a disregard of real classical worlds at the physical level.

8.2.2 Dennett's criterion as undermining the many-minds theories

There is a sense in which the proponent of the many-minds theory can say these classical definite worlds, as experienced by individual minds, are an *illusion*. Certainly, Eq. (8.1) is seen as describing each 'brain branch' corresponding to some definite classical outcome, which encodes why each mind is *under the impression* that the world is classical. The physical realm, according to the wavefunction realist, is not classical at all. Indeed, Lockwood claims that the decoherence basis (as Żurek employed the term), although it is dynamically special in the sense that this basis is robust under the diluting of coherence over the environment, it is not special in any deep ontological sense. One can always pick another basis and represent the objective situation in such a rotated basis, since Hilbert space respects the "democracy of bases".[173]

Hence, the many-minder can uphold – banging fists on the table – that 'there is *really* a superposition!'. Classical worlds are simply illusions of the brain. This directly challenges David Wallace's claim that it is reality itself that is divided up into real classical realms. Indeed, Wallace wishes to side-step the many-minds interpretation already in the introduction of his book:

[169] Hans-Dieter Zeh, "Quantum Theory and Time Asymmetry," *Foundations of Physics*, Vol. **9**, pp. 803–818.[89]

[170] Matthew J. Donald, "A Mathematical Characterisation of the Physical Structure of the Observer," *Foundations of Physics* (1995), Vol. **25**, pp. 529–571.[31]

[171] Donald, "Quantum Theory and the Brain," p. 53.

[172] *Ibid.*

[173] Lockwood, "'Many Minds'," p. 186.

> I wish to avoid [the connotation of the 'many-minds' theories] that somehow a detailed theory of the mental is relevant to the understanding of quantum mechanics, or that there is no real multiplicity in nature, just the illusion of multiplicity in our minds."[174]

Wallace goes on to say that Everettian quantum mechanics really is a many-minds theory, in the sense that when there are many worlds *out there*, there are also many versions of yourself *out there*—many versions of you in which arms, legs and brain states are treated alike. This is true, but also a red herring; it clearly misses the point that is made by the many-minders, namely that the multiplicity is not really *out there*, but only present inherent to our mental realm, while the physical realm is simply described by the wavefunction without any concept of multiplicity. One could challenge the need for some extra dualist ontology due to the failure of supervenience of individual minds on the physical state (as proposed by Albert and Loewer). Or one can insist on difficulties surrounding the preferred basis problem for each particular observer by giving a detailed description of brain states (a bullet that Donald is trying to bite). But one cannot dismiss the basic premiss of the many-minds theorists by turning the argument on its head. The many-minders simply deny Wallace's claim that there is a multiplicity of worlds in the physical realm. This cannot be countered by saying that when there is a multiplicity of worlds at the physical level there will also be multiplicity at the mental level.

This is not the end of the story, though. Some of the many-minds theories do not pose a threat to Wallace's project. For instance, Wallace need not worry about the dualistic version of Albert and Loewer, since he is already departing from that path by assuming *only* the wavefunction as the fundamental ontology and denying that any extra ontology needs to be assumed in order to solve the measurement problem. But Donald's view poses a serious threat, if it turns out to be an acceptable modelling of the brain (which it, as it stands today, certainly seems to be). Donald makes this explicit, too, as he argues that the reality of worlds is nothing special to the world *out there*:

> Nothing is "real" except the switching structures of individual observers (each considered separately), the initial condition, the underlying quantum field theory, and the objective probabilities defined by the hypothesis. Out of these "elements of reality", each separate observer must construct his experiences and learn to guess at what his future may bring.[175]

Wallace is sceptical about the many-minds theories because it somehow rests on a detailed description of how the brain works.[176] In my opinion, such a position is reasonable

[174] Wallace, *Emergent Multiverse*, p. 3.

[175] Matthew J. Donald, " Progress in a Many-Minds Interpretation of Quantum Theory," ArXiv: http://www.arxiv.org/abs/quant-ph/9904001v1 (1999), p. 19.[33]

[176] Wallace, *Emergent Multiverse*, p. 3.

to take, but it does not discredit the many-minds theories at all. It relies on Wallace's functionalist stance, which he defines as follows:[177]

> As a matter of conceptual necessity, mental properties are supervenient on structural and functional properties of physical systems, and on no other properties. Hence, it doesn't matter what the brain is made of, only how it works.

Here, 'conceptual necessity' is explicated as the indication that[178]

> according to functionalism, mental properties are to be understood as by definition being present in systems with certain functional properties. We can distinguish this from the weaker claim that, in fact, mental properties are present in a system if and only if that system has certain functional properties; much of what I will say goes through in this case also.

