Reconstructing Reality

Environment-induced decoherence, the measurement problem, and the emergence of definiteness in quantum mechanics
A critical assessment

Hanneke Janssen

*

Master's thesis - Theoretical Physics
Radboud University Nijmegen

*

Supervisors: Jos Uffink & Klaas Landsman

…Bovenal vreesde hij de verbeelding, die metgezellin met haar twee gezichten, waarvan de één een vriendelijke, de ander een vijandelijke uitdrukking draagt; vriendelijk, hoe minder geloof je aan haar schenkt, vijandig, wanneer je onder haar lieflijk gefluister in slaap valt …

Iwan Aleksandrowitsj Gontsjarow, *Oblomov* (1859)

Voorwoord

Voor u ligt de scriptie waarmee ik na acht jaar mijn studie natuurkunde aan de Radboud Universiteit Nijmegen afsluit. Deze scriptie is geschreven in het kader van de onderzoeksmaster theoretische natuurkunde, maar de inhoud wijkt wat af van wat voor zo'n scriptie doorgaans gebruikelijk is. Ongeveer drie jaar geleden heb ik namelijk besloten mij toe te gaan leggen op de filosofie en grondslagen van de natuurkunde, in het bijzonder die van de quantummechanica.

De aanleiding voor deze "carrière-omslag" was een groeiend gevoel van ontevredenheid met de studie; hoewel ik in staat bleek om alle vakken met een positief resultaat af te sluiten bleef er toch het gevoel knagen dat ik iets miste. Dat ik weliswaar volgens de geldende criteria de stof voldoende meester was, maar dat ik het toch niet helemaal had begrepen. Niet *echt* begrepen. Het is fijn om al je tentamens te halen, maar echt inzicht hebben in een theorie, waarom het formalisme is zoals het is, wat de diepere samenhang is tussen de minimale aannames die je erin stopt en de verregaande conclusies die je eruit trekt, wat voor wereldbeeld er ten grondslag ligt aan de theorie, in hoeverre dat standhoudt, en wat een theorie aan ons wereldbeeld toevoegt – dat waren allemaal zaken waar ik door mijn groeiende vaardigheid in het oplossen van partiële differentiaalvergelijkingen geen bevredigend antwoord op kreeg.

Het oplossen van partiële differentiaalvergelijkingen ben ik intussen wel verleerd. Evenmin heb ik een antwoord op de bovenstaande vragen gevonden. Toch ben ik blij dat ik ervoor gekozen heb het roer om te gooien. Want naast het besef dat het naïef is om te denken dat iemand na acht jaar studie het soort dieper inzicht kan verkrijgen waar ik aanvankelijk op gehoopt had, heb ik vooral ontdekt dat dit soort filosofische mijmeringen de vruchtbare grond zijn waar een weelde aan wetenschappelijke inzichten aan kan ontspruiten. Maar meer nog heeft de filosofische zienswijze mijn kritisch denkvermogen verscherpt, waardoor ik meer en meer de beperkingen van deze inzichten ben gaan inzien. Deze scriptie is daar feitelijk een illustratie van.

In het bijzonder ben ik er niet meer van overtuigd dat de natuurkunde ons het "ware gezicht van de werkelijkheid" toont. Onvoorwaardelijk geloof in de wetenschap heeft plaatsgemaakt voor een wat gematigder houding die uitgaat van het vermogen van de fysica om de beperkingen van ons "gezond verstand" aan te tonen – mits men ook zijn gezond verstand gebruikt om fysische theorieën op hun waarde te schatten. In dat opzicht heeft de natuurkunde voor mij niet aan kracht ingeboet, omdat zij nog altijd laat zien dat de realiteit veel complexer en vooral ongrijpbaarder is dan onze dagelijkse ervaringen ons doen geloven. Hier begint de verwondering, en ik prijs mij gelukkig dat ik die door mij met deze vragen bezig te houden weer hervonden heb.

Daarvoor moet ik in de eerste plaats Jos Uffink bedanken, omdat zijn college over de grondslagen van de quantummechanica voor mij de bevestiging was dat ik mij verder in dit onderwerp wilde verdiepen. Ik ben hem, en eveneens Dennis Dieks, dankbaar dat ik als student van een andere universiteit in Utrecht zo hartelijk werd ontvangen, dat ik de kans heb gekregen om me aldaar verder in het vakgebied te bekwamen door het volgen van colleges, colloquia en de mogelijkheid om met studenten en staf in discussie te treden, en bovenal dat ze mij de positie van Managing Editor bij Studies in History and Philosophy of Modern Physics toe hebben vertrouwd. Ik heb het altijd als een groot voorrecht beschouwd dat ik nog tijdens mijn studie bij heb mogen dragen aan de totstandkoming van een dergelijk internationaal toonaangevend tijdschrift op het gebied van mijn voornaamste wetenschappelijke interesse. Ongetwijfeld is Studies een van de belangrijkste factoren geweest die hebben bijgedragen aan mijn belangstelling, inhoudelijke kennis en vertrouwdheid met het vakgebied.

Daarnaast ben ik Jos, en mijn andere begeleider Klaas Landsman, bijzonder dankbaar voor de toewijding waarmee ze steeds weer de zoveelste nieuwe versie van een of ander hoofdstuk van mijn scriptie hebben gelezen en becommentarieerd. Over het algemeen waren ze er een stuk tevredener over dan ikzelf, maar toch heeft hun kritiek in belangrijke mate bijgedragen tot het eindresultaat. Soms werd het allemaal enigzins verwarrend, maar de worsteling om die verwarring te boven te komen werd altijd beloond met een of ander belangrijk inzicht. Deze scriptie heeft in feite alles te danken aan de momenten dat Klaas en Jos mij over mijn twijfels heen hebben weten te trekken en mij de tekortkomingen, maar veel vaker nog de waarde, van mijn ideeën in hebben doen zien.

Ik ben ook de mensen van de afdeling theoretische hoge-energiefysica in Nijmegen dankbaar voor het bieden van een aangename werkplek, en dat ze al die tijd een "filosoof" (dat is iemand die zich bezighoudt met het meetprobleem) zonder al te veel morren in hun midden hebben geduld.

Rest mij nog om familie en vrienden te bedanken. Hoewel ze niet inhoudelijk aan de totstandkoming van deze scriptie bij hebben kunnen dragen, ben ik iedereen, in het bijzonder mijn goede vriendin Lucia, dankbaar die keer op keer geduldig mijn warrige verhalen over de quantummechanica aan heeft zitten horen en, in mindere tijden, mijn klaagzangen over de last die een scriptie kan zijn. Mijn lieve ouders tenslotte ben ik om veel dingen dankbaar, maar hier volstaat het hun niet aflatende en onvoorwaardelijke vertrouwen in een goede afloop te noemen. Zoals zo vaak, hadden ze natuurlijk ook nu weer gelijk.

Nijmegen, Juli 2008.

Contents

Voorwoord **iii**

Introduction **1**

1 Interpretations of quantum mechanics **5**

1.1 The measurement problem 5

1.1.1 Problems and solutions 6

1.2 Decoherence and the emergence of definiteness 24

1.2.1 Interference vs. indefiniteness 24

1.2.2 The ignorance interpretation of mixtures 27

1.2.3 Measurement as an emergent phenomenon 30

1.3 Concluding remarks 32

2 Environment-induced decoherence **37**

2.1 Preliminaries 37

2.1.1 A brief historical and conceptual overview 38

2.2 Decoherence and the measurement problem 41

2.2.1 Environment-induced decoherence. 42

2.2.2 Solving the preferred-basis problem 45

2.2.3 Environment-induced superselection rules 49

2.3 Decoherence and the emergence of classicality 53

2.3.1 Robustness and decoherence as a source of classicality. 53

2.3.2 Reducing classical physics to quantum mechanics. 54

2.3.3 Decoherence and the emergence of classicality 55

2.4 Concluding remarks 61

3 Philosophical issues **63**

3.1 General philosophical background 63

3.1.1 Decoherence and the (dis)appearance of a classical world in quantum theory 64
3.2 Reconsidering the decoherence argument 68
3.2.1 Decoherence: what it means. 68
3.2.2 . . . and what it takes. 72
3.3 Some (partial) answers and more questions 74
3.3.1 The selection criteria and the status of observables 74
3.3.2 The ignorance interpretation (revisited) 81
3.3.3 Approximate decoherence 85
3.3.4 Subsystems and the environment 91
3.4 Concluding remarks 97

4 Remarks on the many-worlds interpretation 101
4.1 The main issues 101
4.2 Zurek's "Existential Interpretation" 103
4.2.1 Zurek's argument 103
4.2.2 Commentary 108
4.3 Everett's relative states, many minds and many worlds 121
4.3.1 The central concepts 121
4.3.2 Why the Everett interpretation does not work 123
4.4 A modern Everettian account of classicality 126
4.4.1 Wallace and emergent reality 126
4.4.2 The MWI made plausible? 132
4.5 Concluding remarks 138

Conclusion 141

Outlook 153

Bibliography 157

Introduction

The subject of this thesis is the foundations of quantum mechanics. The nice thing about this subject is that it touches upon a number of different disciplines: physics, mathematics, and philosophy. Initially, I started to look upon the subject as a physicist, but gradually the discussion turned more and more philosophical. Therefore, let me start by pointing out what I mean by "philosophical" here, because it may not conform completely to what physicists usually associate with the phrase.

For me, physics and philosophy have always been intimately related subjects, though not in the sense that bringing philosophy into physics would be an excuse for far-fetched metaphysical speculations. Rather, in my view the role of philosophy is most of all to look critically at physics; to see what the underlying assumptions of a physical theory are, whether these assumptions are justified, to find out what the theory tells us the world is like and most of all what it tells us the world is *not* like. For me, the impact of physics lies especially in the latter point; although physical theories can motivate certain metaphysical speculations, they can never warrant them. But physical theories *can* unambiguously show that whatever the ultimate reality may be, it will not conform to our common sense. And quantum mechanics is a particular striking illustration of the latter point.

But what exactly is the trouble with quantum mechanics? Perhaps there is none; by all standards the theory is extraordinarily successful. It ranges over a strikingly broad spectrum of applications and its predictions, although merely probabilistic, are highly accurate. In a sense, it is the best theory of physics we have ever had. In another sense, however, it is completely unsatisfying. This is best seen by considering the fact that the empirical content of the theory is captured by the Born rule, that says that *a measurement of the observable* $\hat{A} = \sum_i a_i |a_i\rangle\langle a_i|$ *in a state* $|\psi\rangle$ *yields a probability for obtaining the eigenvalue* a_i *of* $\hat{A}$ *equal to* $|\langle a_i|\psi\rangle|^2$. It is sometimes believed that the problem lies in the "probabilities", but I think that is not quite so; a theory need not necessarily be deterministic to make sense. Rather, I think the problem is in the "measurement": the dissatisfaction stems in the first place from the fact that the formalism does not give us anything that goes *beyond* those predictions. After more than half a century of foundational research, and numerous proposals for a coherent interpretation of the formalism, it still remains obscure what kind of reality could underly the quantum phenomena. This is particularly so if one chooses not to look beyond the orthodox interpretation[1]. Although I do not think that there is, at the moment, an interpretation available

---

[1] In chapter 1 I will make precise what I mean with the orthodox interpretation, which I take to be different from the Copenhagen interpretation.

that provides a convincing answer to this question, it seems to me a task of the physicist as well as the philosopher to keep exploring the possibilities. It turns out to be a real challenge, but if we want physics to teach us something about the world than it seems to me that the orthodox interpretation fails to do so. Foundational research has so far not given us a complete and consistent picture of the quantum world, and I doubt that it ever will. But if there is one thing that foundational research has taught us, it is that this world would have to be something very unlike the everyday classical reality of our intuition and experiences.

It must be stressed that the phenomena, and not so much the formalism that captures them mathematically, defy our classical intuition already. I am not going to spell out in detail what these peculiar phenomena are, or what is so peculiar about them[2]. There are plenty of good books available that explain this far better than I can. It is not relevant to the present discussion, in any case. I will simply (but harmlessly, I suppose) assume that quantum mechanics is the one and only true theory that encompasses all these phenomena, largely ignore the facts and confine myself to discussing the theory. But the crucial point that I will explore in the first chapter, is that no interpretation of the theory is available yet that fits all these phenomena in a single conceptual scheme –at least not if the latter is to fulfill the additional requirement that no *ad hoc* assumptions are introduced that merely seem to serve to secure empirical adequacy.

So far, I have discussed quantum mechanics as a threat to realism. This is indeed the most direct and clearest way to appreciate the difficulties. Of course, if one is interested in statistical predictions for measurement outcomes only, and does not worry as to what these outcomes or probabilities refer to, then quantum mechanics is a rewarding subject of study. Nonetheless, I do think that quantum mechanics is not just problematic for a realist, and in the four chapters following this introduction I will occasionally point out why I believe this to be the case[3].

As the above remarks indicate, I chose not stick to one particular epistemic attitude from which to consider the interpretational difficulties of quantum mechanics, but rather move back and forth a bit between realism and anti-realism. This is partly so because I have not settled on a specific philosophical view of science yet and tend to change my mind now and then, but mostly because I do not wish to consider the foundational problems from a fixed perspective. Rather, I hope that switching between these different points of view allows me to give a broad and unbiased overview of the problems. Although the question of what, in the light of quantum mechanics, would be the proper epistemic attitude is not the main focus of this thesis, I will occasionally point out what certain questions and approaches imply with respect to this question, in particular when I turn to discussing environment-

---

[2] Non-classical correlations that seem to imply a strange action-at-a-distance, or the peculiar "wavelike" behavior of particles in interference experiments, for instance.

[3] Consider the fact that also Bas van Fraassen, whose constructive empiricism is a thorough anti-realist (though not instrumentalist) view of science, demands that a theory should provide some kind of world picture in order to be intelligible. As he argues: "Ideally, belief presupposes understanding. This is true even of the mere belief that a theory is true in certain respects only. *Hence we come to the question of interpretation:* Under what conditions is this theory true? What does it say the world is like? [. . .] The question of interpretation – what would it be like for this theory to be true, and how could the world possibly be the way this theory says it is? – does indeed go beyond almost all discussions in science. But it is also broached by scientists, just because the theory about the phenomena can rarely be well understood except as part of what it says the world as a whole is like" (van Fraassen, 1991, p.242).

induced decoherence in chapter 3. That is, I will argue that certain alleged problems and answers presuppose a certain epistemic attitude (even though this attitude is generally not made explicit), and that in some cases these presuppositions conflict with one another.

After these general remarks, let me turn to the main focus of this thesis: the measurement problem and environment-induced decoherence. The measurement problem has been said to "dominate the philosophy of quantum mechanics" (Wallace, 2007), and as we will see in chapter 1 it does indeed touch upon about all other foundational questions. (Although I think that is above all because *all* foundational questions are in a way related.) Above, I quoted the Born rule and remarked that its problem lies in the notion of measurement. This is because of the instrumentalist connotations it bears, but furthermore the problem is that quantum mechanics is formulated in terms of the very processes (measurement interactions) it is supposed to describe. (There is more to it, however, as I will spell out in the first chapter.) This apparently paradoxical situation has led to the introduction of the notorious "collapse of the wave packet", which is often identified with the measurement problem itself. Generally speaking, one might say that the various interpretations that have been proposed aim at providing an alternative or explanation for the collapse postulate. (But one should bear in mind that the collapse postulate itself is an artefact of a particular interpretation, namely von Neumann's.) Environment-induced decoherence has been proposed in the 1980's as a physics-based solution to the problem of collapse, and has gradually gained support in the physics as well as the philosophy of physics community.

Despite the enthusiasm with which it has been received, it has been argued for quite some time that environment-induced decoherence does not solve the measurement problem. Nevertheless, there is a growing consensus that the mechanism of environment-induced decoherence is a "crucial dynamical step" (Zeh in Joos et al., 2003, p.37) in the measurement process and should therefore be taken into account by *any* interpretation. To substantiate or weaken this claim will be the central subject of this thesis. Specifically, I will attempt to answer the following question:

> *What exactly does environment-induced decoherence contribute to a solution of the measurement problem?*

Roughly speaking, the outline of this thesis is as follows:

- In the first part of chapter 1 I introduce the measurement problem and various interpretations that have been formulated in response to it. I will give a specific formulation of the problem that seems to be the primary target of the general class of "decoherence" approaches to which, obviously, environment-induced decoherence belongs. The second part consists of a critical assessment of the "decoherence" approaches in general. Here I will also present the well-known standard critique of environment-induced decoherence (which, accordingly, does not have so much to do with the "environment-induced" part.)
- In the second chapter I will take a closer look at the specific features of environment-induced decoherence. I will largely confine myself to giving a broad overview of the literature, although I will be somewhat more critical when discussing the alleged successes of decoherence with respect to the

"emergence of classicality" if the latter is understood in a more general sense than just the measurement problem.

- I will comment upon the relevance of decoherence for the latter issue in chapter 3. After giving a rather general account of the philosophical implications of the decoherence program, I will reconsider the theory as it was presented in chapter 2 from my personal point of view, and point out various tacit assumptions underlying the approach and the difficulties these give rise to. Along the way, I will also reconsider the objections raised in chapter 1 in the light of the specific methods employed by environment-induced decoherence. The upshot of it all will be that the "environment-induced" part does not only fail to alleviate the concerns raised in chapter 1, but in fact even introduces further difficulties.
- Finally, in chapter 4, I will consider in what sense decoherence is relevant to Everettian interpretations, and vice versa, if we are to solve the measurement problem. In particular, I will focus on the question whether the Everettian approach does *more* than addressing the well-known "problem of outcomes" associated with the decoherence approach (as discussed in chapter 1), since in chapter 3 I concluded that there are a number of other difficulties that should be taken care of as well.

Each chapter will be closed by some concluding remarks, that aim to to distill some general points of concern that will be taken up in the chapters following. In the final Conclusion I will answer the main question by analysing the arguments of the foregoing chapters in terms of a number of specific sub-questions. Finally, in the Outlook I will give a brief survey of some open problems.

Before we move on, a final note about the title of this work. I chose to name it as I did since I think it captures the main thrust of the discussion nicely. As I just explained, and will explain more elaborately in the remaining chapters, I think that the central problem with quantum mechanics is the sheer difficulty of a realist reading of the formalism. This is certainly the case for a particular version of the measurement problem that environment-induced decoherence aims to address, namely the problem that quantum mechanics cannot account for "facts", or the "definiteness of the macro world", as it is often called. The idea of environment-induced decoherence, and to a certain extent also of the Everettian approaches that I discuss in chapter 4, is that such a realist reading of the quantum formalism as pertaining to the macro world *is* possible, even if we do not know what kind of micro-ontology it implies. I think this reasoning is thoroughly flawed, but I will postpone further discussion to the Conclusion. As the subtitle indicates, the tone of the discussion will most of all be critical (and unfortunately not very constructive), since, as I stated at the start of this Introduction, that is what I think a philosophical discussion of physics should be.

Chapter 1

The measurement problem and interpretations of quantum theory

This chapter is an introduction to the traditional interpretational problem of quantum theory: the measurement problem. In the course of section 1.1 I will briefly discuss various interpretations that have been formulated as a response to this problem, avoiding most of the technical details. Although my discussion will not be exhaustive, I hope to make clear that no two different strands of interpretation exist that address exactly the same version of the measurement problem. In the course of the discussion I will formulate a specific version of the problem and introduce a corresponding class of proposed solutions, which I will call "decoherence" approaches.

In section 1.2 I will formulate some points of criticism of the decoherence approaches in general, to the extent it does not bear on the specific methods employed by "environment-induced decoherence" (which will be discussed in the chapters following). This general criticism will set the stage for the more detailed discussion in chapter 3. In subsection 1.2.3 a more recent approach to the measurement problem, that sees the latter as part of the broader question of how the classical world emerges from the quantum, is also introduced and criticized.

1.1 The measurement problem

Since the conception of quantum mechanics in its modern formulation its conceptual implications have been the subject of much dispute. The controversy over its interpretation was not settled with the quiet subsiding of the Einstein-Bohr debate: up to this day the challenge quantum mechanics poses to our understanding of the world gives rise to a vast amount of publications, addressing the same questions over and over again, apparently without much success. These questions are not confined to the domain of the philosopher; more and more often foundational questions enter the mathematical *and* experimental investigations of the physicist. This indicates that these foundational issues are not merely a matter of the philosopher's appetite for extratheorical speculation or his curiosity for what there is beyond the tangible world of facts. There seems to be much more at stake.

What there is at stake is partly a metaphysical issue: it is notoriously difficult to pin down what kind of reality quantum mechanics confronts us with. In particular, quantum mechanics seems to pose a serious threat to some of the cornerstones of physical science; most notoriously, determinism and (or) locality. But there is also a more down-to-earth objection, which is that without the addition of extra interpretational rules the theory suffers from obscurities in its formulation that have to be clarified, regardless of any kind of metaphysical prejudices one may have. Simply put, the difficulty is that without the projection postulate the quantum formalism fails to account for the facts of our experience.

The problem becomes most prominent in the context of measurements, where the quantum world is connected to the observable macroscopic realm, and is therefore usually referred to as "the measurement problem". And in fact, anyone who has pondered over it for a while would agree that there *is* some kind of problem. However, that is about where the consensus ends. A precise formulation of the measurement problem that covers all the various approaches in the literature is hard, if not impossible, to find. In this section I will try to give an overview of the various formulations of the problem, starting from a rather general and often heard criticism of the orthodox interpretation that will gradually be refined to end with a precise formulation of what the measurement problem is that environment-induced decoherence, the approach that will concern us in the following chapters, aims to solve. Along the way I will discuss how various other interpretations of quantum mechanics try to deal with the various questions that can be raised in connection with the measurement problem[1].

1.1.1 Measurement in quantum mechanics – problems and solutions

The Copenhagen and orthodox interpretations of quantum mechanics

What I will call the orthodox (or standard) interpretation of quantum mechanics is not really an interpretation. It corresponds to the pragmatic implementation of the formalism as it is used by the practising physicist; it takes the quantum state as a tool to compute probabilities for observable phenomena, and it makes no attempt to explain or motivate the projection postulate (to be discussed below). It turns out to be well suited to the practical needs of the physicist, but in this case, this pragmatic attitude does not merely reflect a disinterest for the profound metaphysical challenges raised by quantum mechanics: as I will argue below, I think that it also reflects a stubborn ignorance of a serious conceptual flaw in the axiomatic foundations of the theory.

This orthodox interpretation is not to be confused with the Copenhagen interpretation (a collective term for the, only partially overlapping, ideas of Bohr, Heisenberg and Pauli), of which it is rather a caricature deprived of all philosophical content. Although the Copenhagen Interpretation leaves many questions untouched, it nev-

---

[1] Dickson (2007) and Wallace (2007) provide a similar survey of the measurement problem and interpretations of quantum mechanics, although their presentation differs from mine.

ertheless provides a philosophical framework that goes far beyond the instrumentalist orthodox position (that is most often associated with Dirac). In contrast to most of the other interpretations that will be discussed in what follows, the members of the Copenhagen school developed their views before the measurement problem came to dominate the philosophical debate about quantum mechanics. (In fact, the measurement problem originated from Einstein's and Schrödinger's critique of the Copenhagen Interpretation, see for instance (Fine, 1996)). Nevertheless, Bohr does have some important things to say about the role of measurement and observation in quantum mechanics, and in fact many of these ideas are in a sense revived in the decoherence-based interpretations that I will discuss in chapter 4. I cannot possibly cover all the historical and philosophical issues concerning the Copenhagen interpretation in full depth, as the literature devoted to it is huge[2]. Here I will confine myself to the main characteristics of the Copenhagen interpretation that are of special relevance to the present discussion.

The first is the"doctrine of classical concepts". Bohr insisted on the use of an unambiguous language as a precondition for objective knowledge. To him, this unambiguous language was the language of classical physics. As a consequence, the measurement apparatus, being the part of the physical system that connects the quantum realm to observation, needs to be described in classical terms. There is thus a strict division between the object and the apparatus – the "Heisenberg cut" – corresponding to the distinction between the quantum and classical mode of description. But since this distinction is epistemological (or semantical), and not ontological, the location of the cut is (to a large extent) arbitrary. Second, Heisenberg's uncertainty relations imply that there are limitations to the use of these classical concepts which are to be determined by the experimental setup. Properties of a physical system that can be defined only in mutually exclusive experimental arrangements (such as the position and momentum of a particle) are said to be "complementary", meaning that they are on the one hand mutually exclusive but, on the other, they are also both essential parts of the complete physical description[3]. Finally, in spite of the conceptual split between the observed and the observer (or apparatus), together they form an indivisible whole, referred to by (the later) Bohr as a "phenomenon". In contrast to the situation in classical physics, the mutual interaction between the measured system and the apparatus is never negligible, due to Planck's quantum of action (what Bohr called the "essential discreteness" of atomic processes). Since this discreteness is "completely foreign to classical theories" this interaction is beyond the grasp of our classical description (and therefore our understanding). The consequence is a non-causal appearance of the phenomena.

This is the Copenhagen interpretation crammed in a nutshell. It should be read as a semantical and epistemological stance, guiding our understanding of quantum mechanics and the conceptual difficulties associated with it, rather than as a formal account of a solution to those difficulties. The orthodox interpretation is what remains when the quantum formalism is lifted from this philosophical back-

---

[2] References may be found in Landsman (2007), who also discusses the Copenhagen interpretation in considerable detail. A concise introduction is Faye (2002).

[3] The account of complementarity given here is the most comprehensible, but Bohr used the phrase in different contexts as well, often in contradictory ways. In fact, the notion of complementarity is perhaps the most obscure element of Bohr's philosophy. See the previously mentioned references for a more elaborate discussion.

ground (possibly whilst maintaining the idea of a "classical" measurement apparatus without further philosophical motivation). Perhaps Bohr's philosophy raises more questions than it answers and a simplified formal account may be essential for practical applications; but, as I will argue below, the orthodox interpretation tries to deal with a seemingly nonsensical consequence of the theory by introducing an artificial *ad hoc* manoeuvre that only leads to further obscurity - and the questions it consequently gives rise to are no longer backed up by a philosophical framework.

Measurements in the orthodox interpretation

In the Introduction I mentioned the Born rule:

> A measurement of the observable $\hat{A} = \sum_i a_i |a_i\rangle\langle a_i|$ on a state $|\psi\rangle$ yields a probability for obtaining the eigenvalue a_i of $\hat{A}$ equal to $|\langle a_i|\psi\rangle|^2$.

I criticized this rule on grounds of its strong instrumentalist connotations: if the theory is formulated in terms of measurement outcomes only, and if this minimal empirical content is all it offers us as a description of reality, than that theory is hardly a description of reality at all. It is just a calculational tool to predict measurement outcomes, without any conceptual background that can serve as an explanation of these phenomena. Thus, the theory brings with it a thorough instrumentalism that is unacceptable to many.

Furthermore, the crucial role for "measurements" seems misplaced in a thoroughly axiomatized physical theory. For what determines whether the Born rule can be applied? What is a measurement? Is it an ordinary physical interaction? Or does it include observation by a conscious observer? Without a clear-cut answer to these questions, the theory is simply incomplete even from an instrumentalist point of view[4].

This means that the first central question one would need to answer is the following:

> **Measurement:** What is a measurement?

Now the Born rule is not the only place where the problematic concept of "measurement" enters. The Born rule is often neglected in discussions of the measurement problem, since many authors seem to think that the *real* problem is in the projection postulate, or collapse postulate (von Neumann's "first intervention")[5]:

> If the outcome of a measurement of an observable corresponding to an Hermitian operator $\hat{A}$ is the eigenvalue $a_k \in Spec\{\hat{A}\}$, then immediately after the measurement the system will be in the eigenstate $|a_k\rangle$

---

[4] As van Fraassen (1991) argues, it is possible to develop such criteria that designate certain interactions as "measurements". The objection that the phrase "measurement" implies that the orthodox interpretation is anthropocentric is therefore misguided.

[5] The collapse postulate marks an essential difference between the Copenhagen and the orthodox interpretation. As Faye (2002) points out, there is no such thing as a 'collapse of the wave function' in Bohr's view, since Bohr believed that the wave function is merely symbolical, a statistical algorithm, but does not represent some real entity. (It does not have a "pictorial representation" as Bohr puts it.)

> corresponding to that eigenvalue[6].

The point is that if the “measurement” in the collapse postulate would just be an ordinary physical interaction, the latter is not supposed to appear in the axiomatic formulation of the theory at all. In fact, one *can* model measurements physically, but no reasonable model seems to achieve what the collapse postulate proclaims. The collapse postulate is therefore more than a practical ”rule of the thumb”, but is there –in the basic axioms of the theory– to secure certain empirical facts that otherwise cannot be accounted for. However, if one does not want to accept that there is something inexplicable about measurements that justifies their axiomatic treatment, the following question needs to be addressed:

> **Formalism:** Can quantum mechanics be reformulated without referring to “measurements” in its axioms?

Note that this question does not necessarily have to be addressed from a physical point of view, i.e. by giving a detailed analysis of the measurement process. From a more philosophical perspective, the projection postulate could perhaps be abandoned by providing a suitable interpretation of the quantum state (i.e. ascribing more meaning to it than as a calculational tool to compute probabilities). As will be argued in more detail below, such an interpretation should in the first place clarify the meaning of the superposition principle. In fact, the von Neumann formalism only ascribes a clear empirical meaning to an eigenstate of an observable, but not to a superposition of eigenstates. (That this is the core of the problem is clearly recognized by modal interpretations, cf. the discussion below.) So before turning to the physics of measurement interactions I will first have a look at the philosophical side of the problem.

The eigenstate-eigenvalue link

The essence of the projection postulate is that after measurement, the quantum state has to be an eigenstate corresponding to the measured outcome. The main motivation for this is that measurements are (in principle) repeatable, and if a measurement is repeated on the same quantum system *right after the first* (before the system has evolved significantly) one will find the same outcome *with certainty.* Any interpretation will have to deal with this issue (Maudlin (1995) calls this the “problem of effect”)[7].

Usually it is argued that finding measurement outcomes with certainty means that the quantum state *is* the corresponding eigenstate. However, although it is true is that *if* the state is an eigenstate of a certain observable, measuring this observable will yield as outcome the corresponding eigenvalue with certainty, this does *not* mean that we, conversely, need to assume that obtaining a particular value in a measurement implies that the quantum state is the corresponding eigenstate. The

---

[6] Here it is assumed that $\hat{A}$ is maximal, i.e. to every eigenvalue corresponds a unique eigenstate. In case of degeneracies (higher-dimensional eigenspaces corresponding to one or more of the $a_i \in Spec\{\hat{A}\}$) the collapse postulate as formulated here does not apply and must be replaced by Lüders rule, which states that the resulting state after measurement will be the orthonormal projection onto the eigenspace of the measured eigenvalue a_k.

[7] But see (van Fraassen, 1991, sect. 8.5) for arguments why this principle is spurious.

assumption that there is a necessary, one-to-one connection between eigenvalues as measurement outcomes and eigenstates is called the *eigenstate-eigenvalue link* (henceforth e-e link). Appealing as the symmetry of the argument is, it is by no means forced upon us either by logic or by empirical considerations.

The class of interpretations that aim to solve the measurement problem by dropping the e-e link are either hidden-variable theories or *modal interpretations*[8]. I will discuss hidden-variable theories later. An essential difference is that hidden-variable theories are motivated by realism; they posit an underlying determinate structure that is supposed to represent the (unknown) 'cause' of the empirically accessible events. Modal interpretations rather seek to find general rules that fix a maximally allowed set of possible determinate (or 'modally possessed') properties but generally have no ontological commitment[9]. Moreover, these rules do not yield the actual possessed values of a system, but interpret the wave function as a catalogue of *possibly* possessed values[10]. The observable that is deemed 'determinate', and the set of possible values it therefore may take, is generally state-dependent. (But this may depend on the exact definition of "modal interpretation" one maintains.) Unfortunately, modal interpretations suffer from their own technical and conceptual problems (like how to deal with the problem of effect), but the point I wish to make is that the e-e link is an additional assumption behind the collapse postulate which is often not made explicit.

The e-e link is often identified with the collapse postulate. This is quite wrong. Banning the e-e link is a way to avoid the need for a collapse, but there are interpretations that maintain the e-e link but not the collapse postulate. Examples of the latter are the class of relative-state interpretations (Everett-, many worlds or many minds interpretations). The relative-state interpretation was originally formulated by Hugh Everett in 1957 and popularized (among others) by DeWitt and Wheeler, who famously claimed that it showed how "quantum mechanics is capable of yielding its own interpretation". Of course, they were wrong, but what Everett *did* show was that if one assumes that there is a "state vector of the universe" one can assign a state to some subsystem (a pointer, for instance) and correspondingly a (uniquely defined) "relative state" of the rest of the universe allowing the state assignment of the subsystem to be treated in a (statistically) consistent manner. However, the universe remains in a superposition state and no collapse whatsoever occurs. Since Everett never made clear what this is supposed to mean exactly, DeWitt and Wheeler adapted Everett's framework to yield what nowadays is better known as the "many worlds" interpretation. What it adds to the relative-state formulation is radical metaphysics: every term in a superposition represents an equally real world in which this outcome is realized. Consequently, the universe is continuously

---

[8] Modal interpretations drop the eigenvalue to eigenstate link but maintain the eigenstate to eigenvalue link. (A possible exception is Bohmian mechanics, which I will therefore characterize as a hidden variable theory. But see Bub (1997) for a different point of view).

[9] I will maintain this distinction between modal interpretations and hidden-variable theories in what follows. But the distinction is not generally accepted. Some authors (Maudlin, 1995; Dickson, 2007) characterize modal interpretations as hidden variable theories, whereas others (Bub, 1997) hold exactly the opposite. In the characterization of Bub, a modal interpretation is a theory that poses a (possibly state-dependent and therefore time-dependent) preferred determinate observable which can take values for a given state. (Consequently, Bub also counts the orthodox interpretation (with the unit operator as the preferred observable) and the Copenhagen interpretation (according to which the preferred observable is determined by a specific measurement set-up) as 'modal'. See also (van Fraassen, 1991).)

[10] That is, under the restrictions imposed by the Kochen and Specker theorem (see below).

splitting into different 'branches' which cannot communicate with each other in any way (except on the rare occasion of quantum interference)[11]. Despite the rather awkward world view it presents, the many-worlds view has gained in popularity in the past decades, mainly because of its application to quantum cosmology (in which there is no "external observer" to ground the empirical meaning of the state vector in probabilities of measurement results). Another partial reason for this renewed interest is that the relative-state picture seems to combine pretty well with the mechanism of decoherence – something I will discuss in chapter 4, where I will also consider the formalism of the Everett interpretation and formulate some points of criticism.