This is an ontological claim: structural properties of the physical world constitute mental properties. This leaves reductive physicalism as the only viable position and in this way discredits the many-minds theories that take that view. This functionalist stance is made more precise by the use of Dennett's criterion, which directly undermines the many-minds positions that do not accept mental properties in structural terms.[179]

The many-minds positions are undermined because Dennett's criterion (as applied to decoherence) allows one to speak of worlds which are *actually there* by providing a reason why this is so: the worlds are actually there *because their patterns exist* in the quantum theory and these patterns are admitted in the higher-level ontology of classical mechanics. Hence, these macroscopic worlds are real, given by the theory. By fiat of Dennett's criterion, worlds are not an illusion.

[177] Wallace, "Structure," p. 100.

[178] *Ibid.*

[179] Zeh's 'modified psycho-physical parallelism' is a many-minds theory that is not undermined by Dennett's criterion, though. It seems to be compatible with functionalism about 'multi-mental' states. Particular functional properties about being in a decohered mixture of states with certain properties can be functionally identified as the presence multiple mental states. Hence, one would need to give an argument in addition to functionalism to undermine Zeh's position.

9 Conclusion

"*The hidden variable theories are, to me, more cumbersome and artificial—while the Copenhagen interpretation is hopelessly incomplete because of its a priori reliance on classical physics (excluding in principle any deduction of classical physics from quantum theory, or any adequate investigation of the measuring process), as well as a philosophic monstrosity with a "reality" concept for the macroscopic world and denial of the same for the microcosm.*"

— Hugh Everett III, in a letter to Bryce DeWitt.[180]

My claim is that Wallace's Dennett's criterion is doing more than Wallace says it does. Wallace's intended use of Dennett's criterion is this: to promote quasi-classical patterns (subsets of collective $3N$-dimensional configuration-space trajectories of the universal wavefunction) to real classical worlds (higher-level entities within the universal wavefunction that approximately obey dynamical equations of the same mathematical form as those obeyed by a classical N-particle system). Then, the idea is to say that a 'quasi-classical world is a pattern, and the existence of that pattern as a real classical world depends on the usefulness – in particular, the classical explanations and definite predictions (the reason for which is the robustness of the pattern under decoherence in the quantum theory) – of classical physics which admits the classical pattern in its ontology.' I have analysed three conceptual themes of this criterion, namely the supposed objectivity of the emergence of worlds, the quantitative as well as qualitative relations between quantum mechanics and classical mechanics that it presupposes, and the reality-status that it attributes to these worlds.

I argue that the functionalist ideas in Dennett's criterion themselves are in need of justification before it can be accepted, particularly in the form of a specific stance on realism (or at least on what we think scientific theories to be). I furthermore suspect there can be multiple positions that solve the measurement problem along many of the same lines as Wallace, but without the realist talk in Dennett's criterion and without its presupposed connection between the usefulness of theories and the reality of their patterns. One example I have given is the many-minds position, where the quantum formalism is taken very seriously, as Wallace wants to do, but where the definite, classical, worlds are not part of the reality of the universe, but part of the beliefs held by observers.

Quasi-classical patterns emerge weakly from the quantum formalism. These quasi-classical worlds behave as if they are independent and seem to be unexpected with regard to the fundamental ontology that is the universal wavefunction. This is solely ensured by decoherence. Answering the first part of the main question of this essay, by using the analyses of Chalmers and Bedau on the concept of emergence, these quasi-classical worlds can then be seen as weakly emergent entities in the same way that

[180]Everett to DeWitt, May 31, 1957, "Chapter 16: Correspondence: Everett and DeWitt (1957)," *Everett interpretation*, J. A. Barrett, & P. Byrne, pp. 252–256.

thermodynamical quantities such as temperature and pressure weakly emerge from statistical physics. Starting from statistical physics, Maxwell's demon poked a hole in the second law of thermodynamics, by approximating it too a very large degree while allowing for fluctuations. Analogously, by discovering quantum physics we have poked a hole in the classical law of definiteness, since the decrease of coherence is equally defined at the coarse-grained level as the increase of entropy is.

Decoherence is an objective process; it is a result of the way that the interaction Hamiltonians are written, and these Hamiltonians are supposed to correspond to the real forces in the world. The structures that resemble classical worlds at the coarse-grained level of the decoherence basis are thus also objectively there—as long as your take the quantum state literally (as Wallace intends).

As a result, the emergence of quasi-classical patterns itself does not rely at all on the pragmatic factors (nor anything else) raised in Dennett's criterion: quasi-classical worlds objectively emerge weakly as a result of decoherence. The emergence of quasi-classical worlds in the third chapter of Wallace's book is independent of the pragmatic virtues appealed in Wallace's second chapter.

Nonetheless, Wallace appears not to be satisfied with this weak emergence of quasi-classical patterns that are as real as all the other patterns within the wavefunction. He wants to establish that because they are allowed in the ontology of useful higher-level theories they are somehow 'more real'. According to Wallace, the goal of his second chapter is to answer the question "Why does 'quantum mechanics, taken literally' entail that we live in a multiverse?" His answer is[181]

> If we apply to quantum mechanics the same principles we apply right across science, we find that a multiplicity of quasi-classical worlds are emergent from the underlying quantum physics. These worlds are structures instantiated within the quantum state, but they are no less real for that.

I agree with the conclusion that emergent things, as long as they are built out of other real things, are no less real for being emergent. However, the quasi-classical worlds are also not *more* real than any of the other patterns that can be found in other bases—at least not according to the quantum formalism. Wallace is quite right in asserting that classically analogous structures have a certain objectivity in the sense that they 'are there', but that goes for any pattern, also the indefinite patterns in bases wildly different from the classical basis.

As a side-note, I also feel like the remark "the same principles we apply right across science" is too easily made. Although I am willing to follow Wallace's naturalism (in his introduction Wallace tells us this is his only philosophical assumption), I am sceptic that the functionalist claims that are encoded in Dennett's criterion follow from naturalism. Usually, physicists do not apply (something like) Dennett's criterion to establish the reality of macroscopic objects like tigers or the existence of properties like temperature.

[181] Wallace, *Emergent Multiverse*, p. 63.

To return to the quasi-classical patterns being 'more' or 'less' real, I have argued that the quasi-classical patterns only become more 'real' when seen from a certain subjective perspective. This is subjective because one needs to stipulate a preferred basis on regardless of dynamical considerations. Similar to Dennett's 'intentional stance', I claim a specific 'classical stance', a thoroughly *human* view is needed before quasi-classical worlds can be part of 'reality' in Dennett's sense. The classical stance is equivalent to regarding the coarse-grained classical basis (which approximately coincides with the decoherence basis) not only as a useful basis, but as the fundamentally *true* basis from which relevant patterns can emerge. From this stance, one can look at the quasi-classical structures as the true structures and identify them with real classical worlds. I think Wallace would also say that there is no objective basis in which branching occurs, there is only a subjective basis (the decoherence or classical basis) that provides us with relevant branching.

I believe further research should point out if this subjective view is a justifiable position. As a starting point, I have provided an anthropic argument that this classical stance is the best we can do *being the classical human agents that we are*, restoring to some extent the objectivity of the view. The justification of the choice of the set of basis vectors that correspond to the decoherence basis can be anthropically motivated by saying that this set of basis vectors is the only set that is *relevant* to the existence of human beings. The reason is that if we decompose the universal wavefunction in terms of another set of basis vectors, one can say that parts of the quasi-classical patterns that correspond to human beings (these are recognized in the decoherence basis) are projected onto the new basis, but one cannot say that parts of these human beings are projected onto the new basis. The same goes for the patterns of quasi-chairs and quasi-tigers. It is in this sense that human beings and other classical objects are 'un-real' in a basis which is not the classical basis.

Coming to intertheoretic reduction, Wallace also needs to fit higher-level ontology into the ontology of underlying theories. The specific functionalist way in which he does this is an intriguing one. Through instantiation the patterns of the lower-level theory and the higher-level theory can be made to match in a quantitative way—after all, instantiation is a *mapping*. Instantiation also allows for some approximations, which is to be expected during theory-reduction. Furthermore, following Michael Berry and Robert Batterman, no singular limits arise in the infinite-time limit such that no new qualitative physics is to be expected. To relate the qualitative parts of the classical theory and the quantum theory, instantiation is of no use. For this, an appeal to patterns must be made, and this is done through Dennett's criterion, which allows one to use the classical concepts as explanatorily powerful. This qualitative matching, again, can only be done from the classical stance, in which the quasi-classical patterns become useful for predictions and explanations.

In short, the classical stance is the only stance one can take for human language (i.e., classical concepts) to find traction. The patterns in other bases (which are just as real but maybe less stable) simply fall flat to our perceptional apparatus. This leaves one with a rather weaker version of 'real worlds' than Wallace argues for. Nevertheless, I think this is the best one can do in order to speak of 'real worlds' and solving the preferred basis problem in one move such as Dennett's criterion.

However, there is still something that needs to be said about the structures that arise in these other possible bases. These structures are also objectively there according to those that take the quantum state literally. According to the quantum state realist who takes all of the quantum formalism seriously, the parts of those patterns that are projected onto the new basis are (at the fine-grained level) just as real as the original patterns: they are just different patterns.