There is however an obvious reason why the e-e link is most generally associated with some kind of collapse (i.e. a non-unitary change of the quantum state). To see why, consider the measurement process in its simplest form (the von Neumann measurement scheme)[12]:

One starts with a quantum system $\mathcal{S}$ which is in one of the n eigenstates $\{|s_i\rangle\}$ of the observable $\hat{S}$ of interest and a measurement apparatus $\mathcal{A}$ which is in the so-called "ready-state" $|a_0\rangle$, one of the $n+1$ eigenstates (labelled $|a_i\rangle$) of the pointer observable $\hat{A}$. (I assume that the pointer states are non-degenerate.) The eigenstates of the pointer observable correspond to the macroscopically distinct (directly observable) states of the measurement apparatus. (The position of a pointer, for instance.) The measurement interaction (or experimental setup) is designed in such a way that the pointer states become correlated with the state $|s_k\rangle$ of the quantum system during the measurement interaction:

$$|s_k\rangle|a_0\rangle \longrightarrow |s_k\rangle|a_k\rangle \tag{1.1}$$

This may simply be viewed as the quantum mechanical translation of the classical measurement process: properties of the system are correlated to certain properties of the measurement apparatus which 'amplifies' them into the observable domain. In general however, the quantum system is not in an eigenstate of $\hat{S}$, but in a superposition. Due to the linearity of the Schrödinger evolution (which describes the measurement interaction) one has in this case

$$\left(\sum_i c_i|s_i\rangle\right)|a_0\rangle \longrightarrow \sum_i c_i|s_i\rangle|a_i\rangle \tag{1.2}$$

Now the problem is what to do with the entangled superposition of the system and apparatus states on the right-hand side. It certainly does no longer conform to the classical intuitions one has about measurement apparatuses. Each $|a_i\rangle$ represents

---

[11] In many-minds interpretations it is not the world, but the mind of the observer (who becomes aware of a certain outcome) that splits.

[12] As far as I can tell, all interpretations agree on this part (except perhaps for Bohmian mechanics, which is quite a different story). Other approaches that invoke a physical argument (such as GRW or decoherence, see below) also agree that this is the problem, but their arguments essentially boil down to the claim that the problem derives from a wrong set of premises: for instance GRW assume that all physical processes, including measurements, are governed not by the Schrödinger equation but by a modified version thereof. However, one must keep in mind that the von Neumann scheme is a simplification that does not reflect the more sophisticated measurement theory in terms of positive operator-valued measurements (POVM's) used by physicists. To keep the discussion as simple as possible I will not consider these complications (which are mostly irrelevant to the topics discussed here). But see Wallace (2007) for a review of the measurement problem from a more general point of view than the one presented here.

a value of the pointer, but what we get in quantum measurements is not some kind of "simultaneous superposition of pointer values". In fact, there just doesn't seem to be a straightforward way to make sense of the superposition (1.2). I will call this the *minimalist measurement problem* (mmp).

> **The minimalist measurement problem (mmp):** How to make sense of the expression (1.2)?

The inability to straightforwardly answer this question is what I meant at the start of this section by saying that the bare formalism (without projection postulate) has a 'nonsensical consequence'[13]. I call it minimalist because it is (purposely) formulated in a rather obscure way to keep it as general as possible. Phrased as such, it is in fact the measurement problem as it is most frequent encountered in the literature. The point is that everyone seems to disagree about what exactly is the problem. In what follows I will attempt to shed some light on the various stances one can take with respect to the mmp.

Interpretations of the wave function

There is a philosophical side to the mmp that I would formulate as follows:

> **Interpretation:** Can one formulate a suitable interpretation of the state vector that attaches a clear and consistent (epistemic or ontological) meaning to (1.2)?

There is a large class of approaches that do not consider the problem of Interpretation to be essential to the measurement problem. These approaches can be characterized as "physical" since they claim that the solution to the mmp lies in the physics of the measurement interaction, rather than in the interpretation of the quantum state. For these approaches, the fact that we do not understand what the superposition principle means, is not considered to be problematic. The problem with (1.2) is that the macroscopic, directly observable *pointer* is in such an incomprehensible superposition state after the measurement interaction. So the question they are concerned with is how to avoid that quantum "indefiniteness" penetrates into the macroscopic domain, i.e. finding a way to model the measurement pro-

---

[13] The mmp is of course the formal counterpart of the famous Cat paradox of Schrödinger, who noted that

> "The ψ function of the total system would yield an expression for all this in which the living and the dead cat are (pardon the expression) blended or smeared out in equal measure. The characteristic of these examples is that an indefiniteness originally limited to atomic dimensions gets transformed into gross macroscopic indefiniteness [...] This prevents us from continuing naively to give credence to a "fuzzy model" as a picture of reality"(Schrödinger (1935), quoted in Fine, 1996, p. 65).

These last two sentences are essential, because they indicate that Schrödinger primarily intended to argue that one cannot accept a "fuzzy" picture of the microworld on grounds of the latter's unobservability, because one can always come up with some causal chain of events (a measurement!) that transfers this "fuzzyness" to everyday reality. The purpose of the "physics-oriented" approaches that I will discuss below is precisely to reinstate such a "fuzzy" picture, by arguing that the chain will break somewhere between the micro- and the macroscopic. See also footnote 32 for a brief discussion of the conceptual relevance of the paradox from a different perspective.

cess (perhaps by modifying the laws of quantum mechanics itself) such that this awkward state of affairs is avoided[14].

It must be noted that the philosophical point of view cannot be completely separated from the "physical" approaches. The latter often require additional interpretational assumptions to complete their argument, for example their tacitly accepting the e-e link. In particular, they will have to spell out what kind of entities the wave function refers to. Although this distinction is sometimes neglected, I believe that one has to distinguish at least two possible standard approaches[15]:

1. The quantum state describes an ensemble of identically prepared physical systems. (I will call this kind of interpretations "ensemble interpretations", in accordance with common terminology.)
2. The quantum state describes an individual physical system. (By the lack of standard terminology I will call this "individual interpretations".)

The ensemble interpretations have a long history that starts with Einstein's critique of the quantum theory[16]. They are motivated by the fact that the quantum state yields only statistical information, i.e. it is insufficient to characterize the behavior of an individual quantum system (which *cannot* itself be characterised in terms of quantum mechanics). From the point of view of an ensemble interpretation, the state *can* be used to extract probabilistic information about an individual quantum system, but only in so far as it is part of a (fictious) ensemble. The crucial point is that the quantum state itself is not *assigned* to a single element of the ensemble, as being an *intrinsic* physical description of an individual system, but only to the ensemble as a whole. This distinction will turn out to be essential when discussing the decoherence approach to the measurement problem.

Adherents of an ensemble interpretation generally make no attempts to extend the quantum theoretical framework to account for individual quantum systems, or single events. Approaches that aim to complete the ensemble interpretation with such an underlying theory are known as "hidden variable theories". As is now well-known, any hidden variable account will have to satisfy serious constraints[17]. These are characterized by the no-go theorems of Kochen-Specker and Bell. The idea of the Kochen- Specker theorem is the following: consider a hidden-variable theory in which every quantum-mechanical observable always has a definite value.

---

[14] See also section 1.2.2 in chapter 3 and Landsman (2007). That this change of perspective is not appreciated by everyone, is clear from Leggett's (2002) criticism that the decoherence theorist commits the fallacy of thinking to undo a crime by taking the evidence away. Clearly, tempting as this kind of criticism is, Leggett has a different kind of measurement problem in mind. (This kind of criticism *is* highly relevant to the ignorance interpretation (cf. section 1.2.2), but that does not seem to be Leggett's target.)

[15] "Non-standard" being for instance Bohmian mechanics, in which the wave function only has a dynamical role, see below.

[16] See Fine (1996) for a clear discussion of this history and a philosophical analysis of Einstein's ideas.

[17] I will not go into all the subtle details of the subject. There is a huge amount of literature devoted to the interpretation of the various formulations of the Bell-inequalities, the question whether the no-go-theorems really refute all local hidden-variable theories, the exact meaning of the locality constraint, its relation to the assumption of non-contextually, the connection between the Kochen-Specker and Bell theorems, and the plausibility of the experimental tests of the latter. For an introduction to these issues (with the exception of the latter) and references, see for instance (Redhead (1987); Mermin (1993); Fine (1996); Bub (1997) and Dickson (2007).

The "real" (i.e. hidden) physical state is then supposed to assign a value to all the (non-commuting) observables, defined as orthogonal bases of a given Hilbert space, simultaneously. That is, a state is characterized by such a value distribution in which one vector in each basis in the Hilbert space is assigned the value 1 and the other vectors receive value 0 (i.e. the system has a well-defined value corresponding with exactly one projector in the spectrum of each observable defined on this Hilbert space). Kochen and Specker now show that such a value assignment is impossible. (To be precise: they construct a set of 117 tri-orthogonal bases for which no consistent value assignment exists.)

This does not mean that the Kochen-Specker theorem prohibits all hidden-variable theories. But it does disqualify an important class of what are known as *non-contextual* hidden-variable theories. Non-contextually is the assumption, crucial in the proof of Kochen and Specker, that the value assigned to a particular vector is fixed and therefore independent of what basis set it belongs to. This means in particular that it is assumed that the values for non-commuting observables (such as spins in non-orthogonal directions) are simultaneously fixed by the hidden state, *even though they cannot be measured by the same experimental arrangement.* (A different experimental arrangement being an external factor on which the hidden state may depend.)

In the Bell theorems the general assumption of non-contextually is weakened by replacing it by *locality*; briefly put, it amounts to the idea that the measurement outcomes of one particle of an entangled pair should not depend on the *choice* of what to measure on the other particle. (This is a special case of non-contextually, with the pair of non-commuting observables $\hat{A} \otimes \hat{B}$, $\hat{A} \otimes \hat{B}'$, with $\hat{A}$ the observable measured on particle 1, and $\hat{B}$, $\hat{B}'$ two non-commuting observables (say, different spin directions) on particle 2. The value assigned to $\hat{A}$ for particle 1, prior to measurement, should not depend on which observable for the second particle one chooses.) Bell derives an inequality for the expectation values of these observables for such a local hidden-variable model, and shows that these are (for particular states) incompatible with the statistical predictions of quantum mechanics[18].

The Kochen-Specker and Bell theorems thus show that a local hidden-variable model cannot reproduce the (empirically verified) statistical predictions of quantum mechanics. The best known hidden variables account of quantum mechanics is therefore manifestly non-local. Bohmian mechanics has been studied in great detail as a viable alternative to the standard formulation of quantum mechanics, with the advantages that it is a deterministic theory and that it is supported by an intuitively attractive ontology of particles. It describes the motion of particles governed by a classical equation of motion, supplemented by a non-classical potential energy term called the quantum potential. It is the latter which introduces non-locality: for a several-particles system it contains interaction terms that cause an immediate and non-negligible action at a distance. In spite of its non-local nature, in some sense Bohmian mechanics is the most straightforward interpretation of quantum mechanics one can give; it is based upon a simple rewriting of the Schrödinger equation and as a result, it is empirically equivalent to the standard formulation as far as position measurements are concerned. The non-local features of the theory

---

[18] An important difference with the state-independent Kochen-Specker theorem is that the Bell-inequalities are violated only for specific states and a specific choice of spin-directions. As Mermin (1993) argues, this is a consequence of replacing non-contextually by the weaker assumption of locality.

however imply that it is not Lorentz-invariant, posing severe difficulties for an extension of Bohmian mechanics into a relativistic setting. Another concern is that the empirical equivalence of Bohmian mechanics to standard quantum mechanics is secured by a peculiar assumption about particle distributions which lacks proper motivation. Namely, in order for the particle statistics to satisfy the Born rule, the equilibrium distribution of the particles has to equal the square of the absolute value of the wave function, even though in Bohmian mechanics the wave function acts as a 'guiding wave' that governs the dynamics of the particles and is conceptually unrelated to the particle distribution itself.

Despite the formal and ontological clarity of the Bohmian approach, its non-local character departs radically from our classical intuitions. The Bell and Kochen-Specker theorems seem to imply that this is inevitable for any hidden-variable theory. This does not mean that there are no problems left if one simply dismisses any attempt to give a deterministic formulation of quantum mechanics. In fact, a statistical version of the Bell theorems also applies, which means that it is not possible to associate a definite (local) probability distribution over a pre-determined set of measurement outcomes with the quantum state. Thus, even if one rejects the idea of determinism, the phenomena still demand an non-causal description[19]. This has conceptual implications that will also have to be addressed by the adherents of an ensemble interpretation. Before discussing those, I will first turn to the first class of interpretations mentioned above, i.e. the individual interpretations.

The collapse of the wave function

Although physicists often adhere to an ensemble interpretation for pragmatic reasons, it is not entirely in line with the von Neumann postulates. In fact, the collapse postulate presumes that state of the *individual* systems (perhaps within an ensemble) is modified when a measurement has been made. This is a consequence of adopting the e-e link (although the latter is also an essential assumption for most ensemble interpretations, as I will discuss below). For in that case one cannot be satisfied with the superposition (1.2) at the end of the measurement process: if upon measurement one finds the eigenvalue a_n, the e-e link states that the $\mathcal{SA}$ combination should be in an (eigen-) state $|s_n\rangle|a_n\rangle$. So we need an additional step in the measurement process to accomplish:

$$\sum_i^n c_i|s_i\rangle|a_0\rangle \xrightarrow{\text{``}collapse\text{''}} |s_1\rangle|a_1\rangle \; or \; |s_2\rangle|a_2\rangle \; or \; \ldots |s_n\rangle|a_n\rangle, \tag{1.3}$$

in which the "or" can be interpreted in various ways. It does not necessarily mean that these events are mutually exclusive. (This may sound a bit obscure, but this is basically the stance of relative-state interpretations.) However, even relative-state interpretations aim to answer the central question (which is essentially a more precise formulation of the mmp in case of a individual interpretation of the wave function):

> **Outcomes:** Why does one perceive the expression on the right-hand side of (1.2) as a single occurrence at a time of the possible measurement outcomes represented by the $|a_i\rangle$ states?

---

[19] the idea of indeterministic causality is captured in the concept of a *Common Cause*, see (van Fraassen, 1991) for details.

However, if one accepts the e-e link *and* insists on a one-to-one correspondence between the formalism and our experiences, then one needs to explain what causes the *real, objective and physical* transition (1.3). This is what is generally called the "collapse" or "reduction of the wavepacket":

> **Collapse:** What kind of process causes the jump (1.3) of the superposition to one of the (system-) apparatus' eigenstates of the pointer observable?

Now a first response could be to seek the answer to this question in a detailed physical investigation of the measurement process. However, the mmp arises precisely as a result of our attempt to grasp the measurement process in plain physical terms: eventually all physical processes are governed by the Schrödinger equation, but its linearity leads to sheer contradiction with our experience, where superpositions of "ordinary" objects such as pointers are simply absent. As I will discuss below there is hardly a way to avoid this conclusion, despite the fact that the oversimplified form of the von Neumann measurement scheme seems to call for a more detailed argument. (That is, even if we demand a weaker form of collapse in the framework of the ensemble interpretation, cf. the discussion below.)

von Neumann himself did not make any attempts in this direction. Instead, he approached the problem from quite a different perspective: if the collapse of the wave function cannot be accounted for by the known laws of physics, one could introduce it as some kind of additional evolution in the formalism, which only acts during measurements and plays no role whatsoever in other physical processes. Although he was quite well aware of the fact that measurements are nothing but ordinary physical interactions, von Neumann did not have a problem with the 'extra-physical' character of what he called the "first intervention" (the second being the Schrödinger evolution) which governs the "discontinuous, non-causal and instantaneously acting experiments or measurements" (von Neumann, 1955, p. 349). In the end, all measurements amount to perceptions, and for von Neumann, the crucial point was that actual perceptions fall outside the domain of physics proper. Nevertheless, according to what von Neumann calls the principle of "psycho-physical parallelism", our perceptions ought to have an objective counterpart in physical reality. This is basically von Neumann's motivation for introducing "process 1": the collapse postulate. To justify the fact that this process violates the reversible Schrödinger evolution, von Neumann appeals to a strong mind-body dualism: the collapse of the wave function appears where the measurement process crosses the boundary between objective reality and subjective experience; and thus, its ultimate cause is not open to physical investigation. Furthermore, von Neumann argues that it is merely a matter of convention where the line between "subjective" and "objective", the Heisenberg cut, is drawn. He argues that what is required is merely a kind of consistency, meaning that it should be immaterial whether one applies process 1 to the system alone or to the system and apparatus. He shows that this, indeed, is the case, but it is a matter of principle that the line must be drawn *somewhere*.

So what the argument of von Neumann seems to boil down to, in the end, is a reply to the criticism from the start of this section: that the axiomatic structure of the theory should not contain any terms that are themselves to be described by the theory. Surely, that does not prohibit one from introducing extra elements into one's ontology that fall outside the scope of the theory. For von Neumann, the notion of consciousness fitted perfectly well in this ontology – not withstanding other objections one may have against such a radical move. Whereas the usual non-realist orthodox interpretation simply dismisses all talk of an unobserved reality as 'meaningless', von Neumann's "first intervention" even seems to imply that it is the act of observation itself that *creates* the fact being observed[20].

Clearly, this introduction of "consciousness" into the theory as an active intervention, inducing a *physical* effect on the quantum state, is unacceptable for most physicists (with the exception of a few explicit advocates such as von Neumann himself, and also Wigner and London & Bauer). Therefore, various attempts have been made to account for the collapse of the wave function in a different way. An obvious strategy is to reinstate collapse as a *physical* principle and to formulate it in explicit mathematical terms. Such a formulation should not simply add a collapsing mechanism "at the point of measurement" or "observation". Rather, one seeks a modification of the dynamical law that is generally applicable and accounts for both the 'ordinary' Schrödinger evolution and the collapse which takes over in the amplification to the macroscopic domain. This, basically, is the approach of Ghirardhi, Rimini and Weber (GRW).[21] They proposed to modify the Schrödinger equation with an additional stochastic term that causes 'hits' or indeterministic 'jumps' of the quantum state into an (approximate) eigenstate of position with an extremely low (average) frequency. However, this frequency increases as the number of particles involved increases, with the consequence that the average time for superpositions (of position) to persist becomes extremely short for macroscopic systems. Hence the appearance of collapse in the measurement process, when the quantum system is coupled to the macroscopic apparatus.

With a suitable choice of parameters, GRW theory yields good agreement with both the predictions of quantum mechanics and collapse where it is to be expected[22]. One obtains a purely physical formalism without puzzling superpositions of pointer states or an obscure exceptional status for measurements, however at the cost of sacrificing determinism and the Schrödinger evolution as the fundamental dynamical law. Essentially, we are no longer talking of quantum mechanics but of a different theory.

Quantum statistics

GRW theory is an attempt to reconcile the probabilistic nature of quantum measurements with a non-statistical interpretation of the wave function, and so probabilities are introduced at the level of the dynamical law. However, if one adopts

---

[20] A view that Bohr, by the way, was strongly opposed to (Bub, 1997), in contrast to Heisenberg who stressed this point in his 1927 article on the uncertainty relations.

[21] And their successors. GRW theory is perhaps the best known, and most successful, but not the only stochastic collapse theory. A very extensive overview and detailed classification can be found in the chapter on stochastic collapse models by Stamatescu of the book on decoherence by Joos et al. (2003).

[22] This is not to say that GRW theory has no problems to cope with. Notably, no suitable relativistic extension of the theory has been formulated yet, and, on a more philosophical note, there is a problem with the interpretation of the infinite "tails" of the collapsed wave function since these are not compatible with the strict e-e link. (The latter problem will also concern us when discussing the interpretational implications of decoherence in chapter 3.) For a clear discussion of the "tails" problem and other problems within GRW theory, I refer to Wallace (2007).

an ensemble interpretation of the quantum state instead, the need for an explicit collapse (1.3) evaporates. This results in quite a different view on the measurement problem.

According to the ensemble interpretation, the quantum state describes an ensemble of measurement outcomes, distributed according to the Born rule, which, according to Gleason's theorem, conforms to the unique type of probability measure on Hilbert space. Although it need not be explicitly denied by an adherent of the ensemble interpretation that these measurement outcomes should be accounted for by some hitherto unknown theoretical picture, he dismisses the need to address this question for his interpretation of the quantum formalism. However, the ensemble interpretation obviously needs to address the question of *what* kind of measurement outcomes the ensemble consists[23]: the quantum state on the right-hand side of (1.2) together with the Born rule do not yield a probability distribution over a set of well-defined measurement outcomes until a particular observable (orthogonal set of eigenstates) has been specified.

There are two aspects of this discussion that must be distinguished. The first is that the quantum state itself does not specify what the set of orthogonal projectors is (i.e. what the events are) that the formalism assigns probabilities to. The essence of the orthodox formulation is that the quantum state obtains empirical content by 1) specifying a particular observable (call this *qualitative*), and 2) applying the Born rule to the quantum state with respect to this observable (which is *quantitative*). What this procedure gives us is empirical meaning in terms of probabilities for the occurrence of certain measurement outcomes. Henceforth, I will call the basis that establishes the qualitative empirical content the *interpretation basis*.

The crucial point is that the observer's choice of an interpretation basis must *precede* such a minimalist link of the quantum state to observable phenomena (i.e. probabilities for measurement outcomes), because it is not possible to assign a probability distribution to all observables simultaneously[24]. This means that it is impossible to tell from the quantum state alone even what events might occur with what probability; the observer's choice of what to measure (or to calculate) must be made before the state acquires any empirical content (see also footnote 19).

The heart of the problem is thus that *a choice of basis is not physical*; a different choice of basis is nothing but a different mathematical representation of one and the same physical object. Troubling about orthodox quantum mechanics, is that this choice of representation, although arbitrary, is used to endow the abstract wave vector with its empirical content. In other words: the observables in general do not

---

[23] I do not require that one would have to re-derive the Born rule, but opinions are divided about this issue. Because of the uniqueness of the probability measure implied by Gleason's theorem I will accept the Born rule unconditionally. In cases where the Born rule loses its original interpretative context – as in relative-state interpretations, see chapter 4 – it has to be derived from a specific set of assumptions to endow these probabilities with an interpretation. In that respect, a derivation would be useful for other approaches as well – such as Zurek's (2003, 2007) attempt to derive the Born rule from the notion of 'envariance' within the decoherence programme – but it is not *necessary* if one is willing to accept that the Born rule describes the statistics of measurement outcomes. (I will come back to this in the Outlook.)

[24] Conversely, if a particular preferred observable (fixing a particular decomposition of the quantum state) is specified, the Kochen-Specker theorem that prohibits the assignment of definite values to the eigenprojectors of this preferred observable, is no longer valid. See especially Bub's account of the modal interpretation (Bub, 1997) for the mathematical an philosophical elaboration of this idea.

reflect "bare facts", but become empirically relevant in a particular experimental arrangement only. The formalism, however, is indifferent to these experimental circumstances: an observable presupposes a certain empirical question, but the physical arrangements associated with this question still do show up in the calculations. This is precisely why, in my opinion, the orthodox interpretation is arbitrary and *ad hoc*, even without the collapse postulate. Indeed, the greater merit of the Copenhagen Interpretation over what I call the orthodox view is that it links the choice of observable to a measurement context, at least conceptually. I will call this the general version of the preferred basis problem:

> **Preferred basis (general):** For a given quantum state, what determines the orthogonal set of projectors to which the Born rule assigns the probabilities?

The second aspect is that the state on the right-hand side of (1.2) does not represent a "classical" probability distribution over the events specified by an arbitrary choice of observable. To see this, write down the density operator of the superposition on the right-hand side of (1.2):

$$\rho = \sum_{ij} c_i c_j^* |s_i\rangle\langle s_j| \otimes |a_i\rangle\langle a_j|, \tag{1.4}$$

(To alleviate notation, in what follows I will consider the simplified expression $\rho = \sum_{ij} c_i c_j^* |a_i\rangle\langle a_j|$, i.e. disregard the system-apparatus split.) Now, the coefficients squared $|c_i|^2$, i.e. probabilities, on the diagonal add up to one. However, the density matrix also assigns positive probabilities to projections not belonging to this set. This is problematic, for the off-diagonal terms (interference terms) prohibit a classical interpretation of those probabilities. For instance, if we compute the probability $\mathbb{P}_\rho(\hat{P}_{|a_k\rangle+|a_l\rangle}) = |c_k|^2 + |c_l|^2 + 2Re(c_l c_k^*)$, we find it contains an extra term $2Re(c_l c_k^*)$ compared to $\mathbb{P}_\rho(\hat{P}_{|a_k\rangle}) + \mathbb{P}_\rho(\hat{P}_{|a_l\rangle})$. In other words, the probability that the system is found in the state $|a_k\rangle + |a_l\rangle$ is not equal to the sum of the probabilities corresponding to each state separately, due to the interference terms. It therefore seems that one cannot identify a state like $|a_k\rangle + |a_l\rangle$ with "either $|a_k\rangle$ or $|a_l\rangle$". (Real physical arrangements in which these kind of expressions appear would invite such an interpretation. In the context of the double-slit experiment, for instance, the state $|a_k\rangle + |a_l\rangle$ is associated with the situation that both slits (corresponding with $|a_k\rangle$ and $|a_l\rangle$, respectively) are open. But the probability distribution for particles passing through the two slits differs from the probability distribution resulting when each particle passes through either one slit or the other.)

Thus, if the interference terms are removed, the probability distribution for the states on the diagonal would equal that of a "classical" probability distribution, i.e. the ensemble behaves statistically as if it consists of systems each being in one of the states on the diagonal. Returning to the mmp, the point is that the interference terms

$$\rho - \tilde{\rho} = \sum_{i \neq j} c_i c_j^* |s_i\rangle\langle s_j| \otimes |a_i\rangle\langle a_j| \tag{1.5}$$

should vanish. This is a minimal requirement for interpreting the result

$$\tilde{\rho} = \sum_{i} c_i c_i^* |s_i\rangle\langle s_i| \otimes |a_i\rangle\langle a_i| \tag{1.6}$$

as a probability distribution of events (in this case represented by the correlated stats $|s_i\rangle|a_i\rangle$) occurring with certain probabilities $|c_i|^2$. So another question that an ensemble interpretation needs to address is this:

> **Interference:** What causes (1.6) rather than (1.4) to obtain at the end of the measurement process?

From now on, I will call all approaches that aim to answer the question of Interference "decoherence" approaches. (Of which "environment-induced decoherence", which seeks the solution to this problem in the interaction of the systems with its environment, will be the subject of the following chapters.) The goal of decoherence in the light of an ensemble interpretation is to arrive at a probability distribution of measurement outcomes, i.e. a density matrix diagonal in the basis of a certain preferred observable.

I would like to point out that there are two ways to interpret the meaning of such a diagonal state; in a quantum mechanical sense and in a classical sense. By the first I mean the formal equivalence of such a state with a *quantum* mechanical mixture of the states on the diagonal. The statistics of such an ensemble (in terms of *all* the possible observables) are of course still essentially quantum, i.e. they prohibit an interpretation in terms of simultaneously possessed properties (recall my discussion of the Kochen-Specker theorem). On the other hand, if one considers only a *single* preferred observable, namely the observable defined by the (orthogonal) states on the diagonal, the probabilities are classical, and thus there is no contradiction if one were to interpret the ensemble as consisting of systems actually *possessing* these values.

Note the subtle relation between the questions of the Preferred Basis and Interference above. The question of Interference presupposes that a privileged set of orthogonal states (the eigenstates of a certain observable) has been specified that determines in what basis the density matrix is to become diagonal, i.e. that an answer to the first question is available[25]. (In my presentation of the problem, I assumed a certain preferred basis consisting of system and pointer states $|s_i\rangle, |a_i\rangle$, but this decomposition of the state (1.2) is of course just a matter of representation.) More generally, one might say that the preferred-basis problem is the most fundamental of all versions of the measurement problem presented here, in the sense that each of those is predicated on a possible solution to the preferred-basis problem; it makes no sense to ask about outcomes of measurements as long as it is has not been decided what exactly is being measured.

On the other hand, an answer to Interference is generally expected to come from the physics of the measurement interaction, and is therefore independent of what kind of observable the physicist happens to have in mind. Therefore, if such a physical approach would be successful, an answer to Interference could (but need not) imply an answer to (the statistical version of) the preferred basis problem. However, one might also argue that the interference terms in the density matrix

---

[25] Of course, the density matrix is Hermitian and therefore defines an orthogonal set of vectors through its own eigenbasis. The Born rule naturally assigns these diagonal projectors their coefficients as probabilities. This would seem to address both questions above at once. It is thus tempting to associate a probability distribution with the density matrix itself, but the point is that this eigenbasis will in general not correspond to the set of events one would like to assign probabilities to.

are irrelevant when there is an absolute preferred observable, since the interference terms are eigenstates of a *different* observable. (Note that in our argument above, the projectors of the form $\hat{P}_{|a_k\rangle+|a_l\rangle}$ do not project on an element of the basis $\{|a_i\rangle\}$.) There are more subtleties involved, in particular with respect to decoherence, which I will discuss in chapter 3 and again in the conclusion.

The measurement process

Before I turn to the question of Interference in more detail I discuss some implications of the specific features of measurement interactions for the preferred basis problem. To see what these implications are, let us return to the state vector representation since this is how the preferred basis problem is usually presented in the literature (Zurek, 1981; Bub, 1997; Schlosshauer, 2004).

In its most general form, the preferred basis problem says that the pure quantum state is just a unit vector in Hilbert space that does not carry any empirical content unless a particular observable has been specified. However, when dealing with measurements, this already imposes some minimal structure on the state; to be interpreted as an (ideal) measurement, the quantum state $|\Psi_{SA}\rangle$ should be represented on a product space $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{S}$ of the Hilbert space associated with the measured system and the measuring apparatus, respectively, and decomposed as a single sum, i.e. with perfect correlations between apparatus- and system states, such as in (1.2). Call this a "bi-decomposition". However, even disregarding the point that there may not be such a preferred factorization of the overall Hilbert space[26], the problem remains that such a representation is by no means unique: for any choice of (orthogonal) apparatus states $|a_i\rangle$, a set of relative system states $|s_i^{rel}\rangle$ can be defined such that $|\Psi_{SA}\rangle = \sum c_i|s_i^{rel}\rangle|a_i\rangle$. (This is Everett's "relative state" formalism. See chapter 4 for a more elaborate discussion.) Any term in the superposition represents a correlation of a certain system state $|s_i\rangle$ with the apparatus state $|a_i\rangle$, yet there is an infinity of ways to write down an arbitrary state $|\Psi_{SA}\rangle$ as such a superposition, each (seemingly) representing another kind of measurement.

One may argue that a genuine measurement imposes some additional constraints on the decomposition of $|\Psi_{SA}\rangle$. After all, in general a choice of orthogonal apparatus states will not define an orthogonal set of systems states, i.e. an eigenbasis of an hermitian observable. If it is insisted that the $|a_i\rangle$ as well as the $|s_i\rangle$ are orthogonal, such a "bi-orthogonal" decomposition of $|\Psi_{SA}\rangle$ will in general be unique, and no preferred basis problem arises. However, if two or more of the coefficients c_i in the expansion $|\Psi_{SA}\rangle = \sum c_i|s_i\rangle|a_i\rangle$ are equal, uniqueness no longer holds. Therefore the requirement of orthogonality is in this case not sufficient to overcome the preferred basis problem.

I briefly illustrate this idea (see also (Zurek, 1981)) by means of the EPR state (where single arrows denote (orthogonal) states of the system, and double arrows those of the apparatus):

$$|\Psi\rangle = \frac{1}{\sqrt{2}}\left(|\uparrow\rangle|\Uparrow\rangle + |\downarrow\rangle|\Downarrow\rangle\right). \tag{1.7}$$

---

[26] This assertion usually seems to be motivated by ontological considerations, but is by no means obvious. See the discussion in section 3.3.4 in chapter 3.

This decomposition happens to be bi-orthogonal as well as degenerate. It expresses a correlation between the two orthogonal states of the 'to-be-measured' spin observable $\hat{s} = \mid\uparrow\rangle\langle\uparrow\mid - \mid\downarrow\rangle\langle\downarrow\mid$ with the eigenstates of the pointer spin $\hat{S} = \mid\Uparrow\rangle\langle\Uparrow\mid - \mid\Downarrow\rangle\langle\Downarrow\mid$. However, one can also express this correlated state as

$$|Psi\rangle = \frac{1}{\sqrt{2}}\left(\mid\rightarrow\rangle\mid\Rightarrow\rangle + \mid\leftarrow\rangle\mid\Leftarrow\rangle\right) \tag{1.8}$$

with $\mid\rightarrow\rangle$, $\mid\leftarrow\rangle$ a different orthonormal basis of the quantum system defined by

$$\begin{aligned} \mid\rightarrow\rangle &= \frac{1}{\sqrt{2}}\left(\mid\uparrow\rangle + \mid\downarrow\rangle\right), && (1.9) \\ \mid\leftarrow\rangle &= \frac{1}{\sqrt{2}}\left(\mid\uparrow\rangle - \mid\downarrow\rangle\right) && (1.10) \end{aligned}$$

and similarly for the pointer states. Note that there are *an infinite number of equivalent ways* to write down such a decomposition. Therefore, one cannot conclude that what was measured by the apparatus was the observable $\hat{S}$. We seem forced to believe that the apparatus has measured *both* (non-commuting) observables $\hat{S}$ and $\hat{S}' = \mid\Rightarrow\rangle\langle\Rightarrow\mid - \mid\Leftarrow\rangle\langle\Leftarrow\mid$ *at the same time.*

The above argument shows that even the additional constraints on the decomposition of the $\mathcal{SA}$ state imposed by the measurement interaction, are not sufficient to uniquely fix the system and pointer states that determine the physical interpretation of this measurement interaction. Even if one insists on bi-orthogonality this will still not suffice in the case of degeneracy of the coefficients[27]. It must be noted that although the non-degenerate case is typical in the mathematical sense, this does not imply it is also *physically* common; in infinite-dimensional Hilbert spaces degenerate states are in fact generic (see for instance (Bacciagaluppi, 2000)).

Nonetheless, the measurement interaction does remove some of the generality of the "statistical" version of the preferred basis problem. One can formulate this stronger version as follows:

> **Preferred basis (decomposition):** Of all the possible bi-decompositions of the state $|\Psi_{SA}\rangle$ on a particular fixed product space $\mathcal{H}_S \otimes \mathcal{H}_A$, what determines which one is the "right" one, corresponding to the system and apparatus states being measured?

In chapter 2 we will see how the theory of environment-induced decoherence claims to solve this issue.

The other question that is central to the ensemble interpretation is whether a more complete description of the measurement process can be given that accomplishes the transition from (1.4) to (1.6). Unfortunately, the answer is no. If one models the measurement process by a unitary evolution operator $\hat{U}$ on the combined system-apparatus Hilbert space $\mathcal{H}_{\mathcal{SA}}$ it is easy to see why: the $\mathcal{SA}$ combination is in a pure state at the start of the measurement process (1.2). The diagonal mixture (1.6) is not a pure state. And unitary evolution will never turn a pure state into a mixture.

---

[27] This problem is not confined to the ensemble interpretation of course: in relative-state as well as modal interpretations the preferred basis problem is a major issue. Hidden-variable theories obviously do not need to address the preferred basis problem: in a deterministic universe what is observed is ultimately determined by what *is* at the fundamental ontological level.