I understand that Wallace needs to counter the objection of the philosopher who claims a black-and-white view of there either being one thoroughly justified decomposition into classical worlds *or* these classical worlds are complete illusions. However, I think Wallace should put more emphasis on the shifting between the ontological stance that views the things that are real according to the theory and the classical stance, which emphasizes that from a human point of view, the classical basis is the only basis in which relevant patterns emerge. Objectively, these relevant patterns happen to be real as well, since all patterns in a real quantum state are likewise real.

Subjectively, then, Wallace promotes them to 'more real' due to the usefulness of the theories that admit them. However, I think it would be more helpful here to speak about 'useful' in the same way as Dennet does. We have seen that where Dennett speaks of useful patterns, Wallace deals with useful theories (theories also admit useless patterns). The question what an 'unreal pattern' would be for Wallace teases out this distinction.

Theories are said to be useful if they are explanatory and predictive. But to make the claim that those things that are 'real' depend on the usefulness of these theories assumes a specific position in the realism debate, which is not a consensus view in the philosophical community nor among physicists. This reminds one of the no-miracles argument and inference to the best explanation. It would aid the clarity of Wallace's case if he would assume this position in the introduction of his book.

Moreover, the specific stance that Wallace is taking is left without further explication. I have suggested that this should be a position that highlights some parts of the formalism as more real than other parts; examples of such positions are structural realism or explanationist realism. From such a position one can claim the classical basis to be the preferred one (more real than other bases). The explanationist way is to say that the choice of the classical basis is justified *because* it explains our existence and classical experiences. The structural way is to say that because *only* structures and no objects exist, all we have to do is relate structures or patterns at different scales to each other. In the structural case the quantitative instantiation relation does the job by itself. Note that in both these cases Wallace imports specific philosophical baggage.

I have the impression that Wallace invokes Dennett's criterion to establish a reality of classical worlds with the fullest ontological weight of the word 'real'—this is unnecessary if the only goal is solving the measurement problem. Indeed, one can be a quantum state realist (or even a wavefunction realist, for that matter), follow Wallace's analysis of decoherence and the recovery of the Born rule and still deny the objective multiplicity of macroscopic classical worlds! A many-minds theory can easily be conceptualized in which the measurement problem is solved along much the same lines as *The Emergent Multiverse*, while denying the reality concept inherent in Dennett's criterion. The difference is that Wallace's multiplicity of real classical worlds with definite properties is in fact a multiplicity of beliefs corresponding to the individual branches of the superposition that the observer's physical brain is in—even though the quantum state is taken as literally as possible. There is *really* a superposition.

At the end of the day, I believe that Wallace is unclear about the specific sense in which the quantum state – being the fundamental ontology of the theory –should correspond to things in the world. Although I do not expect him to solve the ontology problem entirely (after all, his book is about the measurement problem), his functionalist view does import a lot of metaphysical baggage. This compels one to be clearer about the reality of the universal wavefunction. It is true that Wallace has claimed not to be a wavefunction realist: "I see wave-function realism as in general an unhelpful way to think about the ontology of quantum mechanics."[182] The reason is that the wavefunction realist commits the same category mistake as a classical theorists who calls herself a 'phase-space-point realist'.[183]

The remark that Wallace's readers must base their interpretation of the wavefunction on is that the many-worlds interpretation is not an interpretation, but rather "just quantum mechanics itself, read literally, straightforwardly—naively, if you will—as a direct description of the physical world, just like any other microphysical theory." But quantum theory is not "just like any other microphysical theory", it suffers from the preferred basis problem. Only from the classical stance can we be naive about it. Throughout the book, it remains difficult to distil good correspondence rules between the quantum state and the world.

Although the reality claims are not clear for the fundamental level, such claims are made about higher-level ontology—through Dennett's criterion. As claimed above, however, things are also unclear at this level, since Dennett's criterion regards explanatory power and predictability as two sides of the same coin, applied as a virtue to establish reality. Antirealists would simply not follow the argument that reality is established through usefulness. This conflicts with Wallace's claim that no additional philosophical baggage besides naturalism is assumed in his book.

I suspect that a straightforward, literal reading of the wavefunction necessitates a specific position in the realism debate anyway, namely a position that downgrades some of the formalism; although structural realism or explanationism do this job. It might be simplest to be faithful to a more empirical way of explaining our definite, classical, macroscopic experiences—as Everett intended.

[182] Wallace, *Emergent Multiverse*, p. 316. And also: David Wallace, "Against Wavefunction Realism," *Current Controversies in the Philosophy of Science* (Routledge, 2017), B. Weslake and S. Dasgupta (eds.).[84].

[183] Private communication.

References

[1] Adler, S. (2001). "Why decoherence has not solved the measurement problem: a response to P.W. Anderson." *Studies in History and Philosophy of Modern Physics*, Vol. **34**, pp. 135–142. URL: https://www.sciencedirect.com/science/article/pii/S1355219802000862?via%3Dihub.

[2] Albert, D. & B. Loewer (1988). "Interpreting the Many-Worlds Interpretation." *Synthese*, Vol. **77**, pp. 195–213.

[3] Albert, D. (1990). "On the Collapse of the Wave-Function." In *Sixty-Two Years of Uncertainty*, pp. 153–165. Edited by Arthur Miller. New York: Plenum Press.

[4] Anderson, P. (2001). "Reply to Cartwright," in reply of "Science: 'A Dappled World' or a 'Seamless Web'?" by Nancy Cartwright. *Studies in History and Philosophy of Modern Physics*, Vol. **32**, pp. 499–500.

[5] Bacciagaluppi, G. (2002). "Remarks on Space-time and Locality in Everett's Interpretation." In: *Non-Locality and Modality*, pp. 105–122. Edited by: T. Placek & J. Butterfield. Kluwer Academic Publishers.

[6] Guido Bacciagaluppi & Antony Valentini. *Quantum Theory at the Crossroads*. New York: Cambridge University Press, 2009. ArXiv: https://arxiv.org/abs/quant-ph/0609184.

[7] Bacciagaluppi, G. (2016). "The Role of Decoherence in Quantum Mechanics." *Stanford Encyclopedia of Philosophy*, Fall 2016 Edition. Edited by E. N. Zalta. URL: https://plato.stanford.edu/entries/qm-decoherence/.

[8] Barrett, J. A. (2011). "On the Faithful Interpretation of Pure Wave Mechanics." *The British Journal for the Philosophy of Science*, Vol. **62**, No. 4, pp. 693–709. URL: https://academic.oup.com/bjps/article-abstract/62/4/693/1417741?redirectedFrom=fulltext.

[9] Barrett, J. A. & P. Byrne (2012). *The Everett Interpretation of Quantum Mechanics*. Princeton, NJ: Princeton University Press.

[10] Batterman, R. W. (2002). *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*. Oxford: Oxford University Press.

[11] Bechtel, W. (1993). "2.2: Axiomatic account of theories." In: *Philosophy of Science*, pp 27–29. University of Chicago Press. Edited by Werner Callebaut.

[12] Bell, J. S. (1971). "On the hypothesis that the Schroedinger equation is exact." TH.1424-*CERN*. In: Barrett, J. A. & P. Byrne, pp. 284–290.

[13] Bedau, M. (1997). "Weak Emergence." *Philosophical Perspectives*, Vol **11**: *Mind, Causation, and World*, pp. 375–399. Oxford: Blackwell.

[14] Berry, M. V. (1994). "Asymptotics, singularities and the reduction of theories." *Logic, Methodology and Philosophy of Science IX*, pp. 597–607. Editors: D. Prawitz, B. Skyrms and D. Westerståhl.

[15] Bohm, D. (1952). "A Suggested Interpretation of the Quantum Theory in terms of "Hidden" Variables I" (& II). *Physical Review*, Vol. **85**, No. 2, pp. 166–179 and pp. 180–193. URL: http://cqi.inf.usi.ch/qic/bohm2.pdf

[16] Brigandt, I. and A. Love (2015). "Reductionism in Biology." *The Stanford Encyclopedia of Philosophy*. Edited by E. N. Zalta. URL: http://plato.stanford.edu/archives/fall2015/entries/reduction-biology/.

[17] Caldeira, A. O. & A. J. Leggett (1983). "Path Integral Approach to Quantum Brownian Motion." *Physica*, Vol. **121A**, pp. 587–616.

[18] Carnap, R. (1966). "Chapter 1: The Value of Laws: Explanation and Prediction." In: *An Introduction to the Philosophy of Science*. New York: Dover Publications, 1995. Edited by Martin Gardner.

[19] Chalmers, D. J. (2006). "Strong and Weak Emergence." Chapter 1 in *The Re-emergence of Emergence*, pp. 244–56. Edited by P. Clayton and P. Davies. Oxford: Oxford University Press. URL: http://www.consc.net/papers/emergence.pdf.

[20] Chalmers, D. J. (1996). *The Conscious Mind*. Oxford: Oxford University Press.

[21] Chakravartty, A. (2017). "Scientific Realism." *The Stanford Encyclopedia of Philosophy* Summer 2017 Edition. Edited by E. N. Zalta. URL: https://plato.stanford.edu/archives/sum2017/entries/scientific-realism/..