This argument is a bit a too simple, however. There are, in fact, numerous objections one could make against the simplified measurement scheme (1.2). For instance, one need not assume that the apparatus starts out in a pure state, or that the pointer states are exactly orthogonal, or that the measurement is perfect (leaving the system states unperturbed.) However, the many 'insolubility proofs' available in the literature, each starting from a much more general and slightly different set of assumptions, all conclude that these assumptions are incompatible with one another[28]. Pessoa (1998) lists these assumptions as follows:

1. The apparatus can be described as a quantum system (either pure or mixed).
2. The $\mathcal{SA}$-system evolves according to the Schrödinger equation.
3. Measurement outcomes correspond to distinct (i.e. orthogonal or nearly orthogonal) pointer states.
4. The measurement interaction is a member of a specified class of measurements (first or second kind, PVM or POVM.)
5. The final $\mathcal{SA}$-density matrix is a mixture that is diagonal in the pointer basis.

The insolubility proofs now show that the conditions 1,2,3,4 and 5 cannot be simultaneously fulfilled. If one wants to arrive at the diagonal mixture (1.6), that is, maintain condition 5, he will have to drop one of the other assumptions. Now the only reasonable candidate seem to be the unitarity condition 2, since without the assumptions 3 and 4 one cannot speak of measurement interactions in a meaningful way. I already discussed the possibility of introducing an explicit collapse; either by von Neumann's 'first intervention' which seems to require an active role for human consciousness, or Ghirardhi, Rimini and Weber's spontaneous collapse theory, which modifies the Schrödinger dynamics at the fundamental level[29].

However, there is a third option, which might be called an 'implicit collapse': it was assumed that the $\mathcal{SA}$-system evolves according to the Schrödinger equation, since the $\mathcal{SA}$-system is closed. *Open* systems, however, do not evolve unitarily. An open system is a system interacting with an environment which is not included in the kinematics (the quantum state) but does affect the dynamics. Put differently: the *total* system including the environment evolves unitarily, but if we "restrict our attention" to a part of this combined system we will find that it does not obey the Schrödinger equation, but a so-called generalized master equation (see for instance (Joos et al., 2003)). That one can circumvent the insolubility proofs this way does of course not imply that the measurement problem has been solved. Nevertheless, that the inherent "openness" of macroscopic systems would explain the transition from (1.4) to (1.6) is the basic idea of environmental decoherence, which will be the central topic of the next chapter.

---

[28] The first insolubility proof was formulated by von Neumann (1955). A comprehensive overview is Brown (1986). A recent and very general proof (which aims at a more general notion of macroscopic distinguishability of the apparatus-states rather than the diagonality of the density matrix) is formulated by Bassi and Ghirardi (2000).

[29] But note that these approaches are concerned with a different kind of collapse that is supposed to reduce the superposition state to one of its terms, not to remove the interference terms. (In other words, it is probabilistic but not statistical.)

1.2 Decoherence and the emergence of definiteness

In the previous section, I have claimed that decoherence approaches (in particular environment-induced decoherence) receive their motivation from an ensemble interpretation of the quantum state. The central question these approaches aim to address is that of Interference, yet no such solution can be satisfactory if the (statistical version of) the Preferred Basis problem is not also addressed. Although the latter problem could possibly be addressed in terms of an answer to the first, in the older approaches to the measurement problem the issue has more often been neglected altogether (such as for instance Everett's original proposal for a relative-state interpretation that I discuss in chpter 4). As I will explain in chapter 2, environment-induced decoherence, especially the approach of Zurek (1981, 2003) is a notable exception, and in chapter 3 I will explain what the consequences are of the fact that in this case the problems of Interference and the Preferred Basis are addressed separately.

In this section, I will focus on the problem of Interference, since this has generally been understood by the general physics community to be the essential tenet of decoherence, and has been criticized the most in the more philosophically oriented literature. The kind of criticism presented here is in fact rather general and well-known (though I will try to give a somewhat personal spin to the argument) and independent of the particular means by which decoherence is obtained. For a critical and more detailed assessment of the methods of environment-induced decoherence in particular, I refer to chapter 3.

1.2.1 Interference vs. indefiniteness

Various authors (Bell, 1990; Maudlin, 1995; Bub, 1997; Pessoa, 1998) have objected that the measurement problem has not been solved at all when the interference terms have been removed. From their point of view, the problem with the interpretation of the superposition state in the orthodox framework does not lie in the statistical discrepancy due to the interference terms, but in the e-e link: once it has been specified what the preferred observable is, the e-e link does not assign any definite value of the associated physical quantity to a superposition of eigenstates of this preferred observable. The problem of Outcomes is the question of how to reconcile this indefiniteness with our definite and unique experiences, and this is what these critics of decoherence see as the measurement problem. Addressing the problem of Interference does in that case not seem to be very helpful. As Bell (1990) puts it: "The idea that elimination of coherence, in one way or another, implies the replacement of 'and' by 'or', is a very common one among solvers of the 'measurement problem'. It has always puzzled me." (p. 36)

At the heart of this criticism is the assumption that the wave function represents a *single* system, not a statistical ensemble of systems. This difference is essential to understand the nature of the dispute. The adherent of the ensemble interpretation does not need to address the problem of Outcomes (at least not in the form and context as it was put here), simply because he takes the occurrence of measurement outcomes more or less for granted. To understand the statistical distribution of these outcomes in a realistic vein may be tremendously difficult, but that is perhaps an issue to be left to physicists (or philosophers) with more imagination. In any case, for an ensemble interpretation the physical systems to which the individual measurement outcomes refer are *not* to be accounted for by the quantum state, and that is what is essential here.

Unfortunately, however, it seems that the essentially statistical justification of decoherence-like approaches (in the general sense of section 1.1.1) is not appreciated by most decoherence theorists. In fact, the prevailing idea seems to be that the vanishing of interference is somehow relevant at the level of the individual systems/facts/outcomes that constitute the ensemble[30]. This is indeed a rather awkward move, and it is in this light that we should read Bell's criticism.

So there are two aspects to this kind of criticism which I would like to distinguish:

- The first is that the question of Interference is motivated by a purely statistical argument that is completely silent about the properties of individual systems within the ensemble (i.e. an ensemble interpretation could never provide an answer to the question of Outcomes, simply because the occurrence of outcomes is taken for granted within such a statistical framework).
- The second is that *if* it is assumed that wave functions describe individual systems, the fact that the wave function decoheres does not imply that these systems acquire some definite property (represented by the states on the diagonal).

Both issues are addressed in one way or the other, but as I will now explain the standard responses are inadequate.

Especially in the older discussions of decoherence (Zurek, 1981, 1982, 1993) the first objection is avoided because the relevance of the question of Interference receives quite a different kind of motivation that is known as the "ignorance interpretation of mixtures" (IgI). The idea is that since a diagonal density matrix is formally (i.e. statistically) equivalent to a classical ensemble, it is also *ontologically* equivalent. This is supposed to link the statistics of the wave function to an interpretation at the level of the individual members of the ensemble. If such an ignorance interpretation would work (but as I will explain, it does not) it might also address the second point.

As I will try to argue below and more extensively in chapter 3, a large part of the conceptual criticisms of decoherence hinges on the conflation of these two conflicting interpretations of the wave function (statistical vs. individual). To make this somewhat more precise; there are two different claims that should be (but unfortunately are usually not) distinguished :

a. The *ensemble* described by the diagonal density matrix $\sum_i |c_i|^2 |s_i\rangle\langle s_i|$ *behaves statistically as* a set of measurement outcomes corresponding to the states $|s_i\rangle$, distributed according to the Born coefficients $|c_i|^2$.

b. The *quantum system* described by the diagonal density matrix $\sum_i |c_i|^2 |s_i\rangle\langle s_i|$ *has a definite property* corresponding to an eigenstate of the preferred observ-

---

[30] Zeh, who has always been committed to a realistic interpretation of the wave function (Camilleri, 2008), puts it emphatically: "One has to keep in mind this individuality of the superposition principle and its consequences for *individually* observable physical properties in order to appreciate the meaning of the program of decoherence" (Zeh, in Joos et al., 2003, p. 11). We will encounter more instances of this kind of thinking in chapter 3.

> able represented by the orthogonal states on the diagonal (i.e. the system has a definite value of the physical quantity represented by this observable).

These two accounts of decoherence cannot be straightforwardly identified because they correspond to different interpretations of the wave function. The claim that there is nevertheless a connection between the two is often illustrated by means of the double slit experiment: when the particles are sent through the two slits, the resulting pattern on the screen indicates interference and hence does not correspond to a distribution of particles having passed through either one slit or the other. On the other hand, when it has been determined that each particle has passed through a particular slit, the interference disappears. Thus, the fact that the path of the individual particles has become determinate is reflected in the vanishing of the non-diagonal terms from the density matrix that describes the ensemble.

But does this example provide sufficient support for the claim that decoherence (disappearance of interference terms) implies definiteness (claim *b.* above)? I think not; in fact, it can be easily shown to be unfounded (if not false), so let me explain. (In what follows, I will take the preferred observable to be position to make the connection with the double slit-experiment more prominent. By "localization" I mean that the individual particles in the ensemble have a well-defined position.)

In its strongest form, the claim is that *decoherence is a necessary precondition for definiteness*[31]. Only if this statement is true can one appeal to the disappearance of interference to *explain* that individual particles have a well-defined position. However, as Pessoa (1998) also points out, this is easily refuted: also in the case of interference do the individual particles in the ensemble leave a mark on the screen at a well-defined position. So one can have localization without decoherence.

Now I think that this argument should be treated with some care, because the advocate of environment-induced decoherence might want to use it for his own benefit. As will be explained in the next chapter, the theory of environment-induced decoherence predicts that interference terms (usually in the position basis) will vanish when the microscopic system interacts with an environment - which, in effect, can be any external (open) system. So, it may be argued that also when the particles hit the screen decoherence takes place and they become localized. However, note that this means that one has to *presuppose* that decoherence implies definiteness, in order to respond to the objection that definiteness is *not* a consequence of decoherence. This seems to be begging the question.

As an attempt to break out of this circle, we could take a closer look at what kind of claims the double-slit experiment actually provides empirical support for. They are the following:

> Path of individual particles is indeterminate → interference pattern present.

and

> Path of individual particles is determinate → no interference pattern present.

---

[31] For instance, Zeh remarks that "this loss of coherence is indeed required by mere logic once measurements are assumed to lead to definite results"(Joos et al., 2003, p.13).

The decoherence theorist, however, claims that

> no interference pattern present → each particle is localized

in order to use the disappearance of interference as an *explanation* of localization. But even if we make the dubious identification "determinate path" with "being localized" this does not follow from either of the two statements above. Because from the first we can conclude that

> no interference pattern present → $\neg$ (path of each individual particle is indeterminate.)

but the negation of "path of each individual particle is indeterminate" is not "path of each individual particle is determinate". This is precisely because the first statement is about the ensemble as a whole, whereas the second is about its individual members. (The negation of "every individual human being has two legs" is not "every individual human being has not two legs".) This kind of reasoning cannot bridge the gap between the two kinds of realities that are attributed to the wave function.

One could possibly go on like this, but I think I have made my point; the claim that the disappearance of interference in a certain eigenbasis is connected to the definiteness of the corresponding property of the individual members of the ensemble, *is not supported by any evidence.* Both its motivation and experimental support are statistical in nature, and therefore the addressing question of Interference as being relevant for the measurement problem, only makes sense within the framework of an ensemble interpretation. (This is not to say that the connection between decoherence and definiteness is not there –but one would have to come up with a better argument to prove it.)

1.2.2 The ignorance interpretation of mixtures

Especially in the older decoherence literature (e.g. Zurek (1982, 1993)) the claim that decoherence implies the existence of definite measurement outcomes is established in a somewhat more roundabout way. The idea is to turn the statistical argument into a stronger ontological claim; namely through the ignorance interpretation that purports to explain the statistics of the ensemble in terms of the properties of the individual systems that constitute it.

In his original discussion of the measurement process, von Neumann (1955) argued that after the measurement correlation between system and apparatus was established, some additional mechanism was needed to turn the superposition state into a statistical distribution of measurement outcomes. Therefore, he introduced his "first intervention" (the collapse of the wave function) to accomplish the transition of (1.4) to (1.6). This is just the formal part of the story, but *physically*, the collapse of the wave function causes the system-apparatus couple to end up in one of the states $|s_i\rangle|a_i\rangle$ at each measurement event. The mixed state (1.6) thus represents a genuine probability distribution, reflecting our ignorance about which of the states on the diagonal actually obtains. The collapse of the wave function thus *precedes* the representation as the diagonal mixed state; it explains why one can interpret these probabilities in terms of actually occurring random measurement outcomes. This is at least how von Neumann intended his argument.

Unfortunately, the argument has often been misinterpreted in the opposite direction; approaches (especially the older ones) that address the problem of Interference do so because it is believed the transition of (1.4) to (1.6) is the actual collapse and all that is desired, ignoring the specific motivation for this transition that von Neumann adopted. Alas, the argument works in one direction, but not the other. Although a genuine probability distribution of states $|s_i\rangle$ is properly represented by a mixture diagonal in these states, not any such state represents a genuine probability distribution. By "genuine probability distribution" I mean that the mixed state (1.6) represents an actual ensemble of states $\{|s_i\rangle\}$ of which a portion $|c_k|^2$ is in the state $|s_k\rangle$ (or equivalently, a single element taken at random from such an ensemble with the coefficient representing our ignorance about its actual state). The "ignorance interpretation of mixtures" purports to interpret *any* mixed state in this sense. As a consequence, according to this view pure states have some kind of ontological priority over mixed states (Ochs, 1981). Such an interpretation therefore wants to blame the probabilistic appearance of measurements on epistemic uncertainty: the ensemble not merely behaves statistically *as if* it were a classical mixture - it *is* such a mixture. But however appealing this claim may seem, it proves to be untenable.

This is particularly clear when one retraces the IgI back to its original context of the measurement problem. Let me spell out the full argument in some detail (I omit the system-apparatus split for notational convenience, that is, I refer to both with a single vector $|s_i\rangle$ and the phrase "system"):

1. The ensemble is in a state $|\Psi\rangle = \sum_i c_i|s_i\rangle$
2. *Problem*: the corresponding density matrix $\rho_{int} = \sum_{ij} c_i c_j^*|s_i\rangle\langle s_j|$ does not represent what is observed: an ensemble of outcomes $|s_i\rangle$ mixed in the proportion $|c_i|^2$ (to match statistically with the Born rule).
3. *Solution*: find some (dynamical) process that turns ρ_{int} into $\rho_{diag} = \sum_i |c_i|^2|s_i\rangle\langle s_i|$.
4. The ensemble described by the density matrix ρ_{diag} is statistically equivalent to the desired classical ensemble.
5. Therefore, the ensemble is *ontologically* equivalent to such a "classical" ensemble.
6. Therefore, the ensemble described by ρ_{diag} consists of systems, of which a portion weighted by the $|c_i|^2$ is in the state $|s_i\rangle$.

Basically, one can see what is wrong with this kind of argument by keeping an eye on the distinction between the claims *a.* and *b.* from the previous subsection. I started from the assumption that the pure state $|\Psi\rangle$ describes a (non-classical) ensemble to end up arguing that the individual members of a (now different) ensemble each are in a pure state $|s_i\rangle$ (and consequently, that the ensemble they constitute is "classical"). This is rather confusing. Of course there is nothing special about the $|s_i\rangle$ states that justifies their interpretation in terms of the individual systems, if such an interpretation is not also warranted for the $|\Psi\rangle$ state one started out with.

(Just start with the $|s_i\rangle$, take a different observable and expand the state in the corresponding eigenbasis.)

It is clear that the argument goes astray at step 5 (this is in fact the often repeated critique of the Ignorance Interpretation in a nutshell). However, the awkward conflict between the premise and conclusion of the argument is usually obscured because the presentation of the IgI generally starts from the diagonal density matrix ρ_{diag}. The claim is then simply that a mixed state as ρ_{diag} should be interpreted as an ensemble of pure states because the latter have a certain ontological priority.

That this formulation of the IgI (omitting the first three steps in the argument above) is problematic, is often explained by pointing at the distinction between "proper" and "improper" mixtures (see for instance (van Fraassen, 1991)). This is a different kind of criticism than the above, since it explicitly takes into account the particular means by which decoherence is usually obtained. Proper mixtures describe genuine physical ensembles, but an improper mixture is what results when a subsystem of an entangled pair is 'taken by itself'. To see what this means, consider the entangled state on the product space $\mathcal{H}_1 \otimes \mathcal{H}_2$ (represented as a density matrix)

$$\rho_{ent} = \sum_i c_i c_j^* |\phi_i\rangle\langle\phi_j| \otimes |\psi_i\rangle\langle\psi_j| \tag{1.11}$$

'Taking a subsystem by itself' (say the subsystem on $\mathcal{H}_1$) means that one considers observables of the form $\hat{A} \otimes \hat{\mathbb{I}}$ only. Since the expectation values of these observables on the state (1.11) equal the expectation values of the observable $\hat{A}$ on the reduced density matrix obtained by tracing out the states $|\psi_i\rangle$ on $\mathcal{H}_2$

$$Tr^{\mathcal{H}_2}(\rho_{ent}) = \rho_1^{imp} = \sum_k |c_k|^2 |\phi_k\rangle\langle\phi_k| \tag{1.12}$$

this reduced density matrix ρ_1^{imp} is usually regarded as representing the state of the subsystem on $\mathcal{H}_1$. Although formally this expression equals that of a statistical ensemble, the ontological meaning is of course quite different. This therefore suggest that one cannot simply identify any diagonal density matrix with a physical ensemble of states.

Apart from this conceptual objection, one can also easily see that the IgI will not work formally. Note that for the second subsystem on $\mathcal{H}_2$ one has:

$$Tr^{\mathcal{H}_1}(\rho_{ent}) = \rho_2^{imp} = \sum_l |c_l|^2 |\psi_l\rangle\langle\psi_l| \tag{1.13}$$

Interpreting the reduced density matrices as representing genuine ensembles, that is that the system on $\mathcal{H}_1$ is in a state $|\phi_k\rangle$ with probability $|c_k|^2$ (and similarly for the state on $\mathcal{H}_2$), would imply that the combined state on $\mathcal{H}_1 \otimes \mathcal{H}_2$ is:

$$\rho_{mix} = \sum_{kl} p_{kl} |\phi_k\rangle\langle\phi_k| \otimes |\psi_l\rangle\langle\psi_l| \tag{1.14}$$

(Note that we did not assume that the distributions of the $|\phi_i\rangle$ and $|\psi_j\rangle$ are statistically independent, i.e. in general we do not have $p_{kl} = |c_k|^2 |c_l|^2$. The actual state of the composite system may very well depend on correlations between the subsystems.) This mixed state is observationally different from the entangled –pure– state we started out with.

It must be noted that what I presented here is what is sometimes called the "strong version" of the ignorance interpretation (Ochs, 1981). It states that the system is in a pure state $|\psi_i\rangle$ which is one of the terms in the mixture $\sum_i c_i c_i^* |\psi_i\rangle\langle\psi_i|$. A weaker version of the ignorance interpretation is to say that mixed states always represent an ensemble, yet it is *not* possible to infer the 'real' pure state of the elements of the ensemble from an expression like (1.6). The latter weak version of the ignorance interpretation seems rather unattractive, but there is a direct and simple argument that forces us to weaken the ignorance interpretation in this respect: the expression (1.6) can be expanded in any arbitrary basis. And if some of the $|c_i|^2$ are equal, we can even decompose $\tilde{\rho}$ in an infinite number of different sets of orthogonal projectors. Without an independent criterion to select a preferred observable that fixes the decomposition, the non-uniqueness of the expansion of (1.6) makes it questionable whether the ensemble interpretation even achieves its modest goal –obtaining the formal equivalent of a probabilistic ensemble of well-determined pointer readings.

Summing up, it seems that the proper way to view the relevance of the transition of (1.4) to (1.6) is merely that it removes the interference terms that have no clear interpretation (and might well be considered empirically superfluous, cf. my discussion of the ensemble interpretation in 1.1.1). In the end however, it cannot explain what *is* observed: the occurrence of definite measurement outcomes.

1.2.3 Measurement as an emergent phenomenon

We have seen that one can characterise interpretations of quantum mechanics in terms of different sorts of distinctions, according to what they regard as the essence of the measurement problem:

1. The "philosophical" approaches that seek a solution to the measurement problem in an interpretation of the quantum state (e.g. modal and relative-state interpretations) vs. the "physical" approaches (GRW, environment-induced decoherence).
2. Individual interpretation approaches that want to explain the problem of Collapse or Outcomes vs. ensemble interpretations that address Interference.

To these, I would like to add a third distinction:

3. Approaches that are directed at the measurement problem only vs. those that want to explain the "emergence of classicality".

For the latter class of interpretations, there is nothing special about measurements; the measurement problem is just a particular instantiation of the more general problem of explaining how the laws of classical physics emerge (perhaps as a limiting case) from the laws of quantum mechanics (which are assumed to be more fundamental). A solution to the question of classicality could imply a solution to the measurement problem in the broader sense of the mmp (for instance, because it can be shown that the superposition principle breaks down for a certain class of physical systems). On the other hand, a solution to the measurement problem does of course not imply a solution to the question of classicality.

Clearly, not all interpretations share the view that the measurement problem should be addressed through the idea of emergence. For instance, modal interpretations are constructed solely in order to solve the measurement problem. This in contrast to, for instance, Bohmian mechanics which is claimed (by the proponents, such as Goldstein (2006)) to explain many more features of classicality than just the definite pointer reading. In section 1.1.1, I have argued that the (usually implicit) idea of the physically-oriented approaches is that one can avoid having to face the difficult task of interpreting quantum mechanics if it can be shown that the kind of phenomena that are considered to be "typically quantum" do not show up at the macrolevel. Similarly, it seems natural that the question of emergence is not the main concern for the "philosophically" oriented approaches.

Nevertheless, it is not always easy to assign to interpretations one of the labels of the third category. This is because there is no clear-cut notion of what 'classical behavior' is, or what phenomena are in need of an explanation the standard tools of quantum mechanics cannot provide. In any case, it is important to distinguish the following general question

> **Emergence of classicality:** Can one derive the laws of classical physics (in an approximate sense) from the quantum formalism?

from the more specific version:

> **Emergence of classical definiteness:** Why are "ordinary" objects never observed in a superposition of definite, "classical" properties?

By "classical properties" I mean the physical quantities of classical physics and everyday reality, which we generally perceive to have a definite value. The example generally referred to in the literature is position: quantum mechanically, it makes sense to speak of a superposition of positions, but with respect to our daily experiences it does not.

An answer to the second question is an obvious precondition for addressing the first. With respect to the measurement problem, only the second is relevant, but environment-induced decoherence has been claimed to account for the emergence of classicality in a broader sense as well. Apart from a brief digression in section 2.3 in the next chapter, I will largely ignore these kinds of arguments.

Note that I have made a number of implicit assumptions. First of all I suppose that we know what is meant by 'ordinary' objects. It is tempting to replace 'ordinary' by 'macroscopic'. But where is the line between 'microscopic' and 'macroscopic' to be drawn? As we will see in the next chapter, it has been claimed that the superposition principle starts to break down at the level of few-atom molecules already. At the same time, for instance, experiments have been performed showing quantum interference of C_{70} molecules (Hackermüller et al., 2004). Although the quantum-classical border must be somewhere at the molecular level, this is still a difference of several orders of magnitude. Indeed, any reasonable clarification of the phrase 'ordinary' here presupposes that an answer to the question has already been given, and I will not attempt to do that here.

Thus, adopting this ambiguous phrase, one could say that "ordinary" objects are, indeed, never observed in superpositions of certain classically definite properties (such as position). I explicitly say ‘observed’ because some approaches claim that classicality is a valid concept ‘for all practical purposes’ only. That is to say: the superposition principle is universally valid, but the consequences of this principle can only be perceived for a very restricted class of physical systems, we therefore call ‘quantum’. (Yet note that the question of ‘Classicality’ as it is phrased here does not preclude the idea that these superpositions are *not there* at all.) Such is for instance the idea of GRW.

As we will see in the next chapters, the position of environment-induced decoherence with respect to the distinction between these two problems -the measurement problem and the emergence of classicality- is a little ambiguous. Environment-induced decoherence was initially put forward as a solution to the measurement problem, but in the past two decades proponents of decoherence started to step back from this position and the “emergence of classicality” has become the major focus of the programme instead. Nowadays it is not uncommon to boldly admit that environment-induced decoherence does not solve the measurement problem. Nonetheless the latter seems to re-enter through the back door, in the guise of a more general problem of localization of macro-objects. I think that this is quite peculiar, because the reason that decoherence was rejected as a solution to the measurement problem was precisely because it has only statistical relevance and does not seem to have anything to say about individual systems or events; nonetheless, the question of “Classicality” is formulated in the same terms. The argument is simply a further elaboration of the idea that the disappearance of interference somehow implies definiteness. I have already rejected this idea, and the fact that it is now embedded in a broader program of the emergence of classicality does not seem to add anything in its favor. In fact, I would say that things have only gotten worse. Because in the previous case of the IgI there was at least *statistical* support for the claim that each element in the ensemble is in a state corresponding to a definite value for the preferred observable. However, when we turn to the issue of Classicality, this statistical support is no longer available, since this approach implictly depends on a individual interpretation.

This objection also holds for more sophisticated formulations of “classical behavior” which do not refer to the measurement problem (for instance, the derivation of classical equations of motion from the quantum dynamics). Apparently, it is commonly accepted that one can obtain information about the (classical) features of an individual quantum system from the characteristics of its state. This idea is well-worth pursuing, but it raises the question of what the real relevance of the disappearance of the interference terms from the density matrix is.

1.3 Concluding remarks

The remarks of the last section should not be taken to mean that my main question -the relevance of environment-induced decoherence for the measurement problem- is in a certain sense misguided. It is just that some of the authors writing about decoherence seem not to want to touch this controversial issue, and therefore focus on the more general, physics-oriented issue of the emergence of the classical from the quantum. Moreover, even in the philosophical literature it is often stated that “decoherence does *help* with the problem” (Dickson, 2007), however without making the meaning of this statement much more precise. It is in the latter respect that I hope to make a modest contribution in the remaining chapters.

If we are to evaluate the relevance of environment-induced decoherence for the measurement problem, we should first of all try to establish what the exact problem is the theory aims to solve. I think it would be a bit unfair to accuse a theory of failing to solve a problem it was never supposed to address in the first place. At most it would lead to the rather trivial, though not unimportant, conclusion that the theory does not address the *really essential* questions. In fact, I think that the usual criticisms of decoherence boil down to exactly this objection. With the ignorance interpretation of mixtures proved untenable, the central problem of Outcomes is left untouched, and therefore the theory cannot account for the occurrence of "facts" in standard quantum theory (without collapse postulate), it is said.

This immediately gives rise to a couple of points that need to be clarified. The first is whether environment-induced decoherence is in fact capable of solving the problems it *does* explicitly aim to address, viz. the problem of Interference and the Preferred Basis problem (in either of two versions). I tend to be more sceptical about this than most authors, but that's something to be discussed later. Second, the idea that decoherence nevertheless *helps* with the problem suggests that these questions do have some kind of conceptual relevance of their own. However, especially with regards to the problem of Interference the precise nature of this alleged relevance is so far still an open question. Third, the conclusion, drawn by the philosophers, that environment-induced decoherence cannot account for the appearance of 'facts'or or 'definiteness' is definitely not shared by all the workers in the field. (Although they tend to tone down their claims, by invoking a somewhat weaker notion of definiteness, about which I will say more in a later chapter (section 3.3.1).) As I suggested in the last section it still lingers on in the idea that environment-induced decoherence leads to the appearance of classicality.

Postponing the first of these issues to chapter 3, I would like to close this chapter by evaluating briefly what the foregoing discussion tells us about the other two issues. First of all, however, let me repeat once more what the decoherence theorist takes to be the measurement problem and the right kind of approach to address it.

The decoherence theorists' measurement problem

The decoherence theorist wants to address the measurement problem by staying entirely within the standard quantum formalism. In particular, it is claimed that no interpretation of the wave function is needed. I take this to mean that they take the orthodox stance that the wave function yields probabilities for measurement outcomes, for without an interpretation one cannot link the formalism to any kind of facts of the matter. (And I think it is clear from the technical discussions that no peculiar interpretational moves are being made as long as it does not come to the question of how to interpret the diagonal density matrix.) Second, for the same reason, the eigenstate-eigenvalue link is maintained.

More elusive is the question whether decoherence adopts a individual or ensemble interpretation of the wave function. Assuming the first, the implication of this "standard" view is that the measurement problem is the problem of Collapse; the state on the right-hand side of (1.2) does not represent a measurement apparatus with definite properties unless it somehow "collapses" into one of the eigenstates representing such a property. Since the von Neumann collapse is (mistakenly) identified with the transition of (1.4) to (1.6), this notion is weakened in the sense of an "apparent collapse" if the ignorance interpretation of mixtures is rejected. (That is to say, if one applies the individual interpretation of the wave function consistently, this means that the diagonal density matrix somehow represents a single system that "appears" to be in one of the states on the diagonal.) What this "appearance" is supposed to mean within a standard account of the wave function is the first hurdle the decoherence approach has to take, and it is here where authors that take this question seriously generally depart from the standard account and turn to, for instance, an Everettian point of view. (See chapter 4.)

On the other hand, if one adopts an ensemble interpretation the "measurement problem" does not take the standard "Schrödinger Cat" form[32]. The question is not about obtaining "definiteness", "facts" or "outcomes", for the ensemble interpretation *presupposes* their (unexplained) existence (perhaps predicated on the specification of a preferred basis). The problem becomes that of obtaining a formal expression that complies with the actual statistical distribution of measurement outcomes. In this respect decoherence (addressing the problem of Interference) seems to be a legitimate approach.

I have a couple of things to say about this, however. First of all, the fact that decoherence is supposed to lead to the "classical" appearance of macro-objects, as it is usually stated in the literature, suggests that the statistical interpretation is not at all what the decoherence theorist has in mind. Furthermore, as I already noted in section 1.1.1, I feel that there exists a certain tension between this kind of statistical reasoning and the supposition that environment-induced decoherence solves (or needs to solve) the Preferred Basis problem. It seems that if the problem is purely statistical, formulated within the orthodox framework, yet at the same time linked to a particular observable or measurement context, the presence of the interference terms is not really relevant. For, after all, these interference terms do not affect the statistics of the measured observable. The fact that so much importance is attached to the vanishing of interference gives the impression that the approach is rooted in a deeper interpretational problem.

So let us assume that the right way to view the decoherence approach is in the light of a individual interpretation of the wave function. In that case, I think there is a problem which is central for all the "physics-oriented" approaches towards a solution of the measurement problem (i.e. those approaches that are happy with the orthodox interpretation, except for the collapse postulate.) The orthodox interpretation of the wave function is that it gives the probabilities of measurement outcomes (in terms of a specified observable). The goal of the physics-oriented approaches is to establish the conditions under which the quantum state, in the light of the eigenstate-eigenvalue link, gives rise to the appearance of these outcomes. In order to do so, however, these approaches need to make their formal results empirical. In particular, they often need to justify certain steps in their derivation by means of an empirical argument, *prior* to having established the formal equivalent of the measurement outcomes that *are* the empirical content of the orthodox

---

[32] In this respect, it is interesting to note that the Schrödinger Cat paradox was originally suggested to Schrödinger by Einstein, as an argument in *support* of a statistical interpretation, and *against* Schrödinger's own interpretation (which I would characterize as individual), as a kind of *reductio ad absurdum* argument. (See also footnote 13). Schrödinger in fact agreed about the absurdity of it all (Fine, 1996). Remarkably, in later discussions the paradox –i.e. the absurd consequence – has been taken much more seriously than it was ever intended.

formalism. (My criticism of the usual "empirical" treatment of approximate decoherence in section 3.3.3 of chapter 3 hinges on this.) If this is the case, the very idea of such a physical approach seems to rest on a *petitio principii.*

Where we are

What does this imply for the points I mentioned at the start of this section? As I said, I will postpone the first question to chapter 3, after we have considered the specific means by which environment-induced decoherence is supposed to achieve its goal. The other two issues will also need to be reconsidered then, but there are some preliminary conclusions I would like to draw.

First of all, I think that if my criticism of the previous paragraph is justified, it answers the second question in the negative; despite its physical relevance, decoherence does not have any *conceptual* relevance. Not more than any other well-confirmed quantum-mechanical calculation, in any case. That the empirical predictions of decoherence can be verified empirically simply indicates that it is good physics - that quantum mechanics is an accurate physical theory that, if properly used, yields the right kind of empirical predictions. It is a good thing to confirm this once more, but it is not an answer to the kind of questions the decoherence theorist is asking. So something more is needed to justify the claim that decoherence helps with solving the measurement problem. And I think that this kind of justification is to be found in the way decoherence addresses the Preferred Basis problem, and how this supports various interpretations of quantum mechanics that do go beyond the orthodox framework.

In fact, the idea that the real relevance of decoherence lies in the way it addresses the preferred-basis problem has been emphasized by philosophers such as Bub (1997) as well as decoherence theorists such as Zurek (1981, 2003) (albeit incidentally and not as forcefully). The Preferred Basis problem does not touch the issue of Outcomes, or so it seems. However, the fact that environment-induced decoherence is supposed to select a certain "preferred basis" has often been interpreted as meaning that environment-induced decoherence also implies definiteness in terms of these states. If that is correct, decoherence would lead to a set of definite measurement outcomes, yet would be unable to account for the occurrence of a particular outcome. This is not merely a problem of indeterminism; if outcomes occur "out of the blue" one should at least be able to say *when* this happens. If the answer is supposed to be "at the moment of observation" we are back to square one. In any case, I will argue that decoherence does not solve the preferred basis problem, at least not in the sense that the decoherence theorists have in mind. The issue is rather subtle; I will discuss it in section 3.3.1 of chapter 3, after having explained what it actually means that environment-induced decoherence selects a preferred basis.

The idea that decoherence implies definite outcomes even though it cannot account for them is what I mean by the third of the three points at the start of this section, i.e. when I say that the problem of Outcomes is not a dead end for most decoherence theorists. Furthermore, I think that it is fair to say that decoherence *was* in fact intended to answer the question of Outcomes - to explain the appearance of a collapse- as I argued above; either through an ignorance interpretation or in a more roundabout way.

So far the general remarks. In the following chapters I will focus specifically on the methods and ideas employed by environment-induced decoherence, to see what precisely is being claimed and whether these technical details shed some new light on the criticism presented here.

Chapter 2

The theory of environment-induced decoherence

The topic of this chapter will be the theory of "environment-induced decoherence". My aim is to give an overview of the subject, with special emphasis on those issues that are most relevant for the foundations of quantum mechanics and in particular the measurement problem. So after briefly sketching the historical background of decoherence, I will focus on the concept of "environment-induced superselection rules" and the way in which the preferred-basis problem is addressed. I will also, albeit briefly, consider the role of decoherence in the quantum-to-classical transition.

2.1 Preliminaries

Basically, the word 'decoherence' means the destruction of phase relations in the state of a quantum mechanical system as a result of some dynamical process; in other words, such a process will cause off-diagonal terms in the density matrix to disappear. As a consequence, interference effects and the typically quantum mechanical correlations due to entanglement vanish. In this bare form, decoherence is an unavoidable fact of the physicists' everyday life. Experiments that exploit phase relations in entangled states (as for instance in quantum computation) are in general so hard to carry out just because of this loss of coherence between different eigenstates.