[22] Chen, E. K. (2017). "Our Fundamnetal Physical Space: An Essay on the Metaphysics of the Wave Function." *The Journal of Philosophy*, Vol. **114**, No. 7, pp. 333–365.

[23] Davies, P. C. W. (1984). "Inflation in the universe and time asymmetry." *Nature*, Vol. **312**, pp. 524–527.

[24] Dawid R. & K. Thébault (2014). "Against the Empirical Viability of the Deutsch-Wallace-Everett Approach to Quantum Mechanics." *Studies in History and Philosophy of Science Part B*, Vol. **47**, pp. 55–61.

[25] d'Espagnat, B. (1976). *Conceptual Foundations of Quantum Mechanics*. 2nd edition. Reading: Benjamins Publishing.

[26] Deutsch, D. (1999). "Quantum theory of probability and decisions." *Proceedings of the Royal Society of London*, Vol. **A455**, pp. 3129–3137.

[27] Dieks, D. & A. Lubberdink (2011). "How Classical Particles Emerge From the Quantum World." *Foundations of Physics*, Vol. **41**, No. 6, pp. 1051–1064.

[28] Dennett, D. (1991). "Real Patterns." *The Journal of Philosophy*, Vol. **88**, No. **1**, pp. 27–51. URL: http://ruccs.rutgers.edu/images/personal-zenon-pylyshyn/class-info/FP2012/FP2012_readings/Dennett_RealPatterns.pdf.

[29] Dirac, P. (1930). *The principles of quantum mechanics*. Oxford: Oxford University Press.

[30] Donald, M. J. (1990). "Quantum Theory and the Brain." *Proceedings of the Royal Society* A, Vol. **427**, pp. 43–93.

[31] Donald, M. J. (1995).“A Mathematical Characterisation of the Physical Structure of the Observer.” *Foundations of Physics*, Vol. **25**, pp. 529–571.

[32] Donald, M. J. (1998). “Discontinuity and Continuity of Definite Properties in the Modal Interpretations.” In: *The Modal Interpretation of Quantum Mechanics*. Dordrecht: Kluwer Academic Publishers. Editors: D. Dieks, P. E. Vermaas.

[33] Donald, M. J. (1999). “ Progress in a Many-Minds Interpretation of Quantum Theory.” ArXiv: http://www.arxiv.org/abs/quant-ph/9904001v1.

[34] Everett, H. III (1957). “The Theory of the Universal Wavefunction.” *PhD dissertation*, Princeton University. In: Barrett, J. A. & P. Byrne, pp. 72–172. URL: https://archive.org/details/TheTheoryOfTheUniversalWaveFunction.

[35] Gambini, R., L. Lewowicz, J. Pullin (2015). “Quantum mechanics, strong emergence and ontological non-reducibility.” *Foundations of Chemistry*, Vol. **17**, pp. 117—127.

[36] Ghirardi, G., A. Rimini and T. Weber (1986). “Unified dynamics for microscopic and macroscopic systems.” *Physical Review*, Vol. **34**, pp. 470-491.

[37] Graham, N. (1973). “The measurement of relative frequency.” In: *The Many-World Interpretation of Quantum Mechanics*. Princeton Editors: Bryce deWitt & Neill Graham.

[38] Gopnik, A. (2000). “Explanation as orgasm and the drive for causal understanding: The evolution, function and phenomenology of the theory-formation system.” In: *Cognition and explanation*. Edited by F. Keil & R. Wilson. Cambridge, MA: MIT Press, pp. 299–323.

[39] Gell-Mann M. & J. Hartle (1993). “Classical equations for quantum systems.” *Physical Review D* , Vol. 47, No. 8.

[40] Greaves, H. (2007). “Probability in the Everett Interpretation.” *Philosophy Compass*, Vol. **2**, No. 1, pp. 109–128.

[41] Hensen, B. J. (2010). “Decoherence, the measurement problem, and interpretations of quantum mechanics.” *ArXiv* quant-ph/0105052v1. URL: http://philsci-archive.pitt.edu/5439/1/Decoherence_Essay_arXiv_version.pdf.

[42] Hooft, ’t, G. (2016). *The Cellular Automaton Interpretation of Quantum Mechanics*. Springer Open Access. *Fundamental Theories of Physics*, 185. URL: https://link.springer.com/content/pdf/10.1007%2F978-3-319-41285-6.pdf.

[43] Hovda, P. (2008). “Quantifying Weak Emergence.” *Minds & Machines*, Vol. **18**, pp. 461–473.

[44] Howard, D. (2007). “Chapter 6: Reduction and Emergence in the Physical Sciences: Some Lessons from the Particle Physics and Condensed Matter Debate.” In: *Evolution and Emergence: Systems, Organisms, Persons*, pp. 141–157. Oxford: Oxford University Press. Edited by Nancey Murphy and William Stoeger.

[45] Howard, D. (2002). “Who Invented the “Copenhagen Interpretation”? A Study in Mythology.” *Philosophy of Science*, Vol. **71**, No. 5, pp. 669–682. URL: https://www.journals.uchicago.edu/doi/10.1086/425941.

[46] Janssen, H. (2008). "Reconstructing Reality." *Master Thesis, Radboud University Nijmegen.* Supervisors: Jos Uffink & Klaas Landsman.

[47] Jackson, F. (1982). "Epiphenomenal Qualia." *The Philosophical Quarterly*, Vol. **32**, No. 127, pp. 127–136.

[48] Johnson, S. (2001). *Emergence.* New York: Scribner.

[49] Joos, E. & H. Zeh, (1985). "The Emergence of Classical Properties Through Interaction with the Environment." *Zeitschrift für Physik B - Condensed Matter*, Vol. **59**, pp. 223–243.

[50] Joos, E., H. D. Zeh, D. Giulini, C. Kiefer, J. Kupsch, I. O. Stamatescu (2003). *Decoherence and the Appearance of a Classical World in Quantum Theory.* Heidelberg: Springer.

[51] Kent, A. (2010). "One world versus many: The inadequacy of Everettian ac- counts of evolution, probability, and scientific confirmation." In: *Many Worlds? Everett, Quantum Theory, and Reality*, Chapter 10, pp. 307–355. Edited by S. Saunders, J. Barrett, D. Wallace, A. Kent. Oxford: Oxford University Press.

[52] Kitcher, P. & W. Salmon (1987). "Van Fraassen on Explanation." *The Journal of Philosophy*, Vol. **84**, No. 6, pp. 315–330.

[53] Kitcher, P. (1993). *The Advancement of Science: Science Without Legend, Objectivity without Illusions.* Oxford: Oxford University Press.

[54] Lipton, P. (2004). *Inference to the Best Explanation.* London: Routledge, 2nd edition, original 1991.

[55] Lockwood, M. (1996). "'Many Minds' Interpretations of Quantum Mechanics." *British Journal of Philosophy of Science*, Vol. **47**, pp. 159–188.

[56] Machamer, P., L. Darden, C. Craver. "Thinking about Mechanisms." *Philosophy of Science* Vol. **67**, No. 1, 2000, pp. 1-25.

[57] Magnus, P. D. & C. Callender (2004). "Realist Ennui and the Base Rate Fallacy." *Philosophy of Science*, Vol. **71**, No. 3, pp. 320–338.

[58] Maudlin, T. (2010). "Can the world be only wavefunction?" In: *Many Worlds? Everett, Quantum Theory, and Reality*, pp. 121–143.

[59] Maudlin, T. (2011, 1994). *Quantum Non-locality and Relativity.* Malden: Blackwell. Third edition.

[60] Muller, F. A. (1992). "On the Principle of Relativity." *Foundations of Physics Letters*, Vol. **5** , No. 6. Plenum Publishing Corporation.

[61] Mulder, R. A. (2017). "Quantizing Lévy Flights: Extending the Caldeira-Leggett model for velocity-dependent coupling forces in dissipative systems." Master thesis at the *Utrecht Institute for Theoretical Physics.* Supervisors: M. A. Caracanhas & C. Morais-Smith. URL: `https://dspace.library.uu.nl/handle/1874/351831`.

[62] Nagel, E. (1961). *The Structure of Science.* Harcourt: Brace & World.

[63] Norsen, T. (2017) *Foundations of quantum mechanics. An exploration of the physical meaning of quantum theory*. Springer International Publishing.

[64] O'Connor, T. & H. Wong (2015). "Emergent Properties." *The Stanford Encyclopedia of Philosophy*. Edited by E. N. Zalta. URL: https://plato.stanford.edu/entries/properties-emergent/.

[65] Tegmark, M & J. A. Wheeler (2001). "100 Years of Quantum Mysteries." *Scientific American*, Vol. **284**, pp. 68–75.

[66] Pessoa, O. (1988). "Can decoherence help to solve the measurement problem?" *Synthese*, Vol. **113**, pp. 323–346. Dordrecht: Kluwer Academic Publishers.

[67] Price, H. (1994). "Psychology in perspective." In: *Philosophy in Mind: The Place of Philosophy in the Study of the Mind*, pp. 83–98. Kluwer. Edited by: M. Michael, and J. O'Leary-Hawthorne. URL: http://prce.hu/w/preprints/psychology-in-perspective.pdf.