However, the phrase 'decoherence' is often supposed to mean more than just this. First of all, the word 'decoherence' has become closely intertwined with what is nowadays considered to be its main cause: the overall and (almost) unavoidable interaction of a quantum system with its environment. In the present chapter I will generally not distinguish between these two meanings, but if necessary I will do so by adding the phrase "environment-induced" (or "environmental") to designate the specific cause of decoherence. (As I will discuss later (sect. 3.3.4 in chapter 3) in more detail, the phrase "environmental" allows a somewhat liberal interpretation: what is essential is that a division is made between degrees of freedom which are "of interest" (the system) and a remaining set of degrees of freedom (the environment) that interact with the system but nevertheless are "irrelevant" and therefore ignored.)

Second, there is a strong connection between this research program that studies the physical effects of environmental interactions, and the more philosophical aim of answering the kind of foundational questions that center around the "measurement problem", or, more generally, the "emergence of classicality". Specifically, decoherence is often supposed to imply that the quantity represented by the states on the diagonal appears to have a definite value. This idea can be made more precise in terms of what is usually called "environment-induced superselection", or "einselection". In recent years, the focus has gradually shifted from superselection – an effect which is particularly relevant for the measurement problem – to the more ambitious program of deriving the appearance of a 'classical' realm from these environmental interactions.

In what follows I will try to give a representative overview of the current state of affairs in decoherence research[1]. What the real relevance of these results is will be a topic for a critical evaluation in the next chapter.

2.1.1 A brief historical and conceptual overview

Because the theory of decoherence has not been the invention of a single person at a single instant, its origins are not very clear. From one point of view decoherence is a theory of the effect of environmental interactions on a quantum system. As such its (classical) roots are sometimes traced back to models in classical statistical physics, as early as 1914, when Borel formulated the Sirius problem (Joos et al., 2003, p.41): he showed that the displacement by a few centimeters of a small mass as far away as the star Sirius will inevitably and almost instantaneously disturb the particle trajectories in a vessel of gas here on earth. This disturbance is strong and rapid enough to render the determinacy of particle positions for the *isolated* system a "pure abstract fiction". This particular example is claimed to be a precursor to decoherence because it shows that even in classical systems, the influence of the environment on a system cannot always be neglected, *no matter how small their mutual interaction.*

The Sirius problem is interesting because it radically departs from what is standard practice in physics: to 'cut off' the object of interest from the rest of the world, and apply the laws of physics to the object in isolation. The Sirius problem shows that this idealization cannot always hold unconditionally. The first suggestions that this might also be relevant to the measurement problem can be found in the early papers of Zeh (1971, 1973). This is often regarded as the origin of the decoherence programme[2].

Interacting quantum systems differ fundamentally from classical systems, however, due to the concept of entanglement, which has no classical counterpart. The essence

---

[1] Because I would like to confine myself to the established kind of research, I will not include Zurek's programme of "Quantum Darwinism". See chapter 4 for a critical discussion of the latter.

[2] Stamp (2006) traces the idea of decoherence back to the work of Ludwig in the 1950's, and Kiefer (in Joos et al., 2003) even to Landau in the 1920's.

of decoherence models is that they study essentially the (local) effects of entanglement[3]. Interactions governed by entanglement are basically about phase exchange, whereas classical interactions (virtually) always require energy exchange. (This does not mean that energy exchange is always absent from the decoherence dynamics however. Cf. Stamp (2006).) This results in the often repeated claim that in decoherence the system is perturbing the environment, rather than the other way around[4].

If one emphasizes the diagonalizing effect of decoherence, it bears close resemblance to the theory of superselection rules (either implemented axiomatically in the algebraic formulation of the theory or derived in the limit of infinite times and infinitely-sized systems, as in Hepp's (1972) approach of quasilocal observables). And if one regards decoherence as the result of ignorance about a large number of degrees of freedom, it is in some sense comparable to early approaches to the measurement problem, which purport a kind of 'thermodynamic' point of view. (For instance Daneri et al. (1962).) A subtle difference with these approaches is that in decoherence the environment is modelled as possessing a specific state, whereas in thermodynamic approaches the state of the environment is assumed to be "indeterminate" (i.e. modelled statistically). The idea of the latter approaches seems to rely on a kind of *cryptodeterminism* (Pessoa, 1998); indeterministic aspects of quantum measurements would result from an inexact specification of the state of a macroscopic system, or from the complicated and intractable interactions between the system and the measurement apparatus (and perhaps its environment.) It must be emphasized that this is not what environmental decoherence claims (it even conflicts with the decoherence view, cf. section 3.3.2), since decoherence does not aim at restoring determinism.

Nevertheless, both the thermodynamic approaches and decoherence are motivated by the ignorance resulting from the observer's limited observation or information-gathering abilities. Despite this conceptual similarity, environmental decoherence should not be confused with mechanisms like 'phase randomization' which also exist in systems with a large number of degrees of freedom and which often leads to the claim that effective diagonality is a consequence of the uniform distribution of phases in the component states of a macroscopic object when only ensemble averages are important[5]. Rather, as I said, the central mechanism driving decoherence is entanglement: the idea is that typical 'quantum' characteristics of a system are contained in its correlations with the environment which cannot be intercepted.

After Zeh's first publications, it took another decade before Zurek (1981, 1982) and Joos and Zeh (1985) made their first attempts to shape these suggestions into a well-developed theoretical framework[6]. In two papers published in the early 1980's,

---

[3] Although Stamp (2006) argues that in oscillator-bath models, "dissipation and decoherence both arise from *energy exchange* between system and bath" (p. 479, my emphasis).)

[4] For instance, (Joos and Zeh, 1985, p. 224) say: "The destruction of interference terms is often considered as caused in a classical way by an 'uncontrollable influence' of the environment on the system of interest. [...] But the opposite is true: The system disturbs the environment, thereby dislocalizing the phases."

[5] As Joos (Joos et al., 2003, p. 177) points out, this kind of dephasing between different subsystems only results from an "incomplete description", which he, apparently, does not consider an essential part of decoherence. More importantly, dephasing applies to phase relations between the subsystems of composite systems and as such does not say anything about loss of coherence between different eigenstates *within* a component system.

[6] Camilleri (2008) gives an interesting account of the early years of decoherence. It appears that

Zurek took the first steps towards his research programme in decoherence (called 'einselection', but we will not use this term for Zurek's results exclusively) he is pursuing up till the present date. Where Zeh's and Zurek's early papers focused on the measurement process, Zurek's more recent aim is to investigate the emergence of classicality in a broader context.

In modern decoherence research, *generalized master equations* are an important tool. A generalized master equation is basically the dynamical generalization of the simple decoherence models presented in section 2.2.1 below: the combination of quantum system, apparatus and environment evolves according to the Schrödinger equation, but this unitary evolution is 'interrupted' by eliminating the environmental degrees of freedom by means of a partial trace. A generalized master equation describes the dynamical evolution of an open quantum system, in which this elimination is already incorporated. (As a consequence, it contains non-unitary terms.) The mature formulation of generalized master equations and its application to decoherence processes was established by Caldeira and Leggett in the 1980's. Their method was however not incorporated yet in the earliest publications of Zurek who used a simplified but more transparent picture to point out the significance of the interactions for the existence of a preferred basis. (See section 2.2 below.)

What makes (the modern formulation of) einselection (in the eyes of many physicists) a more sophisticated theory than previous attempts to solve the measurement problem, is that it is not merely a qualitative, general argument: it yields numerical values for decoherence rates (time scales and effective width of wave packets) which can be empirically tested. In contrast to the early thermodynamic or phase-averaging arguments, einselection predicts the occurrence of observable quantum effects in meso- or even macroscopic systems (under laboratory conditions). Recently, some of these theoretical predictions have been confirmed, be it in most cases only qualitatively[7]. These experiments include the observation of so-called Schrödinger-Cat states: superpositions of macroscopically different states, for instance oppositely directed current-states in a superconducting SQUID. Unfortunately, there still exists disagreement whether the predictions of einselection have been conclusively verified by these experiments. Most decoherence theorists are positive that the experimental results are in excellent agreement with the theoretical predictions, but that might be a case of wishful thinking[8].

But even if the theory can be made to agree with the experimental tests, the question of the foundational significance of environmental decoherence is an altogether different issue. In fact, the price one pays for a realistic model that agrees with experiment is often an increasing lack of intelligibility. Bringing theory and experiment in close agreement is a goal worth striving for, but focusing solely on these problems may hide the conceptual issues at stake. Therefore, in what follows I will

---

Zeh's work was not well-received because of its strong commitment to an Everett-type interpretation. Zurek (who did not now of Zeh's publications when he started to develop his ideas) paved the way for general acceptance by freeing the dynamical claims from a particular philosophical program.

[7] For a critical discussion of recent experimental developments in decoherence and its consequences, see (Leggett, 2002; Stamp, 2006).

[8] As an example; in the preface of the second edition of his book on decoherence, Joos (2003) states that "the phenomenon is now experimentally clearly established and theoretically well understood in principle.". But for instance Leggett (2002) and Stamp (2006) are more critical. Stamp states that the experimental quantitative decoherence rates are *always* found to conflict with the theoretical predictions (which are usually several orders of magnitude smaller)

focus on the general line of argument in environmental decoherence, although this rarely applies to the laboratory situations in which decoherence has been established as a more or less well-confirmed fact. As I said in the concluding remarks at the end of chapter 1, there is a difference between decoherence effects as a physical fact, and the claimed relevance of decoherence for the emergence of classicality, the measurement problem, or the philosophical questions associated with it.

2.2 Decoherence and the measurement problem

Although nowadays many of the researchers in the field feel that the real significance of decoherence lies in the way it addresses certain aspects of the broader question of the quantum-to-classical transition, it was originally proposed as a solution to the measurement problem. Since this is the central topic of my story, we will first have a look at how this particular goal was supposed to be achieved. Nevertheless, it is important to keep in mind that much research in the field of environmental decoherence is no longer devoted primarily to the measurement problem. In fact, many prominent researchers in the field nowadays do not hesitate to claim that "decoherence does not solve the measurement problem" (Zeh in (Joos et al., 2003, p. 21)). (That there still is no consensus on this particular important issue, is probably because it depends on what one considers to be the measurement problem exactly, cf. chapter 1.)

So, in order to avoid confusion, it is important to keep an eye on *what* exactly is claimed to be solved. As we have seen in the previous chapter, decoherence has as its starting point a rather specific (although not uncommon) version of the measurement problem. In brief (for a more detailed argument and criticism I refer to chapter 1, in particular section 1.3):

1. The unitary Schrödinger evolution is universally valid and measurements are just ordinary physical interactions.
2. The simplified von Neumann measurement scheme is incomplete: it neglects the interaction of the quantum system and apparatus with their environment.
3. There is no need to formulate an interpretation of the wave function; decoherence is a purely physical approach that aims to show that typical quantum effects disappear in ordinary circumstances.
4. With these ingredients, decoherence attempts to answer the question of Interference: the kind of (physical) process responsible for the vanishing of interference terms from the density matrix (in the basis of pointer states) is basically the interaction with the environment[9].

The decoherence literature distinguishes the Preferred Basis problem and the problem of Outcomes (see for instance Schlosshauer (2004)), but that there is a difference

---

[9] Two important issues discussed in the previous chapter are not mentioned: whether the e-e link holds and whether one adopts a individual or ensemble interpretation of the wave function. As I already suggested, decoherence approaches do not take a clear position with respect to the latter question. Concerning the eigenstate-eigenvalue link; this is assumed to hold, but in a "modified" sense that is supposed to do justice to the diagonal density matrix as an "apparent" ensemble. This is a subtle issue that I will address in sections 3.3.1 and 3.3.2 of the next chapter.

between the two versions of the first problem (general vs. decomposition) that I formulated in the previous chapter, is usually neglected. The received view is that environmental decoherence is capable of solving either of the two versions of the Preferred Basis problem. More controversially, environmental decoherence is sometimes (especially in more elementary discussions) claimed to solve the measurement problem as a whole. However, we have seen (section 1.2, see also section 3.3.2 in the next chapter) that decoherence leaves the problem of outcomes essentially untouched. After giving a brief outline of Zurek's formulation of decoherence, I will successively consider its answer to the Preferred Basis problem and the meaning of "environment-induced superselection rules". The purpose of the present discussion is to give an overview of decoherence as it is formulated in the relevant literature; in the next chapter I will try to shed some light on these sometimes obscure formulations by giving a summary account from a more personal point of view.

2.2.1 Environment-induced decoherence.

In what follows I will focus on the publications of Wojciech Zurek (unless stated otherwise), since his formulation of decoherence is generally regarded as the prototype and since Zurek introduced most of the terminology. The word used by Zurek to characterize his own work is 'einselection'. (Shorthand for environment-induced superselection.) Einselection is in fact more than the vanishing of interference terms or an attempt to solve the measurement problem. To be precise: in Zurek's terminology the interaction of the environment has three main consequences: decoherence, einselection and the effective classicality of einselected states:

> "Decoherence and einselection are two complementary views of the consequences of the same process of environmental monitoring. Decoherence is the destruction of quantum coherence between preferred states associated with the observables monitored by the environment. Einselection is its consequence -the *de facto* exclusion of all but a small set, a *classical domain* consisting of pointer states- from within a much larger Hilbert space. Einselected states are distinguished by their resilience -stability in spite of the monitoring environment." (Zurek, 2003, p.717)

Here, what Zurek calls decoherence is particularly relevant to the measurement problem; especially if it implies the existence of superselection rules, as Zurek claims. What the concept of einselection is supposed to add to the mathematical formalism of superselection rules is (1) a physical motivation for their existence and (2) the effective classicality of superselected states. (To be made more precise in section 2.3.) At the same time, the theory should maintain the usual predictions of quantum mechanics where appropriate. It should therefore distinguish between typical "quantum" contexts and circumstances that give rise to effectively classical behaviour. This is guaranteed by another prominent feature of Zurek's program, namely the idea that the 'pointer basis', which is essentially the eigenbasis of the superselected observable, is picked out by the apparatus-environment interaction Hamiltonian.

The basic idea of decoherence is quite simple: let us assume that our system of interest $\mathcal{S}$ is initially in a superposition of eigenstates $|s_i\rangle$ (of the to-be-measured observable):

$$|\Phi_{\mathcal{S}}\rangle = \sum_i \alpha_i |s_i\rangle. \tag{2.1}$$

Prior to the measurement interaction this state is coupled to the 'ready state' $|a_0\rangle$ of the pointer *and* the corresponding environment state $|e_0\rangle$.

$$|\Phi_{\mathcal{SAE}}\rangle_i = \left(\sum_i \alpha_i |s_i\rangle\right) |a_0\rangle|e_0\rangle. \tag{2.2}$$

After measurement the pointer-environment coupling evolves the initial apparatus-environment product state into an entangled state on $\mathcal{H}_{\mathcal{A}} \otimes \mathcal{H}_{\mathcal{E}}$:

$$|\Phi_{\mathcal{SAE}}\rangle_f = \sum_i \alpha_i |s_i\rangle|a_i\rangle|e_i\rangle, \tag{2.3}$$

The density matrix of this expression is

$$\rho_{\mathcal{SAE}} = |\Phi_{\mathcal{SAE}}\rangle\langle\Phi_{\mathcal{SAE}}| = \sum_{ij} \alpha_i\alpha_j^* |s_i\rangle\langle s_j| \otimes |a_i\rangle\langle a_j|. \otimes |e_i\rangle\langle e_j|. \tag{2.4}$$

Now, the crucial assumption is that 'for all practical purposes of statistical prediction' (meaning that the physical state of the environment remains unknown) this density matrix can be replaced by the reduced density matrix $\rho_{\mathcal{SA}}$ obtained by tracing over the degrees of freedom of the environment:

$$\rho_{\mathcal{SA}} = \mathrm{Tr}^{\mathcal{E}} \rho_{\mathcal{SAE}} = \sum_{ij} \alpha_i\alpha_j^* \langle e_i|e_j\rangle . |s_i\rangle\langle s_j| \otimes |a_i\rangle\langle a_j| \tag{2.5}$$

This expression would reduce to that of an incoherent mixture

$$\rho_{\mathcal{SA}}^{dec} = \sum_i |\alpha_i|^2 |s_i\rangle\langle s_i| \otimes |a_i\rangle\langle a_i|, \tag{2.6}$$

if somehow we could show that the environment states $|e_i\rangle$ are orthogonal, or at least that their inproducts vanish rapidly (at an acceptable rate compared to the interaction time). This is the principal aim of decoherence theorists: to construct models that show exactly how fast and under which circumstances the environment-pointer states approach orthogonality. Most notably by Zurek (but also by many others), various models have been studied, ranging from simple and unrealistic to complex and somewhat less unrealistic. In general, these models tend to show that orthogonality is achieved very effectively as a result of the large number of environment states. I will illustrate this with a simple (and unrealistic) model, as presented by Zurek (1982). (See also Schlosshauer (2004).)

As an apparatus $\mathcal{A}$ one takes an atom with two orthogonal pointer states ('spins'), denoted $|\uparrow\rangle$ and $|\downarrow\rangle$ respectively. Let the environment $\mathcal{E}$ consist of N such two-state atoms, denoted $|\Uparrow\rangle$ and $|\Downarrow\rangle$. One assumes that neither the system nor the environment has a self-Hamiltonian, and that the interaction between system and apparatus is negligible after a short time, so that only the interaction Hamiltonian coupling apparatus and environment remains:

$$H^{\mathcal{AE}} = (|\uparrow\rangle\langle\uparrow| - |\downarrow\rangle\langle\downarrow|) \otimes \sum_k^N g_k \left(|\Uparrow_k\rangle\langle\Uparrow_k| - |\Downarrow_k\rangle\langle\Downarrow_k|\right) \bigotimes_{k'\neq k} \hat{\mathbb{I}}_{k'}. \tag{2.7}$$

Here, g_k are coupling constants and $\hat{\mathbb{I}}_k$ denotes the identity operator of the kth environmental state. What this expression says, basically, is that the apparatus states couple to each corresponding environmental degree of freedom (which are similar) separately, with a certain strength g_k.

The initial state (2.2) is now replaced by the initial state with the complete environment taken into account[10]

$$|\Psi(0)\rangle = \Big(\alpha_1|s_1\rangle|\uparrow\rangle + \alpha_2|s_2\rangle|\downarrow\rangle\Big) \bigotimes_k^N \Big(\epsilon_{1(k)}|\Uparrow_k\rangle + \epsilon_{2(k)}|\Downarrow_k\rangle\Big). \tag{2.8}$$

The interaction (2.7) yields a time-evolution operator (setting $\hbar = 1$)

$$e^{-iH^{\mathcal{AE}}t} = e^{-i(|\uparrow\rangle\langle\uparrow|-|\downarrow\rangle\langle\downarrow|)t} \prod_k \left(e^{-ig_k(|\Uparrow_k\rangle\langle\Uparrow_k|-|\Downarrow_k\rangle\langle\Downarrow_k|)t} \bigotimes_{k'\neq k} e^{-i\hat{\mathbb{I}_{k'}}t} \right). \tag{2.9}$$

Applying this to the initial state (2.8) gives

$$|\Psi(t)\rangle = \alpha_1|s_1\rangle|\uparrow\rangle|\mathcal{E}_1\rangle + \alpha_2|s_2\rangle|\downarrow\rangle|\mathcal{E}_2\rangle, \tag{2.10}$$

where $|\mathcal{E}_1\rangle$ and $|\mathcal{E}_2\rangle$ are the different environmental states that are correlated to the two measurement outcomes:

$$|\mathcal{E}_1\rangle = \bigotimes_k^N \Big(e^{ig_k t}\epsilon_{1(k)}|\Uparrow_k\rangle + e^{-ig_k t}\epsilon_{2(k)}|\Downarrow_k\rangle\Big); \tag{2.11}$$

$$|\mathcal{E}_2\rangle = \bigotimes_k^N \Big(e^{-ig_k t}\epsilon_{1(k)}|\Uparrow_k\rangle + e^{ig_k t}\epsilon_{2(k)}|\Downarrow_k\rangle\Big). \tag{2.12}$$

Tracing over the environment gives the reduced density matrix

$$\begin{aligned}
\rho_{\mathcal{SA}}(t) &= \mathrm{Tr}^{\mathcal{E}}|\Psi(t)\rangle\langle\Psi(t)| \qquad (2.13)\\
&= |\alpha_1|^2|s_1\rangle\langle s_1| \otimes |\uparrow\rangle\langle\uparrow| + |\alpha_2|^2|s_2\rangle\langle s_2| \otimes |\downarrow\rangle\langle\downarrow| \\
&+ \gamma(t)\alpha_1\bar{\alpha_2}|s_1\rangle\langle s_2| \otimes |\uparrow\rangle\langle\downarrow| + \bar{\gamma}(t)\alpha_2\bar{\alpha_1}|s_2\rangle\langle s_1| \otimes |\downarrow\rangle\langle\uparrow|.
\end{aligned}$$

The coefficient $\gamma(t)$ determines the relative size of the interference terms. It is given by

---

[10] For notational simplicity I take the Hilbert space of the system to be two-dimensional, but this can easily be extended to the more general, but still discrete, case.

$$\gamma(t) = \langle \mathcal{E}_2 | \mathcal{E}_1 \rangle = \prod_k^N \left(|\epsilon_{1(k)}|^2 e^{2ig_k t} + |\epsilon_{2(k)}|^2 e^{-2ig_k t} \right) . \qquad (2.14)$$

The time dependence of the interference terms is a function of the specific features of the model; for this case in particular of the size of the environment N and the coefficients $\epsilon_{1(k)}$ and $\epsilon_{2(k)}$. Note that for any finite N (2.14) is a periodic function; unless the environment is truly infinite, there will always be a $t > 0$ such that $\gamma(t) = 1$. (At the start of the interaction $\gamma(0) = 1$, i.e. full coherence is present.) Zurek (1982) gives some numerical estimates of the time-dependence of the decoherence coefficient $\gamma(t)$, showing that it drops off very rapidly (and increasingly rapid for larger N) and that fluctuations are suppressed and become more and more rare, with a period comparable to the age of the universe or longer[11]. It can be shown that this already happens for relatively small N. Hence, it is argued, this recurrence will never be observed *in practice*[12]. What results is the effectively incoherent mixture:

$$\rho_{\mathcal{SA}} \approx |\alpha_1|^2 |s_1\rangle\langle s_1| \otimes |\uparrow\rangle\langle\uparrow| + |\alpha_2|^2 |s_2\rangle\langle s_2| \otimes |\downarrow\rangle\langle\downarrow|. \qquad (2.15)$$

2.2.2 Solving the preferred-basis problem

Decoherence is often presented as a physical process that diagonalizes the density matrix according to (2.15). However, mere diagonality proves nothing. This is reminiscent of the general form of the preferred basis problem discussed in chapter 1: the reduced density matrix can be written in many equivalent ways on a different basis, so without some additional basis-selection rule no conclusions can be drawn from the form of the density matrix alone. In fact, since the density matrix is Hermitian, there is *always* a complete basis of eigenstates in which it becomes diagonal; the so-called Schmidt decomposition.[13] Zurek (1998) puts it as follows:

> "This eventual diagonality of the density matrix in the einselected basis is a byproduct, an important symptom, but not the essence of decoherence. [...] Diagonality of [the density matrix] in some basis has been occasionally (mis)interpreted as a key accomplishment of decoherence. This is misleading. Any density matrix is diagonal in some basis. [...] [The preferred basis] will be determined by the dynamics of the open system in the presence of environmental monitoring. It will often turn out that it is overcomplete. It states may not be orthogonal, and, hence

---

[11] This kind of behavior is common to most models of decoherence (Bacciagaluppi, 2005; Schlosshauer, 2004) but not self-evident. In particular the decoherence time scale may depend on the initial state of the system and environment. (For instance, if the environment happens to be in an eigenstate of the interaction Hamiltonian (2.7) there will be no decoherence at all.) An example of this and other violations of certain decoherence 'rules of thumb' can be found in (Anglin et al., 1996).

[12] The finite recurrence time is a consequence of the simplifying assumption that the environment Hamiltonian has a discrete spectrum (Joos et al., 2003, Chap.7).

[13] The Schmidt basis is obtained by diagonalizing the density matrix (which has to be done at each instant of time for dynamical systems). If the density matrix is both almost diagonal and degenerate, the Schmidt basis will differ completely from the decoherence basis. Thus, the concepts of decoherence and diagonalization are not always interchangeable. In chapter 3 I will briefly discuss the philosophical consequences of this incompatibility.

> they would never follow from the diagonalization of the density matrix." (p. 7)

Hence it is clear that decoherence involves something more than simply diagonalizing the density matrix. Moreover, the diagonal density matrix can always be expressed in a different basis which destroys its diagonality. As I argued in chapter 1, what is essential is that there exists a certain *a priori* preferred interpretation basis associated with a certain observable; and it is this basis which environment-induced decoherence is supposed to designate as the (approximate) diagonal basis of the reduced density matrix. (See section 3.3.1 for a more detailed discussion of this point.)

As explained in chapter 1, the preferred basis problem must be addressed if we are to solve the measurement problem, for any interpretation. Even if one takes the more pragmatic view that measurement setups are chosen in such a way as to establish the desired one-to-one correlations between the to-be-measured system states and the pointer states (which can be directly 'read off' by the experimenter)[14]. Because, as we have seen (section 1.1.1), this pragmatic point of view is of limited value: for each state, there exists a multitude of expansions which seem to define such a measurement interaction, but for different system- and apparatus observables from those that are actually measured.

The claim that a preferred basis, which turns out to be the diagonalization basis[15], is selected through environmental interactions has been an essential part of the decoherence programme ever since Zurek entered the debate on the measurement problem. Although his interest occasionally shifted to "environment-induced superselection rules" and the alleged ability of decoherence to explain the "collapse of the wave function", Zurek in general attaches more importance to the ability of the environment to select a certain "preferred basis". In his first paper (1981) this is even presented as the sole purpose:

> "Even though [...] we do not face the insoluble question of quantum theory of measurement: 'what causes the collapse of the system-apparatus-environment combined wave function?' we do determine into what mixture the wave function appears to have collapsed." (p. 1517)

Later, the relevance of the preferred ("einselected") basis was no longer considered to be confined to the context of quantum measurements; for Zurek, the stability, or robustness, of these states is the key feature of an emergent classicality. In this section I give Zurek's (1981) original proposal of the stability criterion (and some of its modifications) that is supposed to take care of the preferred-basis problem. How and why this is supposed to be relevant for the emergence of classicality will be discussed in section 2.3.

---

[14] I will call the preference for the correlated (i.e. single-sum) decomposition 'pragmatic' in order to emphasize that other decompositions are always formally possible, but that this simply does not yield the information we are looking for. Although one's preference for the correlated (in particular, bi-orthogonal) form is pragmatic, preference for one of all possible correlated decompositions is not. See section 3.3.1 in the next chapter for a more extensive discussion of this issue.

[15] Almost. Cf. footnote 13.

Zurek's solution of the preferred basis problem

How the environment fixes the pointer states of the apparatus and the eigenstates of the measured observable of the quantum system can be seen fairly easily. The solution of the "preferred decomposition" problem is essentially the introduction of a third subsystem.

In chapter 1 I showed that a degenerate bi-orthogonal state on a Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$ could be decomposed in a (possibly infinite) number of equivalent ways. However, if one introduces a third system (the "environment") with Hilbert space $\mathcal{H}_3$ and orthogonal basis vectors by $|E_\uparrow\rangle, |E_\downarrow\rangle$, things are quite different. (See also Zurek (1981).) Let us assume we have

$$|\Psi\rangle = \frac{1}{\sqrt{2}} \left(|\uparrow\rangle |\Uparrow\rangle |E_\uparrow\rangle + |\downarrow\rangle |\Downarrow\rangle |E_\downarrow\rangle \right). \tag{2.16}$$

If we now try to rewrite this state in the $|\rightarrow\rangle, |\leftarrow\rangle$ basis, the result is

$$|\Psi\rangle = \frac{1}{\sqrt{2}} \left(|\rightarrow\rangle \left(|\Uparrow\rangle |E_\uparrow\rangle + |\Downarrow\rangle |E_\downarrow\rangle \right) + |\leftarrow\rangle \left(|\Uparrow\rangle |E_\uparrow\rangle - |\Downarrow\rangle |E_\downarrow\rangle \right) \right). \tag{2.17}$$

Note that this cannot be written in the form of (2.16). In order to do so, the entangled states on $\mathcal{H}_2 \otimes \mathcal{H}_3$ between the parentheses would have to be expressed as product states, which is generally not possible. So by introducing a third system, the basis ambiguity is removed: rewriting the state of one of the subsystems in a different basis comes at the expense of a more complicated expression for the total state on $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3$ which does not represent a one-to-one correlation between, for instance, the quantum system and the measurement apparatus. In other words: if one insists on this particular form, there is only one option, namely (2.16). In analogy with the bi-orthogonal decomposition, I call the special form (2.16) of the wave function Ψ the "tri-orthogonal" decomposition. The argument that the tri-orthogonal decomposition is unique (unlike the bi-orthogonal decomposition) has been worked out into a mathematical proof by Elby and Bub (1994)[16]. I will come back to the tri-decompositional uniqueness theorem and its relevance for the (decomposition version of the) Preferred Basis problem in the next chapter.

Zurek (1981) now shows that the correlations between the eigenstates of the quantum system and the apparatus can only remain intact under the evolution induced by the environment if the apparatus-environment interaction Hamiltonian commutes with the pointer observable. (In other words: it is this interaction that establishes the desired form of the $\mathcal{SAE}$-state (2.16).) Put the other way around: *given* a particular interaction Hamiltonian $\hat{H}_{\mathcal{AE}}$, only eigenstates of an observable $\hat{A}$ with which it commutes can be used to establish *stable* correlations between system and apparatus in the state (2.3). (In Zurek's words: they can be used to establish a "reliable record".) The observable $\hat{A}$ that satisfies the stability criterion[17]

$$\left[\hat{H}_{\mathcal{AE}}, \hat{A}\right] = 0, \tag{2.18}$$

---

[16] As is clear from my sketchy argument, it is not essential that the states of the environment and apparatus are orthogonal. To be precise, Elby and Bub prove a "n-decompositional uniqueness theorem" ($n \geq 3$) which requires that two of n sets of vectors are linearly independent, and one is non-collinear.

[17] In Zurek (1998) the condition (2.18) has become $\left[\hat{H}_{\mathcal{AE}} + \hat{H}_{\mathcal{A}}, \hat{A}\right] = 0$

is therefore the one and only pointer observable. Next, Zurek discusses a simple model (essentially the model described in section 2.2.1 above, but this time with an environment consisting of a single atom[18]) to show how the final $\mathcal{SAE}$-state attains the desired form (2.3) for the apparatus (pointer) states that commute with $\hat{H}_{\mathcal{AE}}$. But, as I showed above, if one tries to re-express the final state (2.3) in another basis for the apparatus, this will not a decomposition in which the pointer states are correlated one-to-one with the system states.

In the model discussed by Zurek (1981) the apparatus-environment interaction Hamiltonian is diagonal in the pointer states. Hence, the criterion (2.18) is trivially satisfied. In more sophisticated models of decoherence however, there are no observables that commute exactly with the apparatus-environment Hamiltonian and hence the criterion (2.18) cannot be used to select the pointer observable. For instance, pointer states may be subject to internal dynamics due to a self-Hamiltonian $\mathcal{H}_{\mathcal{A}}$ that does not preserve the commutativity condition (Zurek, 1998). So another criterion is needed to select those states that are the least affected by the interaction with the environment. Zurek calls this criterion the "predictability sieve":

> "In order to settle this question, we shall imagine testing *all* the states in the Hilbert space of the system $\mathcal{H}_{\mathcal{S}}$. To this end, we shall prepare the system in every conceivable pure initial state, let it evolve for a fixed amount of time, and then determine the resulting final state – which, because of the interaction with the environment, will be nearly always given by a mixture." (1993, p.294)

The content of this statement will be made more precise in the next section and in section 3.2 of chapter 3.

The criterion by which the final states are ordered after this fixed amount of time has elapsed[19] is their (von Neumann) entropy:

$$H_\rho = -Tr(\rho)\ln\rho \qquad \in [0,\infty), \tag{2.19}$$

in which ρ is the density matrix that has evolved from the initial 'test' state. These initial states, which are pure (and hence have zero entropy), are ordered according to the entropy of the resulting final state, with the states that show the least increase of entropy on top of the list. Since the entropy of the final density matrix "is a convenient measure of the loss of predictability" the states on top of the list are "the best candidates for the classical "preferred states"(Zurek, 1993, p. 294). The von Neumann entropy can involve a somewhat tedious calculation, therefore it is often replaced by the purity (which however must be maximized instead of minimized):

$$\varsigma_\rho = Tr\rho^2 \qquad \in [0,1] \tag{2.20}$$

The operator ρ that appears in these formulas should not be confused with the reduced density matrix $\rho_{\mathcal{SA}}$ in (2.15). The purpose of the stability criteria is

---

[18] Zurek has to make a specific and rather awkward assumption about the interaction time to accomplish the perfect correlation between the system-apparatus state and orthogonal environment states because of this. Of course, an infinite environment remains essential to arrive at the more general result (2.10).

[19] Zurek (1993) also mentions the possibility of ordering the initial states with respect to the time it takes to evolve into a final state with fixed entropy. But he expects that "the two versions of the sieve should yield similar sets of preferred states."

to "measure the loss of predictability caused by evolution for every pure state" ((Zurek, 2003, p.736), see also Zurek (1993). The ρ that enter the sieve are thus arbitrary pure (in a generalized version (Zurek, 1993) also mixed) states.

When the stable states thus selected constitute a complete orthogonal basis this defines the pointer observable. If an exact pointer observable exist, (i.e. an observable satisfying (2.18)) its eigenstates naturally appear on top of the list, but in most cases there is no such exact pointer observable and there are no absolutely stable states, i.e. those maintaining zero entropy. Moreover, the states selected by the predictability sieve will often form an overcomplete basis (such as coherent states, cf. section 2.3 below) and therefore do not define an eigenbasis of an Hermitian observable[20]. Zurek therefore focuses on the states selected by the sieve, rather than the observables. Yet, if perfectly stable states exist, these would have to correspond to our *a priori* notion of the basis of "classical" pointer states that do not superpose. A large part of Zurek's program consists of verifying the classical properties of states on top of the resulting 'predictability hierarchy'. However, it is not yet certain whether the different criteria of 'predictability' yield the same pointer states (Zurek, 2003)[21].

So Zurek's point is that the pointer basis is completely determined by the form of the apparatus-environment interaction Hamiltonian. As a result, the space of 'classical' observables is restricted by the available interaction Hamiltonians. Joos argues: "any interaction described by a potential $V(r)$ is diagonal in position, hence *position* is always the distinguished "observable" measured by the interaction" (Joos et al., 2003, p.55). Thus, we have an argument why position is the determinate property of our observation, also in typical quantum experiments. (See also Schlosshauer (2004); Zurek (2003)) According to this view, position as a determinate property is no longer an *a priori* requirement for observation but a contingent physical fact regarding our universe.

2.2.3 Environment-induced superselection rules

I have been talking several times of 'environment-induced superselection rules' (or 'einselection') without exactly stating what this phrase means. In Joos et al. (2003), Zeh elucidates and warns for a common misinterpretation:

> "An environment-induced superselection rule means that certain superpositions are highly unstable with respect to decoherence. It is then impossible in practice to construct measurement devices for them. This *empirical* situation has led some physicists to *deny the existence* of these superpositions and their corresponding observables – either by postulate

---

[20] Any overcomplete set *does* generate an algebra of 'quasiclassical' observables, however, but Zurek does not develop his approach in this direction.

[21] Zurek (2003) also remarks that besides predictability, there are other possible criteria for classicality such as distinguishability or redundancy (which plays a crucial role in his "Quantum Darwinism" program, see chapter 4), and that it remains to be seen whether these yield the same pointer observables. In Dalvit et al. (2005) four different criteria are investigated for the case of an underdamped harmonic oscillator and for quantum Brownian motion. The authors find that "different robustness criteria are optimized by different measurement strategies. Yet, the *states* resulting from all of the above strategies are essentially identical. Thus, even when environment-induced superselection does not pick out unique pointer states, different criteria still agree on what appears to be the most classical" (p.9).

> or by formal manipulations of dubious interpretation, often including infinities. In an attempt to circumvent the measurement problem [...] they often simply *regard* such superpositions as "mixtures" once they have formed according to the Schrödinger equation." (p. 19)

Despite these cautious remarks, the idea that environment-induced superselection rules could solve the measurement problem motivated the early formulations of decoherence (Zeh, 1971; Zurek, 1982). Here, I will give a brief overview of the formalism of superselection rules, spell out how Zurek's approach may fit into this mathematical scheme and put the abstract formalism in an empirical context.

The formalism of superselection rules

The basic idea of einselection is that combining the two main consequences of decoherence – the diagonality of the density matrix, and the dynamical robustness of the corresponding pointer basis – results in effective superselection rules. These superselection rules restrict the vast Hilbert space of possible quantum states to a small subset of empirically accessible states. (As Zeh mentions this has no ontological consequences.) In this sense, the approach of decoherence deviates from other approaches to superselection rules where certain superpositions are ruled out formally. The explanation given by decoherence leads to a conception of superselection rules as inexact and not fundamental, but as long as this dynamical conception of superselection rules applies to a different domain than the one to which the strict algebraic superselection rules are supposed to apply, this distinction poses no serious conceptual challenge.

Originally, exact superselection rules were introduced to account for certain conserved quantities in quantum mechanics, such as charge and the boson/fermion nature of particles. They can be derived from *a priori* locality or symmetry principles in algebraic field theory, but only in the limit of infinite systems. (Although mathematical arguments are often based on idealizations like this, it is difficult to see how to connect this formalism to the physical world: mathematically, these superselection rules simply cannot *exist* for finite systems.) In algebraic quantum mechanics, on the other hand, the superselection rule itself – or rather the existence of a non-trivial center $\mathcal{C}$ of the *algebra of observables* $\mathcal{O}$ – from which the superselection rule follows, is postulated.[22] The algebra $\mathcal{O}$ of observables is supposed to be some subset of the algebra of self-adjoint operators. (If these coincide, no superselection rules can exist.) The observables in the center $\mathcal{C}$ are called "superselection observables" or *"classical observables"*. Now every projection $P_{\mathcal{C}}$ in $\mathcal{C}$ defines a subspace $\mathcal{H}_{\mathcal{C}}$ of the Hilbert space $\mathcal{H}$ which is invariant under the action of any operator in $\mathcal{O}$. It follows that (in the discrete case) the Hilbert space is a direct sum of these invariant subspaces:

$$\mathcal{H} = \bigoplus_i \mathcal{H}^{\mathcal{C}}{}_i \tag{2.21}$$

This also implies that *for all observables* $\hat{O} \in \mathcal{O}$ the expectation value in a pure state equals that of an incoherent mixture of superselected states. Moreover, in the

---

[22] Although in a different formulation, one postulates the superselection rule itself (i.e. the truncation of the Hilbert space (2.21) and derives the existence of an abelian subalgebra of classical observables. The two approaches are equivalent.

framework of superselection rules the decomposition of such a mixture is unique. The observables in the center $\mathcal{C}$ can be regarded as being classical in the sense that they have well-defined values for the admissible states, i.e. they share a common basis of eigenvectors by definition (Landsman, 1995).

Attractive as this treatment may seem, it has one major drawback: the algebra of observables (i.e. the restriction to a subset of the algebra of self-adjoint observables) has to be constructed 'by hand'. There are no *a priori* reasons why some self-adjoint operators should count as observables, and others don't, even though the choice is far from arbitrary. The formalism has to take recourse to experience: according to Landsman "the truncation of the original set [...] of beables to a (much) smaller set of observables [...] is made by the 'user' who normally has little choice in doing so." (*Ibid.*, p. 50)

Environment-induced superselection rules

To derive superselection rules from very general principles requires the idealisation of infinite systems. Since actual physical systems are finite, a decoherence-based argument will therefore never yield the exact superselection rules of algebraic quantum mechanics. What decoherence arguments pursue, therefore, is a dynamical theory of 'approximate' superselection rules.[23] The price one pays for this, is that the classical world emerges only in an approximate sense as well –a viewpoint explicitly endorsed by the proponents of decoherence (cf. section 3.1). But there are obvious benefits: such a dynamical argument may be able to explain why superselection rules only hold for the macroscopic domain and secures the universal validity of the superposition principle.

In the discussion of the measurement problem at the start of this section I already presented a model to show how decoherence results in the approximate diagonality of the density matrix in the preferred basis. In the same paper Zurek (1982) generalizes this model to point out the connection with superselection rules more clearly:

> "We shall show that as a result of such interactions the state vector of the system is able to remain pure only if it is completely confined to one of the subspaces $\mathcal{H}_n$ [of the Hilbert space of the considered system]. Arbitrary superpositions, with components belonging to two or more subspaces, shall decay into mixtures diagonal in the state vectors belonging to the separate subspaces. [...] Moreover, [...] the set of the observables that can be measured [...] is limited to those that leave subspaces invariant. Systems which exclude the existence of certain pure states and which restrict the class of possible observables in a manner described above are said to obey superselection rules." (p. 1867-1868)

In Zurek's approach, superselection sectors arise naturally as a result of the interaction between the system and its environment. He employs a fairly general model to

---

[23] The mathematical meaning of this notion would have to be made more precise, but unfortunately the view most decoherence theorists hold of superselection rules is rather intuitive and imprecise. For a discussion of superselection rules in the context of decoherence with a particular eye on the mathematical and philosophical implications, see (Landsman, 1995).

show how the dynamics of decoherence allows an interpretation which is formally identical (more or less) to the algebraic approach. In summary, his argument runs as follows:

First, the Hamiltonian that generates the evolution on the combined Hilbert space $\mathcal{H} = \mathcal{H}_{\mathcal{SA}} \otimes \mathcal{H}_{\mathcal{E}}$ is written down as a tree-part operator $H = \hat{H}^{\mathcal{SA}} + \hat{H}^{\mathcal{E}} + \hat{H}_0^{\mathcal{SE}}$ with $\hat{H}^{\mathcal{E}}$ and $\hat{H}^{\mathcal{SA}}$ the self-Hamiltonians of the environment and the system-apparatus combination, respectively, and $\hat{H}_0^{\mathcal{SE}}$ the "diagonal" part of the interaction Hamiltonian (i.e. diagonal in the corresponding subspaces).[24] (The off-diagonal part, with terms like $|sa_i\rangle\langle sa_{i'}| \otimes |e_j\rangle\langle e_{j'}|$, is assumed to be negligible.)

$$H_0^{\mathcal{SE}} = \sum_{i,j} \gamma_{ij} |sa_i\rangle\langle sa_i| \otimes |e_j\rangle\langle e_j|. \tag{2.22}$$

Note that this is a generalization of eq. (2.7) but with the total environment expressed as a single subsystem. Following the same procedure as before (including the partial trace and the limits $N \to \infty$ and $t \to \infty$, in which N is the number of subsystems in the environment) Zurek shows that correlations between states that correspond to different eigenvalues of $H_0^{\mathcal{SE}}$ are effectively 'damped out'. Eigenvalues will in most cases be highly degenerate, and therefore the corresponding eigenstates will form coherent subspaces $\mathcal{H}_n$, block-diagonalizing $\hat{H}_0^{\mathcal{SE}}$. As a consequence, one has

$$\mathcal{H}_{\mathcal{S}} = \bigoplus_n \mathcal{H}_n. \tag{2.23}$$

Moreover, as long as the interaction with the environment remains stronger than the coupling to some apparatus, Zurek argues "Only observables which leave every $\mathcal{H}_n$ invariant are admitted: $\hat{A}$ is an observable on a system $\mathcal{S}$ interacting with the environment $\mathcal{E}$ if and only if $|\psi_n\rangle \in \mathcal{H}_n \Rightarrow \hat{A}|\psi_n\rangle \in \mathcal{H}_n$." This way, formal equivalence to superselection rules is achieved.

Secondly, Zurek discusses how his criterion for the existence of a preferred pointer observable is a direct consequence of this general scheme. The pointer observable $\hat{\Lambda}$ is defined as "any observable measurement which allows us to precisely determine the subspace $\mathcal{H}_n$ containing the state of the system." Thus, the pointer observable can be written as

$$\hat{\Lambda} = \sum_n \lambda_n \hat{P}_n, \tag{2.24}$$

where $\hat{P}_n$ project on the respective subspaces. Accordingly, the pointer observable commutes with the interaction Hamiltonian

$$[\hat{\Lambda}, H_0^{\mathcal{SE}}] = 0, \tag{2.25}$$

which means that $\hat{\Lambda}$ and $H_0^{\mathcal{SE}}$ are simultaneously measurable. In other words, only eigenstates of the pointer observable (2.24) will not be perturbed by the 'continuous measurement' by the environment. Zurek therefore claims that the environment is capable of suppressing superpositions of macroscopically distinguishable states $|sa_i\rangle$. (In section 3.2 in the next chapter I will make the meaning of this claim more precise.)

---

[24] Zurek uses $\mathcal{S}$ rather than $\mathcal{SA}$ for notational simplicity. For reasons that will become clear in the next chapter I do not follow this convention. I write $|sa_i\rangle$ to indicate a state on $\mathcal{H}_{\mathcal{S}} \otimes \mathcal{H}_{\mathcal{A}}$ without a preferred decomposition.

Although formally similar, decoherence does not lead to the exact superselection rules of algebraic quantum mechanics. In the case of position this arguably is no serious objection, but the success of decoherence in this respect has led some authors to claim a prominent role for decoherence effects in the superselection of other (kinematically) conserved quantities (such as charge and spin) as well. Nowhere has it really been spelled out clearly how the environment-induced superselection rules relate to the algebraic approach, but both Joos and Giulini (Joos et al., 2003) hint at the possibility that symmetry principles are only relative, and based on classical prejudice[25].

As will be discussed in the next section, 'environment-induced superselection rules' are often considered to be the key to understanding the appearance of a classical realm. From that point of view, pointers of measurement apparatuses constitute just a subclass of the kind of systems decoherence models aim to describe. As pointed out in the previous subsection, in most cases the einselected states will be position eigenstates. Most theoretical models of decoherence indeed show that wave functions of objects in interaction with an environment rapidly deform and become strongly peaked in position (in phase space) for the local state described by the reduced density matrix[26].

The approximate superselection rules generated through the interaction with the environment constitute a welcome dynamical counterpart to the kinematical formalism of exact superselection rules. However, being only approximate, in this case superselection does not guarantee the uniqueness of the decomposition of the density matrix. This is why Zurek needs to invoke an additional criterion to establish a preferred basis.

Both the selection of a preferred basis and the concept of dynamical superselection rules are generally believed to be essential ingredients in a full solution of the measurement problem. The next aim would be to solve not (merely) the measurement problem, but to derive the appearance of the classical realm as a whole from environmental interactions.

2.3 Decoherence and the emergence of classicality

2.3.1 Robustness and decoherence as a source of classicality.

As I have pointed out in the previous chapter, there is something peculiar about trying to derive the appearance of classicality from the quantum formalism if the problem of Outcomes is left unsolved; after all, this implies that we cannot account for the "definite facts" of our experience, and thus it seems that there is no room for systems to possess definite properties in the quantum formalism. Indeed, as I

---

[25] Giulini quite explicitly holds the view that dynamical laws are prior to kinematical concepts. For instance, in a footnote on p. 297 in (Joos et al. 2003) he says: "The 'right' kinematical implementation of a symmetry group eventually relies on fundamental dynamical laws with respect to which it becomes a *dynamical* symmetry group. Hence, fundamentally speaking, there is no kinematical symmetry (in the sense of being independent of dynamical laws)".

[26] By the uncertainty relations, the degree of localization is not unbounded: the wave function (interpreted as probability distribution) will never become equal to the classical ideal of a single point in phase space. The best decoherence can (and in many cases does) achieve are minimum-uncertainty wave packets, i.e. coherent states. See also section 2.3.3

will explain in more detail in section 3.1 in the next chapter, what one –at most– can gain by a decoherence-based explanation of classicality is the mere *appearance* of properties in a classical sense at the macrolevel. Whether this position is philosophically tenable remains to be seen. However, from the physicist's point of view it is perfectly sensible to speak of *effectively* classical behavior without worrying about the underlying ontology. This is the stance I will take in this section.

Zurek selects a special class of to-be-classical states by means of the stability criteria. This also suggests a solution to an apparent paradox: how could environment-induced decoherence ever explain the emergence of the laws of classical physics, if the application of the latter requires (and harmlessly admits) perfectly isolated systems? The standard answer is that einselected *after* decoherence has done its job. Since decoherence is a very effective process, taking place at extremely short time scales, it is claimed that one is justified in *treating* a physical system in an 'einselected' state as perfectly isolated, although it is essential for the laws of classical physics to be (approximately) valid, that is it is in fact not *physically* isolated at all.

Having thus dealt with the first obstacle on the road to classicality, expectations of where it leads are high. Environment-induced decoherence is considered by many physicists to be the key to an understanding of classicality within the quantum formalism, and the most reliable guide to the 'No Man's Land' in between (Anglin et al., 1996). Within the framework of decoherence, there are two kinds of effects that are assumed to be primarily responsible for the emergence of classicality. On the one hand there is the "delocalization" of phase relations, on the other the stability or "robustness" of the pointer states.

In Zurek's programme, the shift of interest from the measurement problem to the emergence of classicality has the consequence that the stability criteria (2.18), (2.19) and (2.20) are also supposed to apply to a single system, instead of the system-apparatus couple, in interaction with its environment[27]. His aim is to show that these criteria (perhaps for particular models) select states that behave, in a certain sense, effectively classical.

2.3.2 Reducing classical physics to quantum mechanics.

Despite the fundamental relevance of the subject, there is no clear-cut definition of what 'classicality' exactly is. Even if there were such a definition, it would only be approximately valid, and would force us to abandon the notion of a truly classical world. This is of course well known from other cases of intertheoretical reduction[28].

But there are more difficulties with the reduction of classical physics to quantum mechanics than this[29]. First of all, it is not clear what the appropriate limit should

---

[27] See section 3.3.1 for a critical assessment of this move.

[28] The recovery of Newtonian mechanics from special relativity, in the limit of small velocities is of course a paradigm example, but it must be noted that this is an exceptional case where a simple account of intertheoretical reduction indeed seems to be fruitful. In general however it is much less obvious how theories reduce to one another, from a philosophical as well as from a physical viewpoint. I will not dwell on the problem of intertheoretical reduction here. A comprehensive point of departure for the interested reader is Batterman (2007).

[29] I say that classical physics reduces to quantum mechanics in the sense that, ideally, the laws of classical physics would be derivable from those of quantum mechanics, i.e. quantum mechanics has a kind of theoretical priority over classical physics. Mathematically speaking, one might as

be, i.e. $\hbar \to 0$ or $N \to \infty$. (As a matter of fact, from a formal point of view the latter is a special case of the former, cf. (Landsman, 2007)). These alternatives presuppose a kind of explanation for the reduction of one theory to another. But one cannot simply assume that classical physics follows as the systems involved become large. In fact, the failure of this simplified picture has led physicists to investigate considerably more complicated mechanisms, such as environment-induced decoherence. By the same token, the laws of classical physics cannot simply be recovered from quantum mechanics in the $\hbar \to 0$ limit, even though this impression is often made in the textbooks. In particular, additional assumptions about the relevant, "classical" observables have to be made in order to develop a mathematically rigorous framework for the quantum-to-classical transition (Landsman, 2007).

From a conceptual point of view the connection between the abstract formalism of quantum mechanics and the familiar concepts of classical physics is not always easy to grasp. The mathematical approach to the emergence of classicality might come to rescue here, but the physical interpretation of this formalism is in need for further clarification. (For an overview, see Landsman (2007).) The obscurity surrounding the emergence of classicality is reflected in the way the notion of 'classicality' is treated in the literature. As we will see below, proponents of decoherence can sometimes be sloppy in their claims. Also, I will assume that everyone has a pretty clear notion of what the classical world looks like (after all, we live in it) and will not bother too much about the lack of an all-encompassing definition. The need for such an intuitive treatment should not be surprising, considering the great divergence of concepts and physical models involved.

2.3.3 Decoherence and the emergence of classicality

I will now discuss the arguments by which decoherence aims to explain the emergence of classicality, in a more or less hierarchical order; I will start with the most fundamental algebraic level and gradually work my way up to the dynamics of "to-be-classical" physical systems.

The algebra of observables

In a certain sense, the heart of the measurement problem (the problem of outcomes) resides in the logical structure of quantum mechanics; quantum mechanics (in its standard formulation) fails to ascribe definite properties or definite values of observables (for all states) to a physical system. Nevertheless, if one is willing to make some sacrifices it is possible to introduce classical "definiteness" at the algebraic level. Such a sacrifice could, for instance, be to no longer demand that *all* observables should possess definite values. In that case, the Kochen-Specker theorem no longer applies[30]. So if the theory of decoherence were able to show exactly *why* certain operators are to be regarded as "definite" or "meaningful" observables, and others are not, this could perhaps complete the dynamical picture

---

well say that (a certain aspect of) quantum mechanics reduces to classical physics, say, in the limit $\hbar \to 0$ or $N \to \infty$. To avoid confusion, I will adopt the first meaning throughout this and the next chapter, since it is more at home with my central point: that decoherence assumes the mere approximate validity of classical physics. See Batterman (2007) for a brief comparative account of these two opposing viewpoints.

[30] Bub (1997) is a thorough elaboration of this idea.

that decoherence provides with an exact and rigorous mathematical framework.

That decoherence could possibly define what the classical observables are, is sometimes suggested in the literature. I will elaborate on this in section 3.3.1 of the next chapter, but basically, the idea is to take the preferred basis selected by the stability criteria to *define* the observable of interest. That is, if the projectors onto the stable states constitute the spectral decomposition of an operator, than that operator is defined to be an observable. However, generally the states picked out by the 'predictability sieve' do not constitute an orthogonal basis and hence do not constitute the spectral resolution of an Hermitian observable, although the latter may be replaced by positive operator-valued measures (POVM's). (Indeed, this seems to be the motivation for Zurek's speculations to abandon the notion of 'observables' altogether and focus on the possible states of the system alone.) I think this idea could prove to be fruitful, if rigorously worked out (both mathematically and conceptually). I will come back to this in the next chapter.

Coherent states

Related to the idea that decoherence leads to the selection of a certain set of "classical" observables, is that environmental interactions lead to effectively classical states (in a certain limit). As we have seen, a central argument in Zurek's approach to the emergence of classicality is that only certain 'preferred' states will be stable with respect to the decohering effect of the environment. These restrictions on the allowed states, in turn, may point at the connection between the Hilbert space formalism and the phase space structure of classical physics.

Quantum states on Hilbert space that are the analogue of classical states in phase space are the so-called coherent states. Coherent states are "minimum-uncertainty wave packets", i.e. under the bounds of the Heisenberg uncertainty relations they display the least spread of position and momentum. They were first introduced by Schrödinger, who recognized their importance for the quantum-to-classical transition. In the classical limit (taking $\hbar \rightarrow 0$) they have two essential properties: first, the expectation values of quasiclassical observables in coherent states become equal to the classical expectation value in the corresponding point in phase space. Second, coherent states corresponding to different phase-space points become orthogonal, which is in accordance with the idea that classical pure states (in the absence of dynamical evolution) have zero transition probability (Landsman, 2007).

The role of decoherence now is to select (through the predictability sieve, for a derivation see Anglin et al. (1996)) precisely the coherent states as the preferred states in a model of quantum Brownian motion (a rather general model applicable to a wide class of systems, describing a system immersed in a heat-bath). Zurek (1998) explains this result as follows[31]:

> "An example of this situation is afforded by a harmonic oscillator, where the dynamical evolution periodically "swaps" the state vector between its position and momentum representation, and the two representations

---

[31] I will ignore the fact that the role of the environment for this claim is a bit puzzling: that the states resulting from the dynamical evolution are related by a Fourier transform implies that the evolution of the system is unitary – which is obviously not the case if it is interacting with its environment.

> are related to each other by a Fourier transformation. In that case the states which are most immune to decoherence in the long run turn out to be the fixed points of the "map" defined by the Fourier transformation. Gaussians are the fixed points of the Fourier transformation (they remain Gaussian). Hence, coherent states which are unchanged by the Fourier transform are favored by decoherence" (p. 7)

The fixed points of the Fourier transformation are selected by the predictability ieve because they minimize entropy production; stable points are the most favorable 'compromise' since "a near-zero entropy production at one instant may be "paid for" by an enormous entropy production rate a while later." (*Ibid.*) The relevance of this result lies in the fact that coherent states are the most 'classical' of all quantum states. This is what Zurek refers to when he says that "states selected by the predictability sieve become phase-space 'points' " (Zurek 2003). Although there is no reason to take this result to be universal, it provides a nice illustration of the relevance a dynamical process might have for the kinematical aspects of the quantum-to-classical transition.

Macroscopic superpositions

Another aspect to the kinematical structure of classical systems which immediately comes to mind when discussing decoherence is the invalidation of the superposition principle. This is generally regarded to be especially relevant for the measurement problem, but environment-induced superselection is claimed to apply to a much more general class of physical systems. A good example of one of the alleged successes of decoherence for the emergence of classicality is that it would yield an explanation for the chirality superselection rule for molecules. Chirality is a localized property; optically active molecules have a left- or right-handed orientation, which do not correspond to energy-eigenstates. Very small micro-objects occur in energy- and angular-momentum eigenstates that are a superposition of chirality eigenstates (in other words, they do not have a "definite shape"), but the border between these domains is to be put far in the microscopic regime. The molecules that are in energy-eigenstates oscillate between chirality eigenstates with a frequency proportional to the energy difference between ground state and first excited state (Joos et al., 2003). A crude estimate yields a decoherence timescale inversely proportional to the square of the size of the molecule. Comparing this to the typical transition time between chirality eigenstates, which is inversely proportional to the energy difference, we see that the decoherence time for most molecules is much shorter. Hence, it is said that the molecule is stabilized in one of the two chirality eigenstates before oscillations can occur.

However, despite this intuitively appealing argument there is still no definite explanation of the chirality superselection rule available. Amman (1991) points out that the model of the molecule coupled to an infinitely large oscillator bath (boson environment) contains major approximations and can only be solved in the idealized case of a two-level system (the ground state and some exited state of the atom.) He concludes:

> "The problem of chirality can then be discussed as a problem of stability: superpositions of chiral states seem to be unstable under outer

> perturbations. Unfortunately, this conception has yet the status of a conjecture. For none of the discussed environments have superselection rules been derived rigorously." (p. 13)

It seems that the claims of the decoherence theorists that they can explain the chirality superselection rule are a bit too optimistic. Nevertheless, it is likely that the environmental interactions play a crucial role in an explanation of this phenomenon.

But it is not at the molecular level where decoherence attains its fundamental importance; after all, we set out to explain the classical appearance of the world. There is of course no exact decoherence-based model available to derive environment-induced superselection rules for all of the classical world; rather, there are specific idealized models and the general claim that, according to Zurek's simple stability criterion (2.18), macroscopic systems are approximately localized position eigenstates, since interaction Hamiltonians generically depend on position. (As discussed above, in some models coherent states are obtained, which are clearly not eigenstates of position. In fact, most environmental interactions also remove coherence between momentum eigenstates, albeit less effectively. See for instance (Joos et al., 2003, sect.3.2).) Since all measurements are eventually position measurements, this is all that is required. Or so it is argued.

Even disregarding the specific weaknesses of this argument (which I will discuss in the next chapter), it is first of all sensitive to the objections that were raised in section 1.2 of the previous chapter. Almost all discussions of environment-induced *classicality* focus on the vanishing of interference terms, but without referring to the well-known problems of measurement anymore. This obscures, but does not resolve, the main objection that the mere vanishing of interference terms does not imply definiteness.

A good example of this kind of reasoning is found in discussions of the Wigner function and the classical ($\hbar \to 0$) limit. (For a brief treatment of the Wigner function representation and its relevance for decoherence, see for instance Zurek (2003) or the book by Joos et al. (2003). The Wigner function of the density matrix ρ is given by the following expression:

$$W(x,p) = \frac{1}{\pi\hbar}\int_{-\infty}^{\infty} dy\ e^{2ipy/\hbar}\rho(x-y,x+y)$$

The Wigner distribution closely resembles a classical distribution function in phase space, which makes it a convenient tool for describing (approximate) classical correlations within quantum mechanics. However, the Wigner function cannot really represent a probability distribution, since it can assume negative values[32]. The result of decoherence is that oscillations in the Wigner function, representing interference terms, get "smeared out" and only the peaks corresponding to the eigenstates of the monitored observable remain. In extreme cases, decoherence can make the distribution non-negative.

However, this does not warrant a probability interpretation (contrary to what for instance Omnès (2003) suggests). In most cases of decoherence, its positivity is only approximate (and I would say that the *conceptual* identification of a function as probability distribution cannot depend on such contingent, dynamical effects).

---

[32] A problem which, however, can easily be circumvented by adopting the Husimi (or Berezin) function instead of the Wigner distribution, see (Landsman, 2007)

Pictorial representations of decoherence in the Wigner-function representation often seem to show the kind of (quasi-) classical characteristics of the system we are looking for, but as long as the Wigner function has no clear-cut interpretation, one should be careful in identifying these appealing illustrations with representations of our classical intuitions -something like 'probability distributions on phase space' for instance. Moreover, the only reason that the Wigner function displays this kind of behavior if subject to environmental monitoring, is exactly the same as for the vanishing of interference terms from the density matrix: namely the tracing out of the environment. The conceptual problems with this kind of reasoning do not simply disappear in a different representation. Instead, the most important contribution the Wigner-function representation of decoherence makes, is mathematical convenience: it facilitates the implementation of the classical limit $h \to 0$. Whereas this limit is hard to handle in general, it is much better behaved for the decohered Wigner function (Zurek, 2003).

Dynamics

Although the collapse postulate was introduced to secure the definiteness of properties at the macrolevel, it is in direct conflict with the classical, deterministic evolution of macroscopic states. For it is essentially the collapse postulate which introduces indeterminism in the evolution of the quantum state. Moreover, if one assumes with the decoherence theorists that the environment is continuously 'measuring' the system, the effect would have to be omnipresent. States of systems interacting with their environments would therefore be continuously 'collapsed' into some eigenstate of the measurement observable, causing a complete freezing of the motion.[33] So, although the projection postulate has been invented to explain classical measurement records in quantum experiments, in this case it would lead to plain contradiction with the classical situation if it were to apply to all physical systems. For quantum systems, however, this effect is known to exist and is called the 'quantum Zeno effect' (Joos et al., 2003, sect. 3.3.1). The existence of the Zeno effect shows that decoherence cannot (and should not) replace phenomenological collapse in all respects. What is needed is an explanation of why the Zeno effect typically does not occur for macroscopic (classical) systems. At first sight, although it abandons the collapse postulate, decoherence seems to imply a Zeno effect as well (Joos et al., 2003; Joos, 2005). The rate of change of the diagonal elements $\rho_{\alpha\alpha}$ (that Joos calls the "properties") of the density matrix of a system is given by the von Neumann equation:

$$i\frac{\partial \rho}{\partial t} = [\hat{H}_{int}, \rho] = \sum_{\beta}(H_{\beta\alpha}\rho_{\alpha\beta} - H_{\alpha\beta}\rho_{\beta\alpha}) \tag{2.26}$$

which is equal to zero if the interference terms vanish. Hence the "freezing of the dynamics". The intuitive picture is that due to the superselection rule, states cannot evolve unitarily into one another since the intermediate states are not accessible. However, the explanation decoherence offers for this phenomenon is more subtle, and is claimed to be able to account (at least qualitatively) for the differences between physical systems regarding their Zeno-like behavior. In fact, it is claimed that the decoherence models explain the existence of an exponential decay law for radioactive systems (which cannot be exact for an *isolated* quantum system).

---

[33] This is intuitively clear, but can also be proven fairly easily. See (Joos et al., 2003).

Detailed models show that for strong coupling to the environment, there is indeed a Zeno effect, but for smaller couplings the effect is absent (or even reversed, resulting in the so-called 'anti-Zeno effect', meaning that measurements actually increase the probability of state transitions). Strong coupling is said to be typical for few-state systems (the quantum case). Large systems are much less sensitive to the coupling to the environment and hence show no Zeno behavior:

> "These models show, that the reason for the Zeno effect lies in the delocalization of phases needed for the natural (isolated) evolution of a system. [...] Systems with only a few states are typical "quantum systems". It is therefore not surprising that they are very sensitive to being monitored by other systems." (Joos et al., 2003, pp. 124-125).

This seems a bit puzzling. Suddenly it is the *microscopic* system which is "sensitive to monitoring by other systems". To understand what is going on here, one must keep in mind the distinction between measurements by external systems and by the environment. The Zeno effect is an effect induced by *real* measurements. It can be shown to be *absent* when the system also interacts with an environment, as long as this interaction is not too strong. Thus, the usual claims in the literature that environmental interactions are analogous to measurements should not be taken literally.

The Zeno effect is a good example of the physical relevance of environment-induced decoherence for explaining the differences between typical quantum systems and systems interacting with their natural environment. As an account of the emergence of classicality, however, it seems of limited value. Nevertheless, stronger claims in this direction can also be encountered in the literature. For instance, it is claimed that "Newton's reversible laws of motion can be derived (to a very good approximation) from strong *irreversible* decoherence" (Joos, 2005, p.9), and Zurek (2003) says that:

> Einselection is responsible for the classical structure of phase space. States selected by the predictability sieve become phase-space "points", and their time-ordered sequences turn into trajectories. In underdamped, classically regular systems one can recover this phase-space structure along with (almost) reversible evolution. (p. 737)

When saying that "einselection is responsible for the classical structure of phase space", Zurek is referring to the selection of coherent states as the preferred states, as I discussed before. But Zurek's claim is not only that the predictability sieve selects the most classical states, but also that their *evolution* is such that the deterministic, reversible trajectories of classical mechanics emerge.

Now first of all it must be clear what is meant by a 'trajectory' here. For if one rejects the collapse postulate quantum states always evolve deterministically: it is not the state itself, but the corresponding measurement outcomes (eigenvalues of a certain observable) obtained from a "measurement" that are indeterministic. This is not what Zurek is referring to. Supposedly, his claim is that the evolution of the coherent states is such that when mapped onto phase space the classical trajectories obeying Hamiltons equations emerge. Unfortunately, Zurek is not very clear about this issue:

> The evolution of a quantum system prepared in a classical state should emulate classical evolution that can be idealized as a "trajectory" – a predictable sequence of objectively existing states. [. . .] Einselected states are predictable: they preserve correlations and hence are effectively classical. (Zurek, 2003, pp. 735-736)

Thus, it seems that Zurek identifies predictability –characterizing trajectories– with stability, which was the criterion by which the classical states were defined in the first place. If this suspicion is correct, the claim that *given* the classical state of a system, it will result in trajectories, would be just a tautology. Surely, much more should be made of this: the connection between the simple stability criteria (2.18), (2.19) and (2.20) and the classical equations of motion is not that trivial. One would have to show that, for certain specific models perhaps, the proper equations of motion in the phase-space picture emerge. But although it has been claimed several times that this can be achieved (Joos, 2005; Joos et al., 2003; Zurek, 2003) no serious attempt to such a derivation seem to have been made. Indeed, it is questionable whether such an undertaking is viable at all.[34]

2.4 Concluding remarks

My aim in this chapter was to give a broad but representative overview of the decoherence literature, without getting lost in too many technical details. The presentation here largely followed the conventional treatment in the literature, and, with the exception of the last section, I have largely abstained from criticism.

The literature about decoherence can be difficult to grasp. The reason for this is twofold. First, there is a large amount of rather sophisticated technical literature about decoherence that does not touch upon the foundational issues (and has no such intentions). This literature I have largely avoided, because it seems to me that these kind of technical discussions are not likely to throw any light on the meaning of decoherence for the measurement problem. Second, there also exists another kind of literature (essentially, the authors that I have been quoting) that does aim to address the kind of questions I posed in the previous chapter. The problem with this kind of literature is that it is full of claims that are not really substantiated, that it is nowhere clearly stated what the questions are that are being addressed, and that it suffers from a thorough lack of self-criticism.

For a more critical evaluation of the achievements of decoherence, the standard formulations found in the literature are not really helpful. So what I will do in the next chapter, is to reformulate the essential characteristics of environment-induced decoherence from my personal point of view, in the hope that this will enable me to pinpoint the hidden assumptions, unjustified claims, and remaining problems. Tentatively, I would conclude that the discussion of the present chapter does not

---

[34] It might be more fruitful to invoke the notion of 'consistent histories' (as advocated by Griffiths, Gell-Mann and Hartle, and Omnès, among others) instead of this rather vague concept of "trajectories". The consistent-histories approach relies on the logical consistency of time-ordered propositions rather than deterministic dynamical evolution, and decoherence is often introduced in this framework to give a dynamical explanation for the consistency of histories. (Rather than introducing them in an *ad hoc* manner.) See Landsman (2007) for a discussion of the merits of this approach.

lead me to abandon the main conclusion of the previous chapter; that environment-induced decoherence cannot account for the appearance of facts within the standard quantum formalism (without collapse).

I have also discussed the relevance of decoherence for the "emergence of classicality" in the more general sense. In contrast to the rest of this chapter, I did not attempt to hold back my criticism of this program, that will not be part of the next chapter. This criticism should have made clear that even if decoherence may be an important mechanism behind the emergence of classicality, and as such is a welcome addition to the more mathematically-oriented approach of the classical limit, the story it tells is far from complete. This is not only because a general argument is lacking – only particular features of classicality for specific systems under special conditions have been derived – but especially because the approach suffers from an inherent vagueness that does not do justice to the well-posed questions that *can* be asked about the emergence of classicality, difficult as the subject may be. Often it is not clearly stated what the problem is or what the results are, and if claims are made regarding the appearance of effective classical behavior in decoherence models, at a closer look there turn out to be hardly any conclusive arguments to support them. Concepts and terminology are used without a clear definition and are mixed up at will, assumptions are made but are not made explicit and are often dropped again if convenience demands: what is meant by stability? Is it a feature of quantum systems (Zeno effect) or of classicality (robustness)? Is it imposed as a selection criterion (predictability sieve) or derived as a consequence indicating classicality (trajectories)? What are the kind of systems that interact strongly with an environment? Do we still need to assume a collapse postulate or not? And so on. To give a full account of the emergence of classicality along these lines, a lot of work remains to be done. And in my opinion this not only concerns calculations on ever more complex (and, hopefully, realistic) models, but also an unambiguous formulation of the theory.

Chapter 3

Interpreting decoherence: philosophical issues

In this chapter, I will discuss the theory of decoherence from a philosophical perspective. This philosophical discussion bears on two different issues. The first, which will be the topic of section 3.1, is the world view underlying decoherence. The most important consequence of this world view for a decoherence-based interpretation of quantum mechanics is that the classical world loses its fundamental status and instead becomes a relative, approximate and anthropocentric concept. My second concern will be the more down-to-earth claims that are made by the decoherence theorists. I believe that these claims also require some philosophical scrutiny, especially if they are to be taken as a (partial) solution to the measurement problem. In section 3.2, I will therefore have a fresh look at the basic results of decoherence, and I will briefly analyze the background assumptions of the theory and how they are related to the various conclusions one may draw from decoherence. Finally, in section 3.3 I will distill some open questions from the foregoing discussion that should be addressed if decoherence is to be considered a serious solution to the measurement problem, and to some extent try to answer them.

3.1 General philosophical background

As noted already in chapter 1, decoherence is usually not presented as an "interpretation" of quantum mechanics, but as a purely physical mechanism; its aim is to make sense of the measurement problem (or more generally: the emergence of classicality) by including the environment into the description of the measurement process, using the standard formalism without state vector collapse. It also claims to avoid any metaphysics to link the abstract notion of the state vector to our concrete experiences. Yet, of course, without *some* interpretation the quantum formalism is just mathematics without empirical content. Nevertheless, the general idea seems to be that the mechanism of environmental decoherence itself is compatible with any kind of interpretation one might attach to the wave function; such an interpretation should make no difference to the technical part of the argument. The point at which the need for an additional interpretation becomes manifest is when one needs to go from the diagonal reduced density matrix to the occurrence of definite and unique measurement outcomes (see chapter 1). (In fact the very idea that a diagonal density matrix would solve the measurement problem is an interpretational assumption in itself, cf. section 1.2.) As I will discuss in some more detail in section 3.3, most criticisms of decoherence focus on this issue. Consequently, an additional interpretational framework (in particular an Everett-like interpretation of the state vector) is often invoked by the adherents of decoherence to cover up this gap in their explanation.

But although decoherence is not an interpretation of quantum mechanics in the usual sense, taking it seriously nonetheless implies a specific philosophical stance towards issues like the tension between the epistemological and ontological aspects of quantum mechanics, the status of the classical world[1], and the need for taking into account the role of the observer. In this section I will focus on the role of the observer and the rejection of a separate classical realm in decoherence. This will set the stage for my discussion of the interpretation resulting from embedding decoherence within a many-worlds framework in the next chapter.

3.1.1 Decoherence and the (dis)appearance of a classical world in quantum theory

In chapter 1, I argued that the essence of the measurement problem as seen by many physicists is not that quantum mechanics tells us that the micro realm is weird, for one might well live with that. The real problem (at least as Schrödinger saw it) is that there seem to be processes, namely measurements, that transfer the "indefiniteness" of the quantum world up to the macroscopic realm. Up to this day, no one has been capable of arguing convincingly why the quantum mechanical description breaks down somewhere in between these regimes. More generally speaking, if one is a realist about scientific theories, it is to be expected that either quantum mechanics or classical physics is a more fundamental theory, and that the laws of one can be derived as consequences or limiting cases of the other.

Early interpretations of quantum mechanics did not consider the quantum formalism as fundamental, and so they focused on maintaining a classical realm (broadly understood). The most prominent example are hidden-variable theories, that aim to explain the probabilistic nature of quantum phenomena in terms of a classical statistical causal (and usually deterministic) model[2]. Thus, the HV theories aim to maintain the classical ideals of causality and definiteness for a deeper underlying theory, but they do not demand that reality is classical at the "surface" (where the phenomena are typically probabilistic.) Another paradigmatic example of the viewpoint that classical physics is more fundamental, at least in its role of translating the abstract quantum formalism into comprehensible physical data, is the Copen-

---

[1] For the sake of the present discussion, in what follows I will equate the existence of an (approximate) superselection rule with the validity of the whole corpus of classical physics. As discussed in the previous chapter, decoherence is often regarded as the physical mechanism behind the appearance of classicality understood in a much broader sense than just superselection, but I think that the evidence in support of this claim is rather poor.

[2] This is a view that is often wrongly attributed to Einstein. What Einstein really demanded was to reconstruct quantum mechanics as a *realist* theory, and it was in this respect that he opposed the instrumentalist inclinations of the Copenhagen interpretation. (See for instance (Bub, 1997; Fine, 1996).)

hagen interpretation. In fact, the status of the classical realm in the Copenhagen interpretation is the opposite of the HV ideal: Bohr never demanded a realistic model, conceivable in classical terms, that underlies the quantum appearances. He did, however, take the validity of the classical description for granted at the level of our ordinary experiences (to which the phenomena eventually belong)[3].

More recent interpretations (or physical approaches to the problem) such as environment-induced decoherence, take quantum mechanics to be the fundamental descriptive apparatus of physical reality – even though we hardly have a clue of what kind of reality this is. As I explained in chapter 1, these interpretations aim to derive the classical world (in particular, the notion of "definiteness") as an emergent concept. Despite the fundamental differences between approaches as diverse as GRW, environment-induced-decoherence and the many-worlds interpretation, they all share the common understanding that this emergent definiteness is to be attained by invalidating the superposition principle, either at the physical level or at the level of our experiences (see the discussion of the many-worlds view in chapter 1). This idea is, in turn, motivated by their acceptance of the e-e link. An important characteristic of the approach of environment-induced decoherence is that it does not impose classical definiteness as an *a priori* requirement on physical reality, but as a specific consequence of the way in which we, human observers, perceive the universe: the classical universe does not really exist, but is an anthropocentric notion, a mere "conceptual tool" to describe the phenomena as they occur to *us*.

That decoherence does not require exact definiteness is expressed in the concept of "approximate" superselection rules. There are two aspects to this inexactness that need to be distinghuished[4]:

1.a. Interference between eigenstates of the preferred observable only vanishes in the $t \to \infty$ limit, and

2.a. Interference only vanishes for a particular class of *local* observables (of the form $\hat{A} \otimes \hat{\mathrm{I}}$) on $\mathcal{H}_{\mathcal{A}} \otimes \mathcal{H}_{\mathcal{E}}$.

The local perspective expressed by the second point refers directly to the observer. The limitations to his observation capabilities essentially cause the "illusion" of classical definiteness. This is an obvious difference with the concept of observation that was criticized in the orthodox interpretation of quantum mechanics. There, the observer induces a *physical* intervention affecting the quantum state, exactly in order to *secure* classical definiteness. For decoherence the act of observation

---

[3] I should emphasize that Bohr's views on the subject are not always clear, but with respect to the status of classical physics he insisted upon a purely *epistemological (or even semantical)* quantum-classical distinction, not an ontological one. Bohr was also strongly opposed to the realistically motivated HV program. Nevertheless, despite this reluctance to ontological claims, it seems that Bohr's insistence on classical concepts was motivated by the belief that the "ordinary" world we experience is rightfully and completely described by these concepts. (See also Landsman (2007) for a discussion of the subtle relation between classical concepts and quantum physics in Bohr's philosophy.)

[4] From now on, I will assume that the environment is infinite, in order not to complicate the discussion with mathematical details that are largely irrelevant for the present purposes. For finite environments, the decoherence functional (2.14) has no limit since it is a periodic function. In that case, in point 1. below one would have to speak of interference terms disappearing only completely at discrete points separated by very large time intervals. (And of course, outside these points the function is *almost* zero almost everywhere, except at small intervals centered around the periodic points of recurrence.)

does not play such an active physical role, because it does not impose classical definiteness as a fundamental constraint on physical reality. Classicality is not a *modification* of the laws of quantum mechanics, but a *consequence* of these very laws if certain features of the observer (in effect, his locality) and certain limiting conditions are taken into account[5].

To elaborate a bit on this point; in the context of environment-induced decoherence, the observer has a double role with respect to the emergence of classicality. The first is rather trivial: if classical physics is not fundamental, but merely a useful approximation to the exact, but unpractical, quantum laws, one will have to specify the conditions that ascertain the usefulness of classical physics. These conditions will naturally contain some anthropocentric notion of "practical indistinguishability". In this context, the relevant question for decoherence is whether we would ever be able to perceive the tiny off-diagonal terms in the reduced density matrix. Probably not: decoherence models arguably show that the off-diagonal terms disappear too fast to be noticeable, for even the smallest dust particle[6]. (And remain negligibly small for billions of years.)

But the concept of locality implies that the theory is also anthropocentric in a less trivial way. To see this, consider the central claim of decoherence theory:

> *Approximate superselection rules (indicating classical definiteness) on the quantum state space of the system emerge (in a certain limit) if a part, usually called the "environment", is not observed.*

Compare to special relativity:

> The relativistic laws of motion are approximately Newtonian if the speed of a body is much slower than the speed of light.

In this simplified account, each theory (quantum mechanics or relativity) reduces another theory (classical physics or Newtonian mechanics, respectively) when a certain condition has been satisfied. But the crucial difference is that in the latter case, this reducing *condition* can be quantified and is independent of anthropocentric criteria as to when "approximate" is approximate enough (call this the reduction *criterion*). The theory of decoherence, however, introduces an anthropocentric notion of "observation" (or rather, non-observation) in the formulation of (part of) the condition that reduces classical physics to quantum mechanics *itself.*

One could therefore say that for environment-induced decoherence, classical reality is a good enough approximation (for all practical purposes) only as a consequence of our way of looking at things. I would characterize this situation as follows:

1.b. Quantum theory is universal, but classical physics may in some cases be a useful approximation (in particular, for typical 'macroscopic' systems that interact strongly with their environment).

2.b. The observer partakes in establishing the appearance of these classical phenomena.

---

[5] Of course, a thorough analysis of this issue requires that one makes precise what is meant by an "environment". This will be discussed in section 3.3.4.

[6] I agree that decoherence shows that interference terms are generally very tiny (at least in those models usually discussed in the literature). But see section 3.3.3 below for a critical discussion of the claim that smallness of interference terms is enough to secure classical definiteness.

These aspects are linked to the two aspects of the approximate superselection rules I distinguished: "our way of looking at things" is essentially local, and the remaining interference terms may be ignored "for all practical purposes" if they are so tiny that we will never be able to see these effects that would reveal the quantum mechanical nature of the universe to us. The difference is crucial. In principle, the idea that, as a consequence of our locality, we are unable to keep track of all the photons that scatter off a macroscopic object on the one hand, and that we are unable to see the remaining (local) interference terms on the other hand, have little to do with each other. After all, the tiny interference terms that remain when the environment is disregarded, are still part of the state of the system one *is* looking at. (The standard decoherence argument seems to neglect this distinction; I will elaborate on it in sections 3.3.2 and 3.3.3.)

Environment-induced decoherence can be characterized in terms of the two points above, but other interpretations may endorse either of the two. The Copenhagen interpretation is the paradigmatic example of an interpretation that maintains point 2.b., but rejects the other: Bohr takes the existence of a classical world and classical concepts for granted, denies that the classical aspects of the world can be cast in a quantum mechanical description, but instead believes (as expressed in his principle of complementarity) that the observer determines what can be the case, or at least which propositions about observable phenomena should be considered "meaningful". The projection postulate pushes the consequences of this philosophical view even further and requires that the observer induces a physical modification of the quantum mechanical laws to secure classical definiteness. Most other interpretations have embraced point 1.b., but renounce the idea that the observer could have any part in the *cause* of classical appearances. (Compare GRW theory[7].)

According to this classification[8], modern formulations of the many-worlds interpretation fall in the same class as decoherence theory. This also explains why decoherence fits the many-worlds approach so comfortably. Many-worlds interpretations take the state vector and the Schrödinger evolution as fundamental, and uphold a "perspectival" notion of facts. Modern versions of the MWI (see the discussion in the next chapter) are even closer in spirit to the decoherence programme, since they recognize that classicality is not only a subjective, but moreover an approximate concept. As we will see later, this has drastic metaphysical consequences. After all, taking quantum mechanics as seriously as decoherence and the MWI do implies a denial of our common-sense ontology that we call the "classical world". One might perhaps feel that these drastic metaphysical implications are not really a victory over the instrumentalist orthodox interpretation to feel comfortable with. Definitely not when the act of observation (or non-observation) re-enters the theory in a hardly concealing disguise. In chapter 4, I will investigate whether this is a price worth paying: if decoherence and the MWI are really capable of clearing quantum mechanics from the conceptual fog that surrounds it, then, perhaps, one

---

[7] Note however, that in GRW what is assumed universally valid is not "quantum theory" in the traditional sense, including the Schrödinger evolution, but quantum theory with the modified stochastic dynamical law. Putting things differently, one might say that quantum theory (including the Schrödinger evolution) is in fact an *approximation* to the *exact* law, which only becomes manifestly different at the macroscopic level. Thus, if one takes "quantum theory" to be standard quantum mechanics, GRW in a sense reject point 1.b. The essential point, however, is that the macroscopic physics is assumed to emerge from the low-level microscopic laws.

[8] But see Butterfield (2001) and Landsman (2007) for related, but slightly different accounts of interpretations of quantum mechanics along these lines.

should accept that the world behind the phenomena is not at all what we have always thought it to be.

3.2 Reconsidering the decoherence argument

Thus, the argument of decoherence asks us to acknowledge the failure of the program to recover an absolute classical reality from completely general arguments, and to seek instead an approximate classicality that emerges only relative to an observer and as a contingent feature of the way we perceive our universe. Whether or not one should gratefully give up the existence of an absolute classical reality and allow a prominent role for the observer in establishing that reality, is perhaps a matter of taste. In any case these are not the kind of issues that necessarily render an argument incomplete or inconsistent. I will now turn to some issues that may (possibly) do. Some of those were already briefly mentioned in the previous chapter; I will now discuss them in some more detail because it is these problems that an additional interpretative framework should aim to overcome.

First some words about notation: in the rest of this chapter and the chapter following, I will use the symbol $\mathcal{S}$ (from "system") to indicate an *isolated* quantum system[9], an $\mathcal{A}$ ("apparatus") for an *open* quantum system, and $\mathcal{E}$ denotes the environment with which $\mathcal{A}$ interacts. (In the next chapter, we will also encounter the observer $\mathcal{O}$, which also denotes an open quantum system, with the specific feature that it is used to measure another open quantum system (i.e. it is the apparatus for $\mathcal{A}$.)) I hope to minimize possible confusion by applying these conventions consistently.

3.2.1 Decoherence: what it means...

In order to pin down exactly what the possible difficulties with the approach are, I start with briefly recapitulating the central points of the previous chapter, in a somewhat different setting. Although simplified, the following presentation is meant to capture the essence of the decoherence argument. One should keep in mind, however, that research in decoherence goes considerably beyond the idealized picture I give here, but these technical subtleties are not likely to resolve any of the conceptual problems.

The core of Zurek's (1981, 1982, 2003) argument is as follows (compare section 2.2.1 in chapter 2): Suppose that we have established a correlation between an observable $\hat{S}$ with eigenstates $\{|s_i\rangle\}$ on the Hilbert space $\mathcal{H}_\mathcal{S}$ of a quantum system, and a corresponding observable $\hat{A}$ with eigenstates $|a_i\rangle$ of the apparatus with Hilbert space $\mathcal{H}_\mathcal{A}$, by means of some measurement interaction Hamiltonian $\hat{H}_{\mathcal{S}\mathcal{A}}$. First suppose that system and apparatus are in an eigenstate :

$$|\Psi_{\mathcal{S}\mathcal{A}}\rangle = |s_i\rangle|a_i\rangle. \tag{3.1}$$

According to the decoherence theorist, this picture is incomplete: the apparatus will generally interact strongly with its environment $\mathcal{E}$, through an interaction

---

[9] By "isolated" I mean that $\mathcal{S}$ does not interact with an ("irrelevant") environment $\mathcal{E}$. Of course, it can occasionally interact with an apparatus $\mathcal{A}$.

Hamiltonian $\hat{H}_{\mathcal{A}\mathcal{E}}$. If this interaction is of the form (2.7), with the "apparatus part" diagonal in the basis $\{|a_i\rangle\}$, then the result of this interaction will be a product state on $\mathcal{H}_{\mathcal{S}} \otimes \mathcal{H}_{\mathcal{A}} \otimes \mathcal{H}_{\mathcal{E}}$:

$$|\Psi_{\mathcal{S}\mathcal{A}\mathcal{E}}\rangle = |s_i\rangle|a_i\rangle|e_i\rangle \tag{3.2}$$

Compare what happens to an arbitrary state on the system-apparatus subspace (i.e. a superposition of correlated states (3.1)):

$$|\Phi_{\mathcal{S}\mathcal{A}}\rangle = \sum_i c_i|s_i\rangle|a_i\rangle \rightarrow \sum_i c_i|s_i\rangle|a_i\rangle|e_i\rangle = |\Phi_{\mathcal{S}\mathcal{A}\mathcal{E}}\rangle \tag{3.3}$$

Hence, states that are not an eigenstate of $\hat{S} \otimes \hat{A}$ will get entangled with the environment (i.e. the resulting state will not be a product state on $\mathcal{H}_{\mathcal{S}} \otimes \mathcal{H}_{\mathcal{A}} \otimes \mathcal{H}_{\mathcal{E}}$.) In fact, we could have omitted the system altogether (see also below), so one may conclude that there is particular set of apparatus states that are (in this sense) "preferred" by the interaction:

1. **Einselection -"no entanglement"**: There is a *unique* basis of apparatus states that do not get entangled with the environment, given a particular apparatus-environment interaction with Hamiltonian $\hat{H}_{\mathcal{A}\mathcal{E}}$.

The fact that most $\mathcal{S}\mathcal{A}$ states will become entangled with the environment, has an important corollary when we realize that we are interested in the $\mathcal{S}\mathcal{A}$-part only; that is, we have to 'trace out' the (orthogonal) environment states (indicated by the curly arrow):

$$\begin{aligned} |\Psi_{\mathcal{S}\mathcal{A}\mathcal{E}}\rangle &= |s_i\rangle|a_i\rangle|e_i\rangle \rightsquigarrow \rho^{\mathcal{E}}_{\mathcal{S}\mathcal{A}} = |s_i\rangle\langle s_i| \otimes |a_i\rangle\langle a_i| && (3.4)\\ |\Phi_{\mathcal{S}\mathcal{A}\mathcal{E}}\rangle &= \sum_i c_i|s_i\rangle|a_i\rangle|e_i\rangle \rightsquigarrow \rho^{\mathcal{E}}_{\mathcal{S}\mathcal{A}} = \sum_i |c_i|^2|s_i\rangle\langle s_i| \otimes |a_i\rangle\langle a_i| && (3.5) \end{aligned}$$

The reduced state $\rho^{\mathcal{E}}_{\mathcal{S}\mathcal{A}} \in \mathcal{H}_{\mathcal{S}} \otimes \mathcal{H}_{\mathcal{A}}$ at the right-hand side of (3.4) is pure. The density matrix at the right-hand side of (3.5), on the other hand, is mixed. (Note that the density matrix representation of $|\Phi_{\mathcal{S}\mathcal{A}}\rangle$ would be $\sum_{ij} c_i c_j^*|s_i\rangle\langle s_j|\otimes|a_i\rangle\langle a_j|$.) Again, we could have carried out the same analysis in terms of the apparatus alone. So, we have:

2. **Einselection - "loss of purity"**: Due to interaction with the environment, superpositions of eigenstates of the preferred observable $\hat{A}$ will become mixed when restricted to the subspace $\mathcal{H}_{\mathcal{A}}$ of the apparatus alone.

This "loss of purity" leads Zurek (1993, 2003) to quantify his selection criterion for the preferred basis in terms of the predictability sieve (p. 48). Equivalently, one could say that the time evolution of a state (in density matrix representation) $\rho_{\mathcal{S}\mathcal{A}}$ restricted to the system-apparatus product space $\mathcal{H}_{\mathcal{S}} \otimes \mathcal{H}_{\mathcal{A}}$, is not unitary. (The non-unitary term in the generalized master equation is (nearly) equal to zero for the class of einselected states.) Thus, only the eigenstates of the system-pointer observable evolve "as if" the environment is not present.

Summarizing the above, the consequence of the apparatus-environment dynamics and the partial tracing for the state space of the apparatus is the following:

3. **Einselection - general**: The interaction with the environment restricts the state space of the apparatus to
    (a) Eigenstates of the preferred observable $\hat{A}$.
    (b) Mixed states.

So far, I focused on the apparatus states only. The fact that these are correlated to certain system states seems to be quite irrelevant for the argument. (That is why the system observable was omitted in the points 1-2 above, but one could just as well replace the preferred apparatus observable $\hat{A}$ by a system-apparatus observable $\hat{S} \otimes \hat{A}$, in which $\hat{S}$ can be any observable on $\mathcal{H}_{\mathcal{S}}$ – depending on the exact $\mathcal{SA}$-interaction $\hat{H}_{\mathcal{SA}}$ of course.)

The role of the measured system is clarified if we consider Zurek's slightly different presentation of his argument (Zurek, 1981, 1982, 1993, 1998). There, the focus is on the *correlations* between the $\mathcal{S}$, $\mathcal{A}$ and $\mathcal{E}$ states in the state vector, rather than density matrix, representation. In contrast to the selection criteria mentioned thus far, this argument, which was initially motivated by the measurement problem, does pertain to the $|s_i\rangle$ as well as the $|a_i\rangle$ states, and fixes a *decomposition* for states like (3.3). Zurek argues that the correlations between system and apparatus states can remain intact if the measurement interaction Hamiltonian $\hat{H}_{\mathcal{SA}}$ is such that the eigenstates $|s_i\rangle$ of the "to-be-measured system observable" $\hat{S}$ couple to the preferred pointer states that do not entangle with the environment. This should be clear from (3.3).

That is, the stability of the pointer states (preferred states) in the sense of point 1. guarantees stability of their correlations with some set of system states. The claim is that this condition is in fact necessary, in particular for having *instantaneous* correlations. Recall the preferred basis problem as it was formulated in chapter 1. There, I showed that there may also exist other (possibly bi-orthogonal) decompositions of $|\Phi_{\mathcal{SA}}\rangle$ which, although represented as a single sum, do not correspond to the desired system-apparatus correlations. The crucial point is that these alternative single-sum decompositions will not remain intact when coupled to the environment: write

$$|\Phi_{\mathcal{SA}}\rangle = \sum_i c_i |s'_i\rangle|a'_i\rangle = \sum_i c_i |s'_i\rangle(\sum_j \lambda_j^{(i)}|a_j\rangle) \tag{3.6}$$

with the "alternative" apparatus states $|a'_i\rangle$ expanded in the basis of pointer states $|a_j\rangle$. Then, after interaction with the environment:

$$|\Phi_{\mathcal{SAE}}\rangle = \sum_{ij} c_i \lambda_j^{(i)} |s'_i\rangle|a_j\rangle|e_j\rangle \neq \sum_i c_{ij}|s'_i\rangle|a'_i\rangle|e'_j\rangle \tag{3.7}$$

that is, the resulting entangled state on $\mathcal{H}_S \otimes \mathcal{H}_A \otimes \mathcal{H}_{\mathcal{E}}$ cannot be expanded in the same basis (nor any other) of correlated SA-states that constituted the "alternative" decomposition before interaction with the environment, and arbitrary environment states $|e'_j\rangle$ This is due to the fact that the entangled apparatus-environment part of $|\Phi_{\mathcal{SAE}}\rangle$ above is not a simple product state on $\mathcal{H}_A \otimes \mathcal{H}_{\mathcal{E}}$. (Cf. section 2.2.2.)

4. **Einselection - correlations**: If the system states are to *remain correlated* to the apparatus states under interaction with the environment, one has no other choice but to take for the latter the preferred states $|a_i\rangle$.

This argument can be cast in mathematical terms by invoking the tridecompositional uniqueness theorem (Elby and Bub, 1994; Bub, 1997). It says the following:

> **Tridecompositional uniqueness theorem (Bub, 1997):** If $\{|\psi_i^1\rangle\}$ and $\{|\psi_i^2\rangle\}$ are linearly independent sets of vectors on Hilbert spaces $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_3$, respectively, and $\{|\psi_i^3\rangle\}$ is any set of mutually non-collinear unit vectors, then the representation of $|\Psi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3$ in the form of a tridecomposition:
>
> $$|\Psi\rangle = \sum_i c_i |\psi_i^1\rangle|\psi_i^2\rangle|\psi_i^3\rangle \tag{3.8}$$
>
> is unique, if its exists[10].

The tridecompositional uniqueness theorem leads to a different formulation of einselection and decoherence. (These are largely equivalent, but there are subtle differences that will be analyzed in section 3.3.1 below and in my discussion of Zurek's "existential interpretation" in chapter 4.) Its main consequences are:

5. The decomposition of the final state $|\Phi_{\mathcal{SAE}}\rangle = \sum_i c_i|s_i\rangle|a_i\rangle|e_i\rangle$ is *unique*; there is no alternative choice of basis vectors $\{|s_i'\rangle\}$ on $\mathcal{H}_\mathcal{S}$ and $\{|a_i'\rangle\}$ on $\mathcal{H}_\mathcal{A}$ that will decompose the same state vector $|\Phi_{\mathcal{SA}}\rangle$ as a single sum as well.
6. Whether the final state $|\Phi_{\mathcal{SAE}}\rangle$ has such a tridecomposition depends primarily on this final state, or, equivalently, given a particular (total) interaction Hamiltonian $\hat{H}_{\mathcal{SAE}}$, on the initial state.
7. The tri-decomposition fixes the *preferred states.* Their exact form therefore depends on the state $|\Phi_{\mathcal{SAE}}\rangle$.
8. The tri-orthogonal decomposition also fixes the environment states. These need not be orthogonal.

The argument so far only involved the environment for selecting the preferred "pointer" basis. The specific form of the apparatus-environment interaction Hamiltonian determines what this basis will be. There is, however, another aspect to the dynamics which, although not completely unrelated, can be considered independent of the former. In (3.4) and (3.5) above, it was assumed that the environment states $|e_i\rangle$, correlated to the apparatus states $|a_i\rangle$, are orthogonal. This assumption is *not* required for the conclusions 1, 2, 3 or 4, but it *is* necessary to remove the off-diagonal "interference terms" from the reduced density matrix $\rho_{\mathcal{SA}}^{\mathcal{E}}$. In that case, the mixed states of point 3 (b) are convex sums of the preferred states (i.e. mixtures of the states of 3 (a).)

There is no *a priori* reason why the environment states $|e_i\rangle$ should be orthogonal. It is not a corollary of the tridecompositional uniqueness theorem, and most decoherence models show that they will generally *not* be orthogonal. What decoherence models (such as on p. 43) *do* show, however, is that the environment states will very rapidly converge to orthogonality in the $t \to \infty$ limit. In terms of the tridecompositional theorem, the total interaction $\hat{H}_{\mathcal{SAE}}$ is such that the final state at $t = \infty$

---

[10] This result generalises to n-decompositions ($n > 2$).

allows not merely a tridecomposition, but, specifically, a tri-*orthogonal* decomposition. (At least, in the simple decoherence models where we assume the system and apparatus states to belong to the eigenbasis of an Hermitian observable.) So, orthogonality of the environment states is a contingent *dynamical* effect.

9. **Diagonality**: For particular interaction Hamiltonians, the environment states will become orthogonal in the limit $t \to \infty$. Consequently, the basis $|a_i\rangle$ defined by the eigenstates of the preferred observable $\hat{A}$ gives the *unique* expansion of (the density matrix representation of) the state on $\mathcal{H}_\mathcal{S} \otimes \mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{E}$ that converges to a reduced density matrix $\rho^\mathcal{E}_{\mathcal{SA}}$ diagonal in the $|s_i\rangle|a_i\rangle$ basis (with the system states $|s_i\rangle$ uniquely fixed by the tri-decomposition).

This conclusion also holds in case there is only an apparatus and an environment (cf. equation (3.5)), even though the tri-decompositional uniqueness theorem does not apply. (The result is that there may be alternative instantaneous decompositions at each moment; but only the preferred states *remain* correlated during the evolution, as the environment states become more and more orthogonal.)

3.2.2 . . . and what it takes.

One can list the assumptions that were (mostly) implicit in the above presentation of decoherence as follows:

Premises:

i. **Interactions** All system-environment interactions that are known to occur in Nature are faithfully represented by the model Hamiltonians studied in the decoherence literature.

ii. **Subspaces** Everything that contributes to the dynamics of the system of interest is completely specified in terms of *three or more subsystems.*

iii. **Correlations** There is a general constraint on the representation of states on this Hilbert space, in terms of the tridecomposition. This constraint expresses a requirement of the existence of *correlations* between the states on the subspaces.

iv. **Infinite time** "For all practical purposes", the state $|\Phi_{\mathcal{SAE}}\rangle$ may be replaced by the state in the limit $t \to \infty$[11].

v. **Locality** We consider observables of the form $\hat{S} \otimes \hat{A} \otimes \hat{\mathbb{I}}$ on $\mathcal{H}_\mathcal{S} \otimes \mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{E}$ only. Therefore, the state $|\Phi_{\mathcal{SAE}}\rangle$ may be replaced by the reduced density matrix $\rho^\mathcal{E}_{\mathcal{SA}}$.

vi. **IgI** The mixed states are to be interpreted as a proper mixture of the pointer states that appear on the diagonal.

---

[11] The "practical purposes" are to be determined by the decoherence timescale, which is dependent on the exact dynamics. When adopting assumption (v) as well, one may replace "state $|\Phi_{\mathcal{SAE}}\rangle$" by "reduced density matrix $\rho^\mathcal{E}_{\mathcal{SA}}$".

The last of these assumptions may be weakened in a sense that will be explained below. Assumption (i) might seem slightly superfluous. This is not quite so, although its role in the issues mentioned here is admittedly rather marginal. The point is that the validity of the results (1)-(9) is in many cases constrained by contingency. This can best be seen from the tridecompositional uniqueness theorem. The result 6 and its consequences 7 and 8 (p. 71), depend on accidental features of the final state. Part of the decoherence literature can be read as an attempt to replace this contingency by a notion of *physical* or *epistemic necessity.* For instance, regarding point (i); by appealing to the kind of interaction Hamiltonians that are present in our universe (or that are known to us), it is explained why all macro-objects (as far as we know) occur in "localized states" (approximate eigenstates of position). (In section 3.3.1 I will consider the need for assumption (i) from a different perspective.)

Precisely which of the above premises are needed depends in a rather subtle way on the kind of perspective one adopts with respect to decoherence (i.e. what one believes its essential tenets are); it turns out that many important results follow immediately even *without* appealing to central notions such as Locality (v) and the orthogonality of environment states. One can capture the essence of these various perspectives on the decoherence argument as follows:

A. Interactions with the environment lead to a *selection of a preferred basis*, through a criterion of "no entanglement", as implied by result 1. This only requires assumption (i).

B. Interactions with the environment lead to a *selection of a preferred basis*, through a criterion of "purity", as implied by result 2. This requires assumption (i) and (v).

C. Interactions with the environment lead to a *selection of a preferred basis*, through a criterion of "stable correlations", as implied by result 4 ("Einselection-correlations"). This requires assumptions (i), (ii) and, trivially, (iii).

D. Interactions with the environment lead to the *disappearance of interference terms* from the local density matrix of $\mathcal{S}$ on $\mathcal{H}_{\mathcal{SA}}$, as implied by point 9 ("Diagonality"). This requires assumptions (i), (ii), (iii), (iv) and (v).

E. Interactions with the environment explain our perception of classically "definite", unique *measurement outcomes.* This conclusion relies on (but does not follow from) assumption (vi), in addition to (i), (ii), (iii), (iv) and (v).

Note that approach A plus Locality (assumption (v)) is equivalent to B. In what follows, I will usually not distinguish these criteria from one another when the distinction is not essential. Moreover, approach C and A are equivalent under the premise (ii) due to the tridecompositional uniqueness theorem, *in case the final state* $|\Phi_{\mathcal{SAE}}\rangle$ *is tridecomposable.* This equivalence does therefore not hold unconditionally. (As I will discuss later (section 3.3.1), the requirement that the final state be tridecomposable imposes constraints on the dynamics that are naturally satisfied in the case of the simple measurement model discussed here, but probably not in more general cases.)

Except for point E, I have avoided including any interpretational claims. This is because I would like to keep a strict distinction between the (sometimes philosophically motivated) assumptions that underlie the physics of the approach, and explicit interpretational moves that serve to make sense of the results. However, the distinction is not always as clear-cut as I present it here. In particular, there exists a common interpretation of decoherence that more or less makes the same claim as E but scrupulously avoids assumption (vi). Instead, it uses a weaker assumption I would paraphrase as follows:

vii. **Observable** The preferred states (defined in terms of any of the selection criteria A-C) define the *interpretation basis* (see p. 18) of the system, apparatus, or system-apparatus state.

I give some comments on the implications of this assumption in the next section.

3.3 Some (partial) answers and more questions

In what follows, I will attempt to shed some light the following four main issues:

1. **The selection criteria, §3.3.1:** How are we to understand the selection criteria (point A, B, or C) for the preferred basis? What is the underlying motivation for adopting a particular rule, and do such rules always apply? What is their epistemic (or ontological) relevance?
2. **The ignorance interpretation, §3.3.2:** Can one justify assumption (vi)?
3. **Imperfect decoherence, §3.3.3:** The justification of premise (iv). Does mere smallness warrant us to ignore the interference terms altogether "for all practical purposes"?
4. **Subsystems and the environment, §3.3.4:** How restrictive are the premises (ii) and (v), and how can they be justified?

3.3.1 The selection criteria and the status of observables

Of the premises above, (vi) is clearly the most controversial. It has certainly been most severely criticized in the philosophical literature, especially in relation to point (v) (recall my discussion in section 1.2 in chapter 1). Decoherence theorists have generally come to accept these criticisms and therefore accept that decoherence alone does not solve the measurement problem. However, as I said, there is a common understanding of decoherence which could be characterized as a weakened version of the view labelled E, invoking assumption (3.2.2) instead of (vi). It harks back to the discussion of section 1.2.1, in particular the idea that decoherence somehow implies definiteness.

Consider for instance, Joos' remark:

> "[M]easurement-like interactions cause a strong quantum entanglement of macroscopic objects with their natural environment. The accompanying delocalization of phases then effectively "destroys" superpositions

> between "macroscopically different" states with respect to a local observer, so that the object *appears* to be in one or the other of those states. [. . .] As long as no collapse is assumed, phase relations are of course still present in the whole system and never 'destroyed' " (In (Joos et al., 2003, p.45)).

Earlier, Joos makes clear that he is not alluding to an ignorance interpretation in the literal, naive, "classical" sense when he writes that "the formal ensemble of states characterizing the local density matrix is not a "real" ensemble in the sense of statistical mechanics" (*Ibid.*, p. 43). Instead, the idea seems to be that by calling the mixture of pointer states "apparent", the usual objections to the ignorance interpretation somehow disappear. Thus, the system is said to appear to have a definite value for the quantity defined by the (eigen)states on the diagonal, even though Joos admits that he cannot account for the particular value that eventually occurs (recall the discussion in sect. 1.3). To address the latter issue is then the central remaining problem of interpretation for decoherence.

The "macroscopically different" states Joos is referring to, are the dynamically preferred states that remain pure under the interaction with the environment (selection criterion B). However, this by itself does not imply that superpositions of those dynamically preferred states are somehow "destroyed", not even in an "apparent", local sense. It is clear from the result (3) in section 3.2.1 that the non-unitary (local) dynamics simply turns the non-preferred pure states into mixed states – but it requires an additional assumption (namely, assumption (3.2.2)) to interpret the latter empirically in terms of the preferred basis.

Of course, this is a rather natural assumption which could harmlessly be added as such to the decoherence program. But it must be stressed that the general preferred basis problem – the fact that a choice of basis is not physical, cf. section 1.1.1 – cannot be solved by appealing to physical interactions alone. Some kind of rule to constrain the arbitrary decompositions of the state vector to those that are physically meaningful should be imposed if we are to address that problem. The merit of decoherence is that this rule can be formulated entirely in terms of the dynamics, and does not involve the observer's whim (as in the orthodox interpretation).

Specifically, this implies that non-orthodox interpretational assumptions already enter the decoherence program *before* one arrives at the point of addressing the problem of outcomes. That is, if one accepts Joos' and other, similar, claims in the interpretation-oriented literature that decoherence gives rise to the appearance of classical definiteness (or classicality in general, cf. section 1.2.3). (There is a sense in which the decoherence program remains "entirely within the standard quantum formalism (i.e. without adding any new elements into the mathematical theory or its interpretation)" (Schlosshauer, 2004). But that purely *physical* program of course *requires* the orthodox framework to interpret its results empirically. I will come back to this issue in the final Conclusion.)

The status of observables in decoherence

It is not at all clear that the decoherence theorists acknowledge the need for an explicit interpretational rule like 3.2.2. Remarks like the one from Joos quoted above and from Zurek on p. 42 seem to indicate the opposite, but there are also suggestions in the literature that seem to argue for such a revision of the interpretational framework of standard quantum mechanics:

> "The [...] description of measurements of the first kind by means of probabilities for transitions [...] (or, for that matter, by observables) is phenomenological. However, measurements should be described *dynamically* as interactions between the measured system and the measurement device. The observable (that is, the measurement basis) should thus be derived from the corresponding interaction Hamiltonian and the initial state of the device" (Zeh, in Joos et al., 2003, p.20).

> "It is tempting to speculate that one could dispose of the observables [...] altogether in the formulation of the axioms of quantum measurement theory. The only ingredients necessary to describe measurements are then the effectively classical, but ultimately quantum, measurement apparatus and the measured system. Observables emerge as a derived concept, as a useful idealization, ultimately based on the structure of the Hamiltonians. [...] Einselection should be included in this program, as it decides which observables are accessible and useful – which are effectively classical"(Zurek, 2003, p.751).

What Zeh and Zurek are suggesting, is that we should replace the "observables" as a primitive interpretative concept by a dynamically emerging heuristic that is tied to the specific measurement interactions involved. Let me start by explaining what this general idea amounts to.

The Born rule assigns probabilities to measurement outcomes. In conventional quantum mechanics, however, one usually does not take the measurement itself into account when calculating probabilities. Instead, probabilities are extracted from the quantum state by specifying a particular orthogonal basis, usually provided by an observable. So the state $|\Psi\rangle$ can be expanded as $|\Psi\rangle = \sum_i c_i |s_i\rangle$ if we want to know the probabilities for the eigenstates $|s_i\rangle$ of a certain observable $\hat{S}$, as well as by $|\Psi\rangle = \sum_i c'_i |s'_i\rangle$ if we want to know the probabilities for the eigenstates $|s'_i\rangle$ of another observable $\hat{S}'$. This choice does not reflect in any way the specific context in which these observables are brought physically to life (cf. my discussion of the (general) preferred basis problem in section 1.1.1).

The above remarks by Zeh and Zurek suggest that the orthodox application of the Born rule is just a shorthand for a more complete description of the measurement process. For simplicity, I take this to be the ideal measurement as characterized by the von Neumann measurement scheme (1.2 in section 1.1)[12]. The interaction $\hat{H}_{\mathcal{SA}}$ is such that the system-apparatus couple end up in a correlated state $|\Phi_{\mathcal{SA}}\rangle = \sum_i c_i |s_i\rangle|a_i\rangle$. A measurement process that results in this particular form of the final state can be completely characterized in terms of the interaction $\hat{H}_{\mathcal{SA}}$ and the initial state of the apparatus, i.e. it is independent of the initial state of the system[13]. To find out the probabilities for the measured system states $|s_i\rangle$,

---

[12] van Fraassen (1991) discusses the implications of more general measurement schemes for this kind of interpretation. See also section 3.3.3.

[13] Here, a particular cut-off time t_m is supposed to be included in the specification of $\hat{H}_{\mathcal{SA}}$; that is, the measurement stops as soon as the system and apparatus have become perfectly correlated.

one only needs to perform a repeated measurement and check the probabilities for the corresponding apparatus states $|a_i\rangle$ if the combined system is in the state $|\Psi\rangle$. But –and this is the crucial point– the scheme does not apply to another system observable $\hat{S}$ with eigenstates $|s_i'\rangle$, and so taking the Born rule seriously implies that one cannot extract probabilities for the latter from the same post-measurement state $|\Psi\rangle$. If we expand the state as $\sum_i c_i|s_i\rangle$ we are in fact referring to the measurement above. And if we expand it as $\sum_i c_i'|s_i'\rangle$ we are in fact referring to another measurement, with another interaction Hamiltonian. The reason that one can take this shortcut and apply the Born rule to the system state directly is clear from von Neumann's (1955) analysis of the measurement process. He showed that applying the Born rule for the observable $\hat{\mathrm{I}} \otimes \hat{A}$ to the state (3.3) yields the same statistical predictions as applying it for the observable $\hat{S}$ on the state $\sum_i c_i|s_i\rangle$ directly, due to the correlations between the system states $|s_i\rangle$ and apparatus states $|a_i\rangle$, and the orthogonality of the latter. But the convenient fact that such a shortcut works should not make us forget what the probabilities given by the Born rule actually refer to, namely to the complete measurement setup.

This is what I think Zeh is referring to when he suggests that "in this dynamical sense, the interaction with an appropriate measuring device *defines* an observable"(Joos et al., 2003, p.20). The reason that one and the same quantum state allows the application of the Born rule for different observables, is that one can generally devise various measurement interactions, each of which couples a different basis of $\mathcal{S}$ to the apparatus[14].

But if we look upon it like this, what is the relevance of the environment? After all, the measured observable can be completely specified in terms of the $\mathcal{SA}$ interaction Hamiltonian. There are two kinds of answers to this question, I think.

The first is found in Zurek's (1981) paper on the preferred basis problem. Zurek takes for granted that *only* a bi-decomposition of the $\mathcal{SA}$ state represents an actual measurement, and so determines the empirical content of the post-measurement state $|\Psi\rangle$, but he points out that this state may allow for several bi-decompositions in case of (accidental) degeneracy. As I explained in section 2.2.2, he solves this problem by appealing to the ensuing interaction of the apparatus with the environment. Thus, only the decomposition in terms of the $|a_i\rangle$ basis that commutes with the $\mathcal{AE}$ interaction Hamiltonian can be stable under the interaction with the environment. So far, this argument is correct, but in the present context it seems rather superfluous to me. Zurek claims to solve the problem of accidental degeneracy by pointing out that it is in fact a matter of *dynamics*. But if we are to take recourse to the dynamics, we could just as well refer to the dynamics of the measurement interaction directly, i.e. invoke the $\mathcal{SA}$ hamiltonian, which selects the measured observable in a completely unambiguous way[15].

---

[14] As such, the present proposal is very much in spirit of Bohr, who stressed "the impossibility of any sharp distinction between the behavior of atomic objects and the interaction with the measuring instruments which serve to define the conditions under which the phenomena appear" (cited in Mermin (1993)). I would paraphrase this difference between the Copenhagen and the orthodox interpretation as that for the Copenhagen interpretation, observables refer to what one would like to *measure*, whereas the orthodox interpretation is merely concerned with what to *calculate*.

[15] Similarly, before discussing Zurek's argument, van Fraassen (1991) remarks that "To deal with the problem of accidental degeneracy, I imposed conditions on the operator that governs the temporal evolution, and not on the character of initial and final state alone. Hence it is not ruled out that the end state may be the same in two measurements, on systems in the same initial state,

The second point of relevance for the environment is more straightforward. We have been concerned with measurement interactions, but the emergence of classicality involves a more general question about empirical content: why is there only a limited set of "classical" quantities that appear to be definite for "ordinary" objects? The answer decoherence offers to this question is that any open system is also in a measurement-like interaction with its environment, which similarly determines a preferred basis of "effectively classical" states. The crucial difference between the quantum and the classical case is then that the apparatus-environment interaction is essentially fixed by Nature (assumption (i)), in contrast to the system-apparatus interaction.

I would like to stress that this particular view on decoherence does not depend on the locality principle (v). The vanishing of interference terms from the reduced (system-)apparatus density matrix does not play any role in defining the preferred states, nor in the assumption that these preferred states constitute the interpretation basis. In fact, I think that the misleading suggestion that the dynamics somehow "forces out" a particular interpretation basis, is closely tied to the suppression of interference between preferred basis in the local reduced density matrix. For instance, Zeh remarks that "only a basis of states that are not immediately destroyed by decoherence defines "realizable observables" (Joos et al., 2003, p.19). But decoherence (including the locality condition) merely turns most pure states into mixed states; of course it does not 'destroy' states and it certainly cannot 'prohibit' a certain choice of basis (unless that is prohibited by our interpretation!)[16] Furthermore, the locality condition is strictly speaking in conflict with the idea that *all* observables allow such a dynamical reconstruction, since the locality principle the assumption that there is a limited class of 'allowed' observables.

Relations between the selection criteria

I believe that this suggested revision of the concept of "observable" would require a more thorough philosophical underpinning. Without going into the details, it seems difficult to me to interpret the resulting theory in a *realist* sense, since it fails to provide some kind of ontological picture behind the phenomena. (The concept of observable should now be considered as a shorthand for a certain interaction that gives rise to certain manifestations in the measurement apparatus, but no more: in particular, they no longer refer to features of the measured quantum system –even less than in orthodox quantum mechanics.) For a more realist-oriented alternative, one could perhaps revise the *physics*, instead of the interpretation of the formalism, and postulate that "effectively classical" (that is, open) systems only exist in dynamically stable states. However, I believe that this assumption would be hard to justify.

In any case, whether one postulates a new interpretational rule or a constraint

---

of two incompatible observables. This is a little curious, though not troubling if the concept of measurement pertains to the type of process as a whole" (p. 216).

[16] The locality condition also seems to underlie Zeh's remark that "this algebraic classification [of realizable observables] remains an approximate and dynamically emerging scheme" (*Ibid.*). Since the Hamiltonians can be defined exactly, the fact that observables are defined in terms of the dynamics does not lead to inexactness. Presumably, Zeh is alluding to the idea that one should look at the (instantaneous) reduced state of the system as it evolves under the environmental dynamics to obtain the relevant observables; this instanteneous state is not exactly diagonal for finite t, but that is in fact irrelevant for the present argument.

on the kinematics of open systems, it seems to me that both these approaches fall short on explanatory power. However, when considering the specific case of a *measurement* interaction there is an alternative approach which is more straightforward. Recall that in chapter 1, I distinguished two formulations of the preferred basis problem: the general one and the “decomposition” version. The latter incorporates assumption (iii), which provides a natural and sufficient argument to define the interpretation basis in terms of the preferred states selected by criterion C. Note that the difference with the general preferred basis problem is that the decomposition problem is not a question about empirical content *per se* (i.e. we need not explain why we perceive the quantum state as we do), but is a more pragmatic question of how the basis ambiguity can be removed under the constraint imposed by assumption (iii). The tridecomposition theorem, in turn, says that the preferred basis should be selected according to criteria A or B precisely *because* in the context of the measurement process they are equivalent to C.

However, the necessity for such a (pragmatic) justification as the “stable correlations” requirement also places an additional burden especially on assumptions (iii) (“Correlations”) and (ii) (“Subspaces”) on page 72. The distinction between the question of “emergence” on the one hand, and the more restricted application of quantum measurements on the other, now becomes essential: for the first problem there is no “apparatus” to be distinguished from the ‘system”, and no straightforward way to restate the question such that the “correlation” constraint acquires a natural justification.

So it seems that one is left with a couple of options.

- The first is to adopt a minimalist version of the measurement problem that merely requires that (in order to apply the Born rule), the relevant observable is linked physically to a particular “measurement context” (interpreted in a broad sense). Then *postulating* that the interpretation basis is defined in terms of the dynamics (by appealing to the selection criteria A or B) will perhaps suffice. This implies a revised concept of observables along the lines sketched above.
- Second, if one focuses on the *decomposition* version of the preferred basis problem (i.e. in the context of a von Neumann measurement), this implies approach C in a fairly unproblematic way.
- Finally, one might want to explain why one perceives macroscopic objects as possessing definite properties (i.e. including, but not restricted to, measurement pointers subject to a specific kind of interaction with a quantum system), on other words, how “facts” come about in a quantum world. Depending on the physical setting of the problem (measurements or emergence) one could take either of the selection criteria, but in addition some kind of argument or interpretational rule to solve the problem of Outcomes is required.

Curiously, it appears that Zurek aims to address the question of classicality by relying on tridecompositional uniqueness (i.e. approach C), even though that only seems to apply to the specific case of measurement interactions[17]. Consequently,

---

[17] Zurek does not clearly distinguish between the various selection criteria: the use of the “predictability sieve” seems to imply a focus on approach B, but he repeatedly (esp. (Zurek, 1993,

Zurek has to justify the relevance of the correlation constraint and invoke an additional third subsystem to guarantee the uniqueness of the preferred decomposition of the state. Exactly how this is done, and what Zurek's motivation for this kind of approach is, will be one of the questions that will concern us in the next chapter.

Dynamical constraints

There is one remaining aspect of the selection criteria that might worry us: Zurek repeatedly claims (Zurek, 1982, 2003, 2007) that according to his selection criterion, the preferred basis is independent of the initial state. And indeed, his selection criterion makes reference to the $\mathcal{SA}$-interaction only. On the other hand, according to the tridecompositional uniqueness theorem, the preferred states are fixed by the final state $|\Phi_{\mathcal{SAE}}\rangle$. So, when the interactions are fixed, the preferred states come to depend on the initial state explicitly. How are these two approaches to be reconciled then?

The solution is simple: in the decoherence models discussed here, the interaction Hamiltonian is constructed to be of the form $\hat{H}_{\mathcal{SAE}} = \hat{H}_{\mathcal{SA}} + \hat{H}_{\mathcal{AE}}$. In Zurek's (1982) model of quantum measurement, these interactions operate in succession. In particular, it is assumed that the $\mathcal{SA}$-interaction terminates after a certain time t_m. (See, for instance, section II.A. of Zurek (2003) for some examples of measurement models with an explicitly defined interaction time.) It is easy to see that, together with an assumption of perfect measurement, this always results in a tridecomposition with for the $\mathcal{A}$ states the (apparatus) eigenstates of the apparatus-environment interaction. It is therefore this specific form of the interaction Hamiltonian that results in a final state that can be tridecomposed, with the same apparatus states, for any initial state. (The different final states only differ in the coefficients in the expansion.)

Yet, this also shows how misleading the idealizations of decoherence models are, or at least the simpler ones that are used to discuss the measurement problem. For instance, all models assume that the apparatus starts out in a pure state (i.e. it does not start interacting with the environment until the interaction with the system has finished)[18]. But the assumption that real measurement interactions satisfy these and the above constraints, is actually quite unrealistic. Relaxing the constraints on the interaction Hamiltonian $\hat{H}_{\mathcal{SAE}}$, for instance, obviously has two unpleasant consequences: first of all, for real measurements, usually imperfect (see section 3.3.3 below), the final state may not have a tridecomposition. And second, even when it does, it will generally lead to a completely different pointer basis (see Donald, 2004). This means, at best, that the tridecompositional uniqueness theorem is irrelevant for realistic measurement situations.

On the other hand, Zurek's alternative, the predictability sieve, seems ready-made for the non-ideal nature of real measurements and interactions. Only in the ideal

---

1998)) emphasizes the importance of correlations, which indicates a preference for C, but without mentioning a third subsystem. Often he refers only to the "stability" of the pointer states, yet without providing an explanation of what happens to superpositions of the latter. See my discussion of Zurek's "existential interpretation" in the next chapter for details.

[18] I am not sure what the consequences for the tri-decomposition are if one relaxes this requirement. For the more general case of the emergence of classicality, the assumption seems to be rather harmless; at least, I interpret the decoherence argument as saying that if one assumes an open system to be in a superposition of preferred states, this situation will not last long.

case of the simplified dynamics used to model the measurement process (as in Zurek, 1981, 1982, 2003) will the sieve select the exact pointer states. Otherwise, when a more complicated interplay of interactions between $\mathcal{S}$, $\mathcal{A}$ and $\mathcal{E}$ is involved, the sieve will yield a hierarchy of more or less stable states (Zurek, 2003). However, it is as yet unclear what imperfect stability means for the interpretation and relevance of the decoherence process for the measurement problem. Zurek acknowledges that there may not be an exclusive pointer basis, but he remains silent about the consequences this has for the preferred basis problem. Perhaps more importantly, the motivation for adopting purity or "stability" as a selection criterion through the equivalence of A and B to criterion C has now gone awry: if the tridecompositional uniqueness theorem does not apply, we have come back full circle to the criteria A or B as "bare" interpretational rules.

There are at least three lessons one can draw from the foregoing discussion: first of all, any interpretation based on decoherence requires an assumption that the dynamically preferred states define the interpretation basis. Furthermore, there are at least three selection criteria A-C available, but these are not fully equivalent. And finally, selection criterion C, although in my opinion the most natural and the one having the greater explanatory power, at least with respect to the decomposition version of the Preferred Basis Problem, has severe limitations: the tridecompositional uniqueness theorem on which it rests will generally fail to define a proper preferred basis (indeed, any basis at all) for the more realistic cases.

I close this section by posing the most prominent questions that remain:

- **Selection:** What kind of selection criterion (A, B or C) for the preferred basis should one adopt?
- **Interpretation basis:** In the light of result 3 of section 3.2.1 ("Einselection -general"), how do the mixed states acquire their empirical meaning from the preferred states (as defined by one of the selection criteria A, B or C)?

Of course, the second question may (but need not) be addressed through the answer to the first question. I state them separately because both aspects of the problem should receive some attention.

3.3.2 The ignorance interpretation (revisited)

I have just argued that the argument by which the decoherence theorists aim to motivate the alleged physical relevance of the preferred basis are unsatisfying, so that an explicit interpretational rule is required to justify the assumption that the preferred basis determines the empirical content of arbitrary (mixed) states. If such a rule would work, it would have the merit that it would invalidate at least one of the objections against the ignorance interpretation, namely, that the decomposition of the density matrix is not unique. What we would obtain, in effect, would be a *weak* version of the ignorance interpretation without its main defect: it is not the (diagonal) density matrix *alone* that determines the ensemble of measurement outcomes to which the ignorance interpretation applies, since the additional information required to tell what the possible (pure) states of the system are, resides in the dynamics (more precisely, in the apparatus-environment Hamiltonian).

But although an answer to the question labelled "interpretation basis" above would solve the basis ambiguity, it does not necessarily mean that the weaker ignorance interpretation (in the above sense) is therefore valid and sound. And as I stressed previously (cf. section 1.2.2), the ignorance interpretation (or a reasonable alternative) is essential to solve the problem of Outcomes, since the mere fact that the density matrix is diagonal in a certain basis cannot explain why one outcome occurs rather than another.

Proper vs. improper mixtures

As explained in section 1.2.2, the ignorance interpretation attempts to blame the probabilistic nature of quantum measurements on epistemic uncertainty, even though the reduced density matrix is an improper mixture, not a proper one. An improper mixture results from taking the partial trace with respect to one of the subsystems of a composite (entangled) state. (In more sophisticated decoherence models, the crucial step of taking the partial trace is hidden in the generalized master equation, but it is there.) The partial trace is merely a calculational tool to predict the expectation values of a measurement performed on one of the subsystems only - in this sense, the reduced density matrix is the restriction of the quantum state to a subsystem. But the flaw with the ignorance interpretation is that it wants to draw an ontological conclusion from this formal construction. It is thus entirely based on the misleading resemblance of the improper mixture (1.12) to a statistical (proper) mixture and a misguided belief in the ontological priority of pure states[19].

Part of the problem is in the e-e link. The ignorance interpretation of the reduced density matrix is clearly motivated by the e-e link, but strictly speaking the e-e link and the ignorance interpretation are incompatible: the total state (including the environment) is still pure, often entangled, and does not correspond to an eigenstate of any physically meaningful observable (let alone the preferred observable)[20]. Even if one does not maintain an ignorance interpretation in the strict sense, but follows Joos in his claim that the system "appears" to be in one of the states on the diagonal (cf. the previous subsection), the argument is still based on the idea that diagonality of the reduced density matrix somehow implies definiteness, which, in turn, is clearly motivated by the assumption that definite properties must be accounted for by the pure states on the diagonal. I will occasionally refer to this idea as a "modified version of the e-e link".

The procedure of taking a partial trace involves ignorance about the traced out degrees of freedom; but it must be stressed that this has nothing to do with ignorance about which one of the elements of an ensemble (which is not there) actually obtains. Yet, a link between these two kinds of ignorance should be familiar from

---

[19] This should do away with claims frequently encountered in the literature that "taking a partial trace amounts to the statistical version of the projection postulate" (Pessoa, 1998, p. 332), or that "[O]ne should always keep in mind that the concept of the density matrix relies ultimately on the assumption that at some stage a collapse [. . .] occurs" (Joos in (Joos et al., 2003, p. 43)). What is meant here, is that the *ignorance* interpretation (and not the formal operation of taking a partial trace) is based on the assumption that one of the terms on the diagonal of the density matrix is actually *realized* with certain probability – this is simply the mistaken identification of proper and improper mixtures.

[20] In fact, for decoherence the problem is even worse because the reduced density matrix is not exactly diagonal. I will discuss this issue in section 3.3.3 below.

classical physics; the result of a coin toss can only be assessed on a probabilistic basis, since the exact conditions (initial velocity and position etc.) are unknown. Averaging over these unknown boundary conditions then yields a probability of 0.5 for each possible outcome, head or tails. It may therefore be tempting to assign the same kind of interpretation to decoherence. However, as Bub (1997) points out, the quantum case is entirely different from the classical (a similar objection is also made by Joos (Joos et al., 2003)): in the case of the coin toss, the outcome could in principle be predicted if all the relevant boundary conditions were known. But in the case of decoherence, knowing the full details of the state of the environment would reveal the interference terms in the global superposition again. Thus, by taking full account of all dynamically relevant degrees of freedom, one passes from a mere probabilistic description of events to a situation *in which there are no events at all.*

Statistical equivalence and empirical constraints

All these criticisms of the ignorance interpretation are well known. From Joos' remark at the start of the previous subsection it is clear that he does not want to appeal to it in a straightforward sense, in any case. But generally, it seems that decoherence theorists are unwilling to reject the ignorance interpretation altogether. So take for instance Zurek (1993):

> "The ability to describe the state of the system in terms of the probability distribution of the same few variables is the essence of effective classicality: it corresponds to the assertion that *the system already has its own state* (one of the states of the preferred basis) which is not necessarily known to the observer prior to the measurement, but is nevertheless, already quite definite" (p. 293, my emphasis).

The 'own state' of the system Zurek is referring to here not only concerns the separability of subsystems of an entangled state (cf. section 3.3.4), but also the assumption that the incoherent mixture resulting from decoherence can (and should) be interpreted as representing our ignorance of the actual state of the pointer. Zurek ignores the difficulties with this view I have pointed out, but (continuing the previous quote) admits that:

> "It is important to emphasize that while this statement is, strictly speaking, incorrect (there is actually a very messy entangled state vector including both the system and the environment) it cannot be falsified by any feasible measurement."

So it seems that for Zurek the point is not about ontology, but about empirical equivalence. This is claimed to be guaranteed because of certain operational constraints on the physically relevant observables. To be precise, the following holds:

> The entangled state $\rho_{\mathcal{AE}} = \sum_i c_i c_j^* |a_i\rangle\langle a_j| \otimes |e_i\rangle\langle e_j|$ is indistinguishable from the state $\rho_{\widetilde{\mathcal{AE}}} = \rho_{\mathcal{A}}^{\mathcal{E}} \otimes \rho_{\mathcal{E}}^{\mathcal{A}} = \sum_i |c_i|^2 |a_i\rangle\langle a_i| \otimes \sum_i |c_i|^2 |e_i\rangle\langle e_i|$ for all observables of the form $\hat{A} \otimes \hat{E}$ with $\hat{A}$ and/or $\hat{E}$ diagonal in the bases $|a_i\rangle$ and $|e_i\rangle$, respectively.

In particular, the two are equivalent if one takes for $\hat{E}$ the identity operator, i.e. ignores the environment altogether.

One could put it differently, and say that the empirical equivalence between proper and improper mixtures that could be taken to be sufficient for an ensemble interpretation, is in fact *violated* for observables that measure certain off-diagonal correlations between the subsystems. However, Zurek's point is that in ordinary circumstances (that is, everyday life as it takes place outside the laboratory), the kind of measurements required to distinguish the the state of a macroscopic system subject to decoherence from a proper one, are not the kind of measurements performed by us, human observers. For pragmatic reasons (these kind of experimental efforts are of no use to ordinary observation) or for reasons of principle (even if we wanted to, these kind of experiments would be impossible to carry out)[21]. So usually it is argued that since, due to decoherence, the state of a macroscopic object becomes entangled with all of its environment, it would be required to intercept all of the environment (that is, the billions of photons that scatter of a macroscopic object) in order to reveal the interference terms. It would definitely be a difficult task to recover interference by such means. A general and rigorous derivation has not been given so far[22], but usually it is argued that such an apparatus would have to be bigger than the known universe or so. Furthermore, what is required is more than 'bare' observation: the experimenter has to establish the correlations between the state of all the photons in the environment and the state of the object. One may perhaps argue that our classical world view can remain intact as long as we are simply not aware of these correlations. After all, there is nothing peculiar about, for instance, the measured values of the spin of a single member of an EPR-pair, as long as we do not compare them to those of his twin brother out there somewhere in the universe.

These remarks indicate that decoherence may in fact have something to offer in support of the claim that the *statistics* of measurement (as prescribed by the Born rule) are properly reproduced without the need for a collapse postulate or an *ad hoc* choice of preferred observable. However, appealing to empirical equivalence and thereby accepting that the ontological claim –that the system *is* in one of the states on the diagonal of the reduced density matrix– is untenable, leaves a crucial gap in the interpretation of decoherence. One in fact acknowledges that the theory cannot account for the occurrence of particular events, or "measurement outcomes". And, as I gave repeatedly pointed out (see section 1.2), this kind of view does not seem to go well with the intentions of the decoherence theorist. In fact, when saying that nothing more than statistical equivalence is needed it seems that the decoherence theorist has given up on his original program (to explain the emergence of classical definiteness) to endorse an ensemble interpretation instead, which takes the occurrence of measurement outcomes for granted without providing an explanation of the latter. The ensemble interpretation may in fact become a good deal stronger in the light of decoherence because an explicit (dynamical) selection criterion would yield a solution of the (statistical version of) the Preferred Basis

---

[21] In the light of the previous subsection, one should perhaps say that all arguments that consider the "off-diagonal" observables to establish the empirical (in)equivalence of the two kinds of mixtures are besides the point. Although I think that this criticism is valid, I will not consider this possibility here. I will take it up in the next subsection, but then applied to the imperfectly decohered density matrix of the system alone. I guess that the argument presented there can be partly extended to the present case in which we are dealing with two correlated systems.

[22] But see Omnès (1992) for a calculation.

problem. But I think that this is only a rather modest contribution for decoherence to make to the measurement problem. (I will elaborate on this in the Conclusion.)

So the picture we arrive at in the end is the following:

- Adopting an explicit basis selection criterion (A, B or C) removes the basis ambiguity, and therefore fixes a preferred representation of the density matrix.
- For a local observer, the improper mixture that is diagonal in the preferred basis will be equivalent to a proper mixture in its statistical predictions.
- Decoherence therefore fits an ensemble interpretation of the quantum state, but it does *not* solve the measurement problem if we take it to be the question of how, when and why particular measurement outcomes come about.

I would like to emphasize that in order to address the latter issue, decoherence would need to embed the formalism in some non-orthodox interpretation of the quantum state. This conflicts with the idea, expressed in chapter 1, that decoherence is a physical mechanism that does not require any additional interpretational rules. But setting aside the difficulties with the diagonal density matrix, the idea is that one only needs to show that typical "non-classical" features of the quantum state (interference) simply do not appear for macroscopic objects. Since the interference terms that prohibit a "classical" interpretation are no longer part of the macroscopic object *itself* anymore, but of the object *together with its environment*, the issue is not about interpretation. Instead, it has become a largely empirical one: the question is whether the theory proves that we can safely, though falsely, believe to be dealing with a proper mixture "for all practical purposes". Therefore, the central remaining questions are:

1. **Interpretation:** What is this additional interpretational rule that bridges the gap between having a diagonal density matrix, and the occurrence of individual events?
2. **Empirical adequacy:** Is the decoherence argument that says that the system-environment state is indistinguishable from a proper mixture, empirically adequate?

An answer to the second question would have to take into account complicated kind of theoretical questions about the practical constraints on observations and measurements. I will make no attempts in this direction. What I will do in the next subsection, however, is to point out that issues of interpretation and empirical considerations are closely intertwined in decoherence.

3.3.3 Approximate decoherence

In section 3.1.1 I noted that there are two different mechanisms responsible for the vanishing of the interference terms: a "locality" condition (point (v) in section 3.2.2) and taking the limit $t \to \infty$ (point (iv)). In the previous section I have dealt with the first of these issues; the objection against the ignorance interpretation is essentially that the mixture obtained after the decoherence interaction is improper because it results from neglecting a certain set of degrees of freedom. (Whether this should literally be understood as a "locality" condition will be discussed in section 3.3.4.)

The second of these points still has to be addressed. It is a crucial (but usually implicit) assumption for environment-induced decoherence that after a very short period of time the environment states can be considered to be orthogonal – even though strictly speaking orthogonality is only achieved in the $t \to \infty$ limit. So the almost-orthogonal reduced density matrix $\rho^{\mathcal{E}}_{\mathcal{SA}}(t)$, for t finite, is replaced "for all practical purposes" by the "decohered" diagonal density matrix $\rho^{\mathcal{E}}_{\mathcal{SA}}(\infty)$, since the difference between the two, contained in the interference terms, is negligibly small. One might wonder, at least I do, whether this assumption is really as harmless as it seems.

It must be noted that my concerns are not primarily directed at the methods used and the results offered by the theory: although grounded in particular models which are often not quite realistic, there is good reason to believe that decoherence is indeed a highly effective process. Calculations, and to a certain extent also experiments, show that coherence drops off exponentially fast[23]. Even if experiments revealing interference of macroscopic objects were feasible, this would not contradict the fact that *for ordinary purposes*, classicality provides a good description of the phenomena. It may be violated in the laboratory, but that is not what we set out to explain.

I will not question these theoretical results. Neither do I object to the assertion that decoherence effects may be manipulated in the laboratory. But I *do* contest the idea that the interference terms present (for finite times) in the reduced density matrix may be neglected on grounds of "empirical indistinguishability". That is, I am willing to accept an argument based on "empirical indistinguishability" to justify the assumption of locality, but not for the problem of approximate decoherence. The crucial difference between the two is precisely that when decoherence is only approximate, the problem of interpretation is not confined to the quantum realm, as was claimed in the closing paragraph of the previous section and in chapter 1, but brought *back* to the level of ordinary, everyday objects. So it seems that I have been jumping to my conclusions too hastily; as already noted at the end of section 3.1.1 the problem of interpreting the interference terms, tiny as they are, *does* pertain to the object itself, and thus needs to be taken care of in the case decoherence is not complete (which for finite times will always be the case).

The failure of the standard argument

There are basically two kinds of claims that depend on the vanishing of the interference terms. The first is that of *statistical equivalence*, as dicussed in the previous section. The reduced density matrix is statistically equivalent to a proper mixture

---

[23] As an example; 'fast' is in the order of $e^{-10^{36}t}$ for a medium sized (10^{-3} m) dust particle in air. Even in absolute vacuum, decoherence due to the 3K cosmic background radiation still takes place at a rate of $e^{-10^{6}t}$. (See for instance Joos and Zeh (1985); Joos et al. (2003).) Of course, time does not only enter via 'pre-decoherence' (i.e. how fast the environment states become orthogonal), but also in the case of (temporal) 'recoherence': in many models (specifically; those with a Hamiltonian with discrete spectrum, cf. chapter 7 of Joos et al. (2003)) of decoherence the damping is not exponential but a periodic function, and recurrence of full coherence after a finite time is to be expected. Calculations show however that these recurrence times are generally longer than the assumed life-span of the universe (e.g. Zurek, 1982).

of the preferred states if and only if the off-diagonal terms vanish (cf. the discussion of the problem of Interference in section 1.1.1). The second is the idea that decoherence implies *definiteness*, which is grounded in an individual interpretation of the quantum state. Neither of these claims necessarily presuppose the ignorance interpretation (cf. section 1.2.1 and 3.3.1). Also note that the criteria A-C for the preferred basis do not in any way depend on the vanishing of the interference terms.

Concerning the issue of statistical equivalence, it is clear that it is sufficient for the interference terms to become negligably small very rapidly. In that case, the exactly diagonal density matrix $\rho(\infty)$ and the corresponding almost diagonal one $\rho(t)$ yield almost identical expectation values for all observables.

However, although this argument is formally correct, I am afraid that it will not work for the decoherence theorist who takes himself seriously. As I argued in the previous subsection, the decoherence theorist seeks a physical motivation for the notion of observable, i.e. in this case an answer to the general version of the preferred basis problem. The consequence is that the decoherence theorist cannot appeal to an orthodox application of the statistical algorithm to argue for near statistical equivalence, for this argument implicitly assumes that one can meaningfully assign probabilities to observables that are not measured. And, strictly speaking, from a decoherence point of view all we have are the preferred states associated with a measurement context (the Born rule simply tells you that probabilities are associated with the coefficients squared). From such a point of view, it simply makes no sense to ground empirical indistinguishability in statistical equivalence of observables that are not specified by reference to the interaction that gives rise to the diagonal density matrix we are trying to establish this empirical equivalence for[24]. So, if one would ground the empirical equivalence of the almost-diagonal and the perfectly diagonal density matrix in mere statistical equivalence, that would mean that we abandon a main achievement of decoherence, namely the dynamical justification for the choice of observable. But as discussed in chapter 1 (section 1.2 and 1.3), the Preferred Basis problem is the central question that needs to be addressed for an ensemble interpretation. And not the problem of Interference!

On the other hand, one could give priority to the problem of Interference; this would boil down to a rule that asserts that the preferred basis is the diagonal basis at each instant of time (taking into account the environmental interaction). However, this alternative interpretational rule (which has been proposed in the context of the modal interpretation[25] has a major flaw: although decoherence models show that the reduced density matrix will become diagonal in the preferred basis, this does not imply that the time-dependent instantaneous diagonal basis will converge to the preferred basis for $t \to \infty$. Namely, in case two or more of the terms on the diagonal are equal, the limiting instanteneous states can be shown to be maximally different from the preferred states defined by the dynamics[26].

---

[24] By this I mean that measuring the interference terms means invoking an interaction that turns the interference terms into the preferred states on the diagonal, i.e. we are in fact talking of a *different* state.

[25] See for instance Bacciagaluppi and Hemmo (1996). Some other authors also seem to take this alternative basis selection rule for the 'natural' interpretation of decoherence, or conflate these two points of view (which are really quite different). For instance, Elby (1994) presents this view as one "about which almost all decoherence-based interpretations agree". (See also my discussion of Wallace's views in section 4.4.2 in the next chapter.)

[26] We say that two bases $\{|\psi_i\rangle\}$ and $\{|\phi_i\rangle\}$ are "maximally different" when $|\langle\psi_i|\phi_j\rangle| = \frac{1}{2}$ for

But perhaps most importantly, as I have repeatedly stressed before (sect. 1.2.3, 1.3 and 3.3.2), it seems that this strong commitment to an ensemble interpretation significantly weakens the alleged relevance of environment-induced decoherence (especially if one abandonds the notion of dynamically defined observables), since the occurrence of facts, and therefore the emergence of classical definiteness, within the quantum formalism is taken for granted by such an interpretation. (I will discuss this in the Conclusion.)

The individual interpretation requires the interference terms to vanish precisely to address the question how definiteness emerges. If no appeal is being made to the statistical argument, it is usually simply claimed that the tiny interference terms are "negligable" and that's it. Clearly, this involves more than declaring the interference terms "meaningless" because they are not being observed or measured (and perhaps cannot be measured in principle), even though this may seem a logical consequence of having the interpretation basis being defined by reference to a particular measurement context (recall my discussion of the ensemble interpretation in the conclusion of chapter 1). If one were to put it like this, such an appeal to empirical constraints would largely undermine the role of the *size* of the interference terms, which is usually considered to be of crucial importance.

The supposition that the vanishing of interference terms implies "definiteness" rests on what I have called the "modified version" of the e-e link. Yet, modified or not, the e-e link is still the motivation behind this approach. So the problem that needs to be tackled is that superpositions of preferred states are "meaningless" or "indefinite" from the orthodox point of view. The idea seems to be that the interpretation of the reduced density matrix as an "apparent ensemble" of some sort is more plausible as the interference terms get smaller, but I think that this is actually unfounded: the size of the interference terms may make a difference for the observable consequences once it has been clearly established what these observable consequences are, but it should be irrelevant for the question whether we are able to attach any meaning to the state of system and apparatus *at all.* (And as such the present problem is rather different from the usual practice in physics, where tiny deviations *can* be discarded on grounds of their irrelevance. That is, even in quantum mechanical applications, where the orthodox interpretation implies that the probability for observing these kind of minor fluctuations is negligable.)

I think that here, we touch upon the dilemma I presented in the closing paragraphs of chapter 1: if one stays entirely within the orthodox-minus-collapse framework, the empirical meaning of the quantum state is that it yields probabilities for measurement outcomes. Yet, at the same time decoherence theorists try to *justify* this interpretation by establishing that at the end of a *real* physical measurement process the system-apparatus couple end up in a state that, in the light of the (modified) e-e link, could indeed be characterized as a set of measurement outcomes with corresponding probabilities. However, we now see that in order to do so, they need to invoke an empirical argument, which is *by definition* formulated in terms of these "measurement outcomes" they are trying to establish. Hence the circularity.

---

each choice of vectors labelled by j, i. For further details I refer to Bacciagaluppi and Hemmo (1994, 1996). Although Bacciagaluppi and Hemmo are able to show that this problem is only essential in the case of exact degeneracy, a more recent analysis of Bacciagaluppi (2000) shows that for a realistic, infinite-dimensional model, his proposal will *generically* produce results which are incompatible with the dynamically preferred basis of decoherence *and* with experience.

In section 1.2.1 I argued that there is no evidence that vanishing of interference terms implies definiteness, but that is not yet a complete refutation. Now, I claim that for *imperfect* decoherence, this claim is untenable. There might still be something to say for it if the off-diagonal terms would vanish completely, but that is not what environment-induced decoherence achieves. The problem of imperfect decoherence is therefore an additional hurdle for interpretations that invoke decoherence, if they are to solve the problem of Outcomes. It is usually assumed that decoherence yields a well-defined set of alternatives, and that an interpretation only needs to explain how one of these alternatives is picked out. I believe that even this preliminary step has not been satisfactorily addressed, but we will see how Everett-like interpretations attempt to deal with it in chapter 4.

Imperfect measurements

All this means that we are left a questionable interpretation of the almost-but-not-completely diagonal density matrix based on a rather vague notion of "empirical indistinguishability". If we nevertheless accept that finite-time effects are not physically relevant, there are more obstacles ahead, independent of the issues of finite time or the finite spatial extent of the observer.

The issue I am referring to is the problem of imperfect measurements. This has been discussed in detail for modal interpretations[27], but not for "orthodox" decoherence. Recall the von Neumann measurement scheme discussed in chapter 1, summarized in equation (1.2). There, we assumed that the measurement was perfect, but perfect measurements are not physically realistic[28]. There are two respects in which the perfect measurement scheme may fail:

1. (*2nd kind measurement*) the states of the quantum system $|s_i\rangle$ may be (unitarily!) disturbed when interacting with the apparatus:

$$\left(\sum_i c_i|s_i\rangle\right)|a_0\rangle \to \sum_i c'_i|s'_i\rangle|a_i\rangle \tag{3.9}$$

   in which the disturbed states $|s'_i\rangle$ need not be orthogonal, or

2. (*erroneous measurement*) the measurement may involve "errors", that is, the apparatus states may couple to the "wrong" system states:

$$\left(\sum_i c_i|s_i\rangle\right)|a_0\rangle \to \sum_{ij} c_{ij}|s_i\rangle|a_j\rangle \tag{3.10}$$

   where $|c_{ii}| \approx |c_i|$ and $|c_{ij}| \ll |c_{ii}|, |c_{ij}| \ll |c_{jj}|$ for $i \neq j$ (small errors).

In fact, formally these two kinds of imperfect measurements are identical. One may rewrite (3.10) as $\sum_j \left(\sum_i c_{ij}|s_i\rangle\right)|a_j\rangle = \sum_j c'_j|s'_j\rangle|a_j\rangle$ with $c'_j|s'_j\rangle = \sum_i c_{ij}|s_i\rangle$.

---

[27] See especially (Bacciagaluppi and Hemmo, 1996)

[28] See (Bub, 1997; Bacciagaluppi and Hemmo, 1996), and in particular (Ruetsche, 1995) for arguments why measurements are generally not ideal.

Despite their formal equivalence, these two expansions invite quite different interpretations of how a measurement can go wrong[29]. But crucially, a "measurement of the second kind" leads to an expansion of the form (3.10), in which the "error terms" with $i \neq j$ need not be small. (As in the case of genuine measurement errors.)

For modal interpretations, the big issue was that the value-ascription rule that invokes the bi-orthogonal decomposition, may result in the selection of the completely *wrong* preferred states in the case of imperfect measurements.[30] "Orthodox" decoherence does not suffer from this problem: the preferred states are selected by means of the apparatus-environment interaction alone, so it does not matter that the system states may have been disturbed, or that there may be a mismatch between system and apparatus states. However, imperfect measurement *does* have the consequence that interference terms in the reduced density matrix $\rho^E_{SA}(t)$ do not vanish, not even in the $t \to \infty$ limit. In the case of erroneous measurements, one may argue that the interference terms are nevertheless small (although they are not a consequence of finite-time effects now, but of imperfections in the $\mathcal{SA}$-interaction Hamiltonian). However, this kind of reasoning will not do if the system states are disturbed through a 2nd kind measurement. Thus, even if the environment states become perfectly orthogonal, the reduced density matrix need not be diagonal[31].

So the conclusions I would like to draw are the following:

- In the context of an ensemble interpretation, the assumption that the off-diagonal terms (small but non-zero at finite t) in the reduced density matrix can be neglected, can only be justified to the extent that one abandons the notion of dynamically defined observables.
- To account for the emergence of "definiteness" (which is a precondition for addressing the problem of outcomes in many decoherence-based interpretations of quantum mechanics) the off-diagonal terms are thrown away on grounds of their empirical indistinguishability. This assumption, however, is unfounded since an interpretation of the (empirical) meaning of the interference terms is still lacking.
- Even granted that interference terms may be ignored when ridiculously small, decoherence fails to account for the case of imperfect measurements, where off-diagonal terms will generally *not* be small or converge to zero for infinite times.

---

[29] One may also rewrite (3.10) as $\sum_j c'_j|s_j\rangle|a'_j\rangle$, but this expansion has no clear-cut interpretation in terms of measurement error. Alternatively, one may consider imperfect measurements as measurements of "unsharp" observables, represented by positive operator-valued measures (POVM's). I will not consider these complications here. I refer to (Bub, 1997, sect. 5.3) and Ruetsche (1995) for details. Ruetsche characterises different kinds of non-ideal measurements and their consequences in a particularly neat way. In particular, it is argued that imperfect measurements may even be a necessary consequence of error-free measurements of certain observables. (In which "error-free" is characterized as the pointer reproducing the right statistics of the quantum system.)

[30] Bacciagaluppi and Hemmo (1996) tried to overcome this problem by means of decoherence.

[31] Of course, when tracing over the system degrees of freedom, the resulting reduced density matrix of the apparatus ρ^{SE}_A *is* diagonal in the preferred basis, but this trivially so if the system observable is taken to be Hermitian. Moreover, even if one would take it as sufficient that the reduced density matrix of the apparatus alone is diagonal, imperfect measurements may invalidate the basis-selection rule due to the possible non-uniqueness and instability of the tridecomposition of the wave function in that case. See section 3.3.1.

These lead to the following question:

- **Imperfect decoherence:** Can an unambiguous interpretation of the almost-diagonal reduced density matrix $\rho^{\mathcal{E}}_{\mathcal{SA}}(t)$ be formulated that is 1) consistent with the dynamical definition of the interpretation basis, 2) empirically adequate and 3) strong enough to deal with the problem of imperfect measurements?

3.3.4 Subsystems and the environment

In section 3.1, I argued that decoherence is manifestly anthropocentric, but that this is less of a problem in decoherence than in the orthodox collapse formalism. Nevertheless, one may wonder whether the observer might not be discarded from the theory altogether (that is, in the terminology of section 3.1, from the reduction *condition*, but of course not from the reduction *criterion*). The notion of "irrelevance", usually invoked to define the environment, seems highly subjective and arbitrary. Perhaps there is a way to make it precise and objective?

The concept of the environment

A closer look at the literature may help to gain some insight into the concept of the environment. In Zurek's writings one encounters three different characterizations of the environment, which are responsible for different aspects of the decoherence process:

1. *The environment as the irrelevant degrees of freedom.* Being "irrelevant" the environment may be left out of the description, expressed by taking the partial trace. In effect, this leads to (almost) *diagonality.*
2. *The environment as those degrees of freedom that interact with the pointer.* If the apparatus interacts with the environment, a dynamically *preferred basis* emerges due to the non-unitary evolution induced by the environment on the local Hilbert space of system and apparatus.
3. *The environment as a source of redundantly stored information.* This is the essence of Zurek's most recent programme of "environment as a witness", or "quantum Darwinism"(Olliver et al., 2004; Blume-Kohout and Zurek, 2005). The main idea is that the environment not only acts as a "sink" of quantum correlations, but is also a *source* of information to the observer. (We do not observe the object itself, but the photons that scatter off). It is believed that this "reliance on second-hand information" leads to *objectivity*, in the sense that different observers can come to agree about measurement results.

The third point of view will be addressed in the next chapter. It is clearly in conflict with the view that the environment is the "unobserved" or "irrelevant" part, but Zurek (Olliver et al., 2004; Blume-Kohout and Zurek, 2005; Zurek, 2007) argues that observation is a matter of a "safe compromise": to intercept enough of the environment to obtain sufficient information about the system, without intercepting so much that this would reveal interference between eigenstates of the preferred observable. I will not question his technical results, but the point is that these kinds of ideas require a reconsideration of what the environment actually *is.* So far, Zurek does not seem to have made serious attempts to reconcile these different points of view.

The first two definitions, although different, are potentially compatible. They both represent essential aspects of the decoherence interaction that are equally important. Actually, the second should be considered as a further condition on the first: as Zurek (1981) argues, there may also be a "rest of the world", which may be left out of the decoherence model altogether. This "rest" does not contribute to the decoherence interaction, or to the appearance of environment-induced superselection rules, in any respect. The crucial point is that there are also degrees of freedom, considered to be "irrelevant" by the observer, that nevertheless interact with the pointer and thus contribute to its dynamical evolution. I will focus on the first of these definitions, but one should keep in mind that there is an additional requirement expressed in the second definition.

It must be noted that the first of the definitions given here still leaves room for two different interpretations. One may paraphrase it as the "internal" vs. the "external" environment view. The latter is closest to classical intuition, and relies on a locality principle: take the environment to be anything that is spatially separated from the system. However, in quantum mechanics intuition often fails, and so it is the case here. The "external" environment view faces a problem that we can paraphrase as follows: decoherence asks us to observe only the system (a certain collection of quantum mechanical degrees of freedom), and ignore its environment. But before decoherence has taken place, there seem to be no definite properties (such as being localized in space) that qualify such a set of degrees of freedom as a system (at least not in the intuitive, classical sense). Depending on which of the selection criteria A-C one adopts, there may not even be a preferred quantity (like position) that is to become definite before a distinction between the system and its environment is drawn. And even if one adopts criterion A or C, which do not invoke a locality constraint to select a preferred quantity, general states will not be eigenstates of this preferred quantity and hence will not have a definite value for this quantity. If the definition of the subsystems *relies* on a notion of classical definiteness, we might wonder whether we are not simply assuming at least part of what we aim to prove. (A similar point is made by Omnès (2003).) In quantum mechanics the idea of a subsystem immersed in an environment is not as straightforward as it might seem from a classical point of view. Although decoherence models generally use a pre-established notion of objects and their mutual interactions, a complete account of the emergent classicality would have to start with a universal wave function and Hamiltonian and identify the subsystems from scratch. This is far from what current models of decoherence are able to achieve.

Alternatively, one may adopt an "internal" environment account which is a more literal, and more anthropocentric, interpretation of the first definition of the environment as "irrelevant". It invokes an idea of coarse-graining as is advocated by for instance Omnès (1992, 2003), who writes:

> "One assumes particularly that a few collective (or relevant) observables can describe the main features of a (generally macroscopic) system, and they are known *a priori.* The system is then split formally into two subsystems: a "collective" one (which is associated with the relevant observables) and an environment, which can be external or in-

> ternal. [...] Observers are supposed to have only a direct knowledge of the collective subsystem. [...] But the question of defining correctly the collective observables for an arbitrary quantum state of the whole system, i.e. to select what is collective and what can be considered as an environment, is much deeper. [...] It represents the real limit of our understanding."(Omnès, 2003, pp. 3-4)

According to this point of view, the environment is interpreted in a broad sense as the microscopic details which are irrelevant for the description of large-scale, or collective, effects that are typical of classical systems[32]. Of course, as Omnès points out, the same danger of circularity as in the "external" account is looming here; we cannot take the distinction between relevant and irrelevant degrees of freedom for granted, if their "relevance" is to be determined by the classical phenomena we seek to derive. However, a possible merit of the coarse-graining approach over the external environment is that the exact choice of coarse graining is (to a large extent) arbitrary, at least conceptually. The role of the observer is to disregard certain degrees of freedom, but the exact choice of coarse-graining is not restricted by any kind of classical preconception. (Of course, it remains to be shown that this arbitrariness does not affect the theoretical results –as suggested by Omnès, this is not so obvious.)

However, for Zurek, this conceptual arbitrariness is precisely the reason for rejecting the view expressed by Omnès:

> "In particular, one issue which has often been taken for granted is looming big, as a foundation of the whole decoherence program. It is the question of what are the "systems" which play such a crucial role in all the discussions of the emergent classicality. [...] Moreover, replacing "systems" with, say, "coarse grainings" does not seem to help at all – we have at least tangible evidence of the objectivity of the existence of systems, while coarse-grainings are completely "in the eye of the observer". (Zurek, 1998, p.22)

His idea seems to be that the concept of subsystems that the external account relies on, is more likely to emerge from a kind of ontological background than are the specific coarse-grainings. At the same time, however, Zurek admits that this ontology is not as clear-cut as in classical physics.

Division into subsystems

A little digression is in order to clarify what exactly is meant here. The idea that the definition of "subsystems" (in particular, the distinction between the to-be-classical system and its environment) poses a severe threat to the conceptual foundations of the decoherence programme is often repeated. This is partly because of the

---

[32] Sometimes Zurek seems to hold the same kind of view:

> "Environments can be external (such as particles of the air or photons that scatter off, say, the apparatus pointer) or internal (collections of phonons or other internal excitations). Often, environmental degrees of freedom emerge from a split of the original set of degrees of freedom into a "system of interest" which may be a collective observable [...] and a "microscopic remainder"." (Zurek, 2003, p. 729)

above mentioned objection of circularity, but also because of the interpretational difficulties associated with the concept of entanglement.

Generally, decoherence will lead to a situation in which system and environment are not in a product state on $\mathcal{H}_\mathcal{S} \otimes \mathcal{H}_\mathcal{E}$. However, there is a conception prevailing in the literature that such a product state is an essential requirement in order to regard quantum systems as ontologically distinct. Joos clearly holds this view. For example:

> "Situations of this kind [i.e. interaction with an environment] lead dynamically to a kinematically holistic, entangled wave function for the entire universe. How can we even find (or define) any subsystems (objects) such as trees or cats?"(Joos et al., 2003, p. 41)

It is mildly paradoxical that in decoherence, entanglement between systems is the driving force behind classical definiteness, whereas at the same time there is more or less a consensus that the existence of entanglement is a major obstacle to a straightforward, "classical" understanding of "systems" in quantum mechanics. The conceptual truggles associated with this double role of entanglement may reveal the key to a deeper understanding of quantum mechanics and its relation to the classical world, but in that case the argument needs some more explenation.

First of all, the problem my be circumvented by *postulating* that states of classical systems are restricted to (and not merely interpreted in terms of) those that are preferred in terms of a particular selection criterion (cf my suggestion in section 3.3.1). If one takes the latter to be criterion A, one finds that the classical states are precisely those states of the system/apparatus that do *not* entangle with the environment. In that case, the preferred states are not only robust (and, ideally, in some respect "classical") but also justify this split into subsystems, because they are simple product states[33].

However, I think that Joos and others misconstrue the problem. Returning to Joos' question, it seems that for him, the definition of what are our objects is to be recovered from the quantum state. To be precise, he *presupposes* a factorisation of the global Hilbert space $\mathcal{H} = \mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{E}$, and notes that the apparatus-environment state is not a product state with respect to this factorisation. However, the real problem of defining subsystems lies precisely in the assumption that a preferred, fixed factorisation of the global Hilbert space exists: in a purely quantum mechanical world, there is no room for the "objects" of our experience, since this notion is based on classical concepts such as localization in space and other definite properties. (Perhaps one could think of a quantum universe made out of elementary particles, but these will never constitute an object in a classical sense if classical definiteness cannot be imposed *a priori*.)

So the real problem is how to justify the split between apparatus and environment. In fact, this split is a necessary requirement in order to establish a preferred basis. I claimed that only selection criterion A requires a locality condition, in the sense that the environment must be "ignored" (traced out) to arrive at a preferred basis. However, also criterion A requires a fixed factorisation of the Hilbert space into an apparatus and an environment part. For *every* account of decoherence, it is at

---

[33] See (Landsman, 2007, sect. 3.4) for a more refined, and somewhat technical, discussion of this point in connection with Raggio's theorem.

least assumed that one can *start* with a concept of "systems", assign a (pure) state to them, and analyze its subsequent dynamics in interaction with its environment. However, if one shares Joos' concerns and deny a fundamental ontological status for the subsystems, one should ask how the overall Hilbert space factorizes in order to assign states to its parts, and without some (ontologically motivated) constraints this factorization is completely arbitrary.

Now, the good news is that the overall interaction Hamiltonian provides exactly the required justification, by assumption (i). We can take the quantum state as ontologically fundamental (which is essential for the Everettian approches discussed in chapter 4), but reject the idea that the state also defines the split between subsystems. Instead, this split is to be recovered from the Hilbert space structure that is implicit in the interaction Hamiltonian.

This way of defining subsystems may possibly fulfil Zurek's ideal of objectivity[34]. What such a physical criterion would achieve is that the only role left for the observer would be his disregarding the degrees of freedom of the environment, rather than *choosing* the degrees of freedom that ought to be ignored. Indeed, from simple decoherence models, it may seem that the desired physical criterion is already available: the decoherence models define the interaction Hamiltonian on a product space $\mathcal{H}_S \otimes \mathcal{H}_A \otimes \mathcal{H}_E$. Taking the Hamiltonian as an ontological primitive, one may recover this particular Hilbert space factorization as the space on which the Hamiltonian takes the simplest form (namely as a sum of three interaction terms), and assume that it (more or less) coincides with the split between system, apparatus and environment.

However, things are not so simple. First of all, one has to assign an ontological role to the Hamiltonians used by the decoherence theorist that exceeds their usual status as merely useful models (assumption (i)). But even if one that as a working hypothesis, there is a second, and perhaps more serious problem this account has to face. First of all, the idea that the Hilbert space structure reflects the distinction between different physical systems seems to be at odds with the fact that the dimension of the Hilbert space is determined by the relevant observables. So physically independent degrees of freedom, such as spatial dimension vs. spin, of one and the same physical system (e.g. an electron), are represented in different Hilbert spaces. Furthermore, it seems unreasonable to demand that there is a strict division into subsystems that coincides precisely with the factorization of the interaction Hamiltonian. In fact, where one puts the boundaries between the subsystems seems largely a matter of convention. And it is by no means guaranteed that shifting the boundaries between $\mathcal{S}$, $\mathcal{A}$ and $\mathcal{E}$ whilst maintaining the same interaction Hamiltonian in the decoherence models, leads to results similar to those previously obtained. In fact, it is unlikely in the light of Donald's (2004) analysis of the existence and stability conditions of "tri-decomposed" states. Donald shows that the tri-decomposition of an arbitrary quantum state (for which such a decomposition exists) will generally not be stable under arbitrarily small variations in the factorization of the overall Hilbert space. These results show that the choice

---

[34] Zurek has made some hints in this direction. For instance, Zurek (1993) remarks that it may be possible to generalize the predictability sieve to select the particular "preferred factorizations" of the overall Hilbert space that, given a particular interaction Hamiltonian, allow for stable states to exists. (A similar suggestion is made by Omnès (2003), and see also (Landsman, 2007, sect. 7.1) for brief discussion and references.) But it seems that recently, he has given up on this program (Zurek, 2007); see the discussion in the next chapter, section 4.2.1.

of factorization can no longer be considered conventional, since it has drastic empirical consequences: the component states on the separate subspaces as defined by the tri-decomposition will be very different for arbitrarily small variations in Hilbert space structure[35]. Hence, three-decompositional uniqueness seems to impose very strict constraints on the freedom one has in defining subsystems, at least for approaches that require assumption (ii). (But a similar objection applies to the bi-orthogonal states of a system interacting with its environment (Donald, 2004).)

The role of the observer

Even if a separation criterion telling us how to distinguish our subsystems can be developed, a remaining question is what makes a particular subsystem "irrelevant" to the observer[36]. There is an answer to this question that can sometimes be encountered in the literature (see especially Landsman (1995)). The argument relates the concept of the environment to the locality of the observer: the environment is taken to be those physical degrees of freedom which interact with a system but are not confined to the spatial vicinity of the observer. Phase relations are dissipated through the environment and will quickly be displaced beyond the reach of the observer. I think, however, that this account is problematic because of the reasons indicated above, but also because, strictly speaking, the assumption that the observer is localized cannot be justified from a quantum mechanical point of view. Although it has the merit that it gives physical insight to the source of "irrelevance" for those degrees of freedom, I think that some philosophical scrutiny is required when taking it as a matter of principle.

Furthermore, in the absence of a better argument for the environment's "irrelevance", the locality principle forces us to adopt the "external" environment view. A separation criterion based on the interaction Hamiltonians may be applicable to both the standard and the coarse-graining approach (setting aside difficulties with the apparent conventionality of the exact location of the split), but only for the standard, external account does one have some kind of justification of the concept of irrelevance.

In the end, the picture that results from my suggestions is the following:

- Adopt an "external" account of the environment based on the locality principle (interpreted literally in a spatial sense).
- The environment interacts with the system, but is otherwise irrelevant to the observer because it cannot be observed: the environmental degrees of freedom are not confined to the observer's spatial vicinity.

---

[35] To be precise: Donald (2004) shows that for any wave function $|\phi\rangle \in \mathcal{H}$, there exists a state $|\psi\rangle = \sum_{k=1}^{K} a_k |\theta_k^1\rangle|\theta_k^2\rangle|\theta_k^3\rangle$ tri-decomposed according to a fixed factorization $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3$ of $\mathcal{H}$ with $|||\phi\rangle - |\psi\rangle|| < \epsilon$ for any $\epsilon > 0$. The same $|\psi\rangle$ can be tri-decomposed according to a different factorization $\mathcal{H}'_1 \otimes \mathcal{H}'_2 \otimes \mathcal{H}'_3$: $|\psi\rangle = \sum_{m=1}^{M} b_m |\zeta_m^1\rangle|\zeta_m^2\rangle|\zeta_m^3\rangle$. However, the component states of these different expansions of $|\psi\rangle$ can be shown to be maximally different, that is $|\langle\theta_k^i|\zeta_m^i\rangle_U| < \epsilon, \forall k, m$ (for an inner product $\langle\ |\ \rangle_U$ suitably defined to accommodate the difference in Hilbert space structure).

[36] In any case, our inability to perceive these degrees of freedom has little to do with their physical status: the particles that constitute the environment can be the same as those of the system. Think also of Zurek's program of "Quantum Darwinism", where the environment is assumed to exist of photons, precisely *because* photons are the carriers of visual perception.

- Adopt an ontology of quantum states and interaction Hamiltonians.
- The latter define a particular Hilbert space tensor product structure, which serves to define subsystems within the global quantum state.
- That way, entangled states can be interpreted as representing distinct systems which are correlated in a non-classical sense, without necessarily having to appeal to an ontological status for the objects themselves.
- Consequently, the system-environment split is represented in the interaction Hamiltonian, and the only role left for the observer is not to observe (Landsman, 1995).

It appears that the problem of defining subsystems and the environment in decoherence is a difficult one to tackle. The approach I have sketched here may not have really alleviated our initial concerns, but at least I am now able to cast them in a number of explicit questions that will be addressed in the next chapter:

- **Environment:** How is the environment to be defined, especially when taking into account the recent idea of Quantum Darwinism?
- **Split:** If there is a strict split between system and environment, ultimately defined by the interaction Hamiltonian, how is this to be reconciled with the apparent (limited) freedom one has in defining the subsystems?
- **Locality:** Why exactly should one consider observables of the form $\hat{S} \otimes \hat{A} \otimes \hat{\mathbb{I}}$ only?

3.4 Concluding remarks

My aim in this chapter was to shine some light onto the sometimes confusing arguments found in the decoherence literature. I have (briefly) discussed the philosophical implications of the decoherence programme, analyzed meaning and validity of the various claims that are made by the decoherence theorists, and made an attempt to uncover the hidden premises.

As for the first part; I have emphasized the close connection in decoherence theory between the observer and the concept of a classical world. This connection seems to be merely pragmatic (classicality as a *useful* description of the world for observers like us), but the limitations to observation that establish this connection act in two different ways (the reduction condition vs. the reduction criterion), and I believe that it is essential to distinguish these mechanisms from one another. In particular, in section 3.3.3 I argued that one *cannot* simply appeal to such empirical constraints in order to get rid of the interference terms in the case of approximate decoherence.

In section 3.2 I have tried to strip the argument from all the technical details and far-fetched claims to its basic assumptions and results. Next, I have analyzed the tenability of these assumptions and claims in the light of the additional complications introduced by a more realistic description. I believe that this analysis shows that many of these claims break down for realistic circumstances.

This is quite typical, since the decoherence literature is abundant with complicated models, aiming to provide a realistic description of the quantum-to-classical transition. It might seem that an oversimplified presentation would lead to unwarranted criticism, and that sophisticated physics is required to give a full account of a solution to the measurement problem by means of a decoherence argument. Nevertheless, my conclusion is rather the opposite. The reason for this discrepancy is the distinction that should be (but usually is not) drawn between the emergence of classicality (or of classical definiteness) and the measurement problem. The sophisticated models in the literature are not supposed to solve the measurement problem, but to show the strikingly classical features of systems interacting with their environments. But when discussing the relevance of decoherence specifically for the measurement problem, one usually returns to the simplified picture (see for instance Wallace (2007)).

My criticism has two sides. First of all, one could of course regard the measurement problem as just a special case of emergence (recall section 1.2.3 in the first chapter). Simply put, the measurement problem is that the "indefiniteness" of the measured object (being in a superposition of eigenstates of the measured observable) gets transferred to the measurement apparatus. So if environment-induced decoherence would account for the "definiteness" of (macroscopic, open) "ordinary" objects, this would solve the problem. However, I find the evidence for the latter claim thoroughly unconvincing. On the other hand, my objections *against* the claim that decoherence implies (the appearance of) definiteness are circumvented in the case of quantum measurements, simply because this particular case allows a more narrow formulation of the problem. It is this more pragmatically formulated problem that the *simplified* decoherence argument is supposed to address. The second part of my criticism consists of the observation that this formulation of the measurement problem, with its particular virtues that allow one to bypass my previous objections, is too much of a theoretical construct to have any connection with reality. Thus, one is facing the task to give an account of the measurement process that is sufficiently realistic, but at the same time preserves those specific features that make it immune to my first criticism. And this, I would argue, cannot be done.

The details are as follows: my objection to the "emergence" argument is that the decoherence theorist fails to explain why the dynamically preferred basis is the interpretation basis. The often repeated claim that superpositions turn into "apparent" mixtures of preferred states is in my view simply false, and the other kind of arguments that I proposed (which always include an additional postulate) seem to lack explanatory power and/or face additional conceptual difficulties. So the upshot is that decoherence does *not* imply classical definiteness. However, the specific case of a measurement interaction has the virtue that it can appeal to the tridecompositional uniqueness theorem to fix a particular *decomposition* of the final (post-measurement) state. (As opposed to a dynamically preferred set of states that have no implications for the decomposition of non-preferred states.) This is what Zurek (1981, 1982) and Bub (1997) consider to be the solution to the preferred basis problem. (In Zurek's later writings the relevance of tridecompositional uniqueness has moved somewhat towards the background, but as I will argue in section 4.2.2 in the next chapter, he still relies on it in order to motivate his selection criterion in terms of "stable correlations".) Unfortunately, the fact that the post-measurement state allows a tridecomposition, in terms of precisely those states that are dynamically preferred (and expected to be classically definite by experience), is a consequence of specific dynamical assumptions which are an artefact of the simplified presentation, and that cannot be met in general.

So if one wants to come up with a more realistic story, instead of tackling a purely theoretical fictional problem, there does not seem to be a particular merit for the specific case of measurement interactions over and above the general issue of emergent classicality. It is worthwhile, therefore, to close with a evaluation of the results achieved by environment-induced decoherence with respect to the latter.

First of all, I agree that there exists a dynamically preferred basis in the sense of points 1 of 2 of section 3.2.1. For this basis to have any particular interpretational relevance, an additional postulate is required. My personal favorite would be to reconsider the status of observables in the light of environment-induced decoherence along the lines sketched in section 3.3.1. This reconsideration is a direct consequence of a new interpretational rule that says that the dynamically preferred basis *is* the interpretation basis. (As I pointed out, however, this approach is not physically or philosophically straightforward, so I remain rather sceptical about this program.) Second, a specific feature of the decoherence models is that the reduced state of the "apparatus" will be (approximately) diagonal in the dynamically preferred basis. Since the decoherence theorists have failed to convince me that this will somehow lead to an "apparent" definiteness of the apparatus in terms of these states, *even* if the preferred basis is defined to be the interpretation basis, I consider this result to be relevant merely from the point of view of an ensemble interpretation, in a very narrow instrumentalist sense. If we are to adopt an ensemble interpretation, however, the selection of a preferred basis is the essential contribution of environment-induced decoherence, and it seems that (approximate) diagonality in that case has lost all its relevance. (Which in a way of course avoids the problem of approximate decoherence.)

There are some remaining issues concerning the definition of subsystems and the role of the environment, which I find more difficult to grasp. Whether one adopts a individual or ensemble interpretation does not seem to make much of a difference here (although one might argue that an ensemble interpretation is sufficiently anti-realistic to be entitled to ignore this difficulty altogether). I have attempted to sketch an approach that defines subsystems in terms of a factorization of the overall Hilbert space that is implicit in the global interaction Hamiltonian, but I doubt that this approach will prove to be useful. Furthermore, the particular physical support for the locality condition (that allows only observables of the form $\hat{A} \otimes \hat{\mathrm{I}}$ on $\mathcal{H}_{\mathcal{A}} \otimes \mathcal{H}_{\mathcal{E}}$) remains unclear. I think that these issues are of crucial importance especially when considering an Everett-like approach to supplement the formalism of environment-induced decoherence. I will return to this in the next chapter.

Chapter 4

Interpreting decoherence: the existential and many-worlds interpretation

In chapter 3 I have investigated what the remaining obstacles are if the measurement problem is to be solved by an appeal to decoherence. The conclusion that decoherence by itself does not solve the measurement problem comes as no surprise; it has for long been known that in order to accomplish such a solution, it should be embedded in one interpretation or another. One of the most popular approaches in this respect is to combine decoherence with (a modified version of) Everett's relative state (or many-worlds) formalism. In the present chapter my aim is to find out whether this approach is capable of answering the other, perhaps less anticipated, questions that were raised in the previous chapter.

4.1 The main issues

The detailed analysis of decoherence in chapter 3 has left us with a number of delicate questions that the standard decoherence formalism could, in my opinion, not satisfactorily cope with. I briefly recall:

1. **Preferred basis:** What kind of selection criterion should one adopt for the preferred basis, and what is the interpretational relevance of this basis?
2. **Outcomes