[68] Putnam, P. (1975). *Mathematics, Matter and Method*. Cambridge: Cambridge University Press.

[69] Rosaler, J. (2015) ""Formal" vs. "Empirical" Approaches to Quantum-Classical Reduction." *Topoi*, Vol. **34**, No. 2, pp. 325–338.

[70] Rosaler, J. (2015). "Local Reduction in Physics." *Studies in History and Philosophy of Modern Physics*, Vol. **50**, pp. 54–69.

[71] Saunders, S. (2010). *Many Worlds? Everett, Quantum Theory, and Reality*. Editors: S. Saunders, J. Barrett, A. Kent, D. Wallace. Oxford: Oxford University Press, 2010.

[72] Sommerfeld, A. (1950). *Lectures on Theoretical Physics: Optics*. New York: Academic Press.

[73] van Fraassen, B. C. (1977). "VII. The Pragmatics of Explanation." *Americal Philosophical Quarterly*, Vol. **14**, No. 2, pp. 143–150. URL: https://carlodelrazocanuto.files.wordpress.com/2010/03/bas-c-van-fraassen.pdf.

[74] van Fraassen, B. C. (1980). *The Scientific Image*. Oxford: Oxford University Press. Reprinted 2013.

[75] von Neumann, J. (1955). *Mathematical foundations of quantum mechanics*. Princeton: Princeton University Press.

[76] Wallace, D. M. (2002). "Worlds in the Everett interpretation." *Studies in History and Philosophy of Modern Physics*, Vol. **33**, p. 637–661.

[77] Wallace, D. M. (2003). "Everett and Structure." *Studies in History and Philosophy of Modern Physics*, Vol. **34**, pp. 87–105. URL: https://www.sciencedirect.com/science/article/pii/S1355219802000850.

[78] Wallace, D. M. (2007). "Quantum probability from subjective likelihood: improving on Deutsch's proof of the probability rule." *Studies in the History and Philosophy of Modern Physics*, Vol. **38**, pp. 311–332.

[79] Wallace, D. M. (2010). "Decoherence and Ontology (or: How I learned to stop worrying and love FAPP)." In: *Many Worlds? Everett, Quantum Theory, and Reality*, pp. 53–72.

[80] Wallace, D. M. (2012). "Decoherence and its role in the modern measurement problem." *Philosophical Transactions of the Royal Society* A, Vol. **370**, pp. 4576–4593. URL: http://rsta.royalsocietypublishing.org/content/roypta/370/1975/4576.full.pdf.

[81] Wallace, D. M. (2012). *The Emergent Multiverse*. Oxford: Oxford University Press.

[82] Wallace, D. M. (2013). "The Everett Interpretation." *The Oxford Handbook of Philosophy of Physics*. Edited by Robert Batterman. Oxford: Oxford University Press.

[83] Wallace, D. M. (2017). "Against Wavefunction Realism," *Current Controversies in the Philosophy of Science* (Routledge, 2017), B. Weslake and S. Dasgupta (eds).

[84] Wallace, D. M. & C. G. Timpson (2010). "Quantum Mechanics on Spacetime I: Spacetime State Realism." *British Journal for the Philosophy of Science*, Vol. **61**, pp. 697–727.

[85] Xin, M., E. N. Olson, R. Bassel-Duby (2013). "Mending broken hearts: cardiac development as a basis for adult heart regeneration and repair." *Nature Reviews Molecular Cell Biology*, Vol. **14**, pp. 529–541.

[86] Zeh, H. (1970). "On the Interpretation of Measurement in Quantum Theory." *Foundations of Physics*, Vol. **1**, pp. 69–76.

[87] Zeh, H. (1999). "The Problem of conscious observation in quantum mechanical description." *Foundations of Physics Letters*, Vol. **13**, No. 3, pp. 221–233.

[88] Zeh, H. (1979). "Quantum Theory and Time Asymmetry," *Foundations of Physics*, Vol. **9**, No. 11-12, pp. 803–818.

[89] Żurek, Wojciech H. (2002). "Decoherence and the Transition from Quantum to Classical-*Revisited*." *Los Alamos Science* 2002, No. **27**.

[90] Żurek, W. H. (1998). "Decoherence, Einselection and the Existential Interpretation (The Rough Guide)." *Philosophical Transactions of the Royal Society*, **A356**, pp. 1793–1821.

The verse on the first page is from *Sappho*, "Come Close," translated by Aaron Poochigian in *Stung with Love: Poems and Fragments* (2009), Penguin Classics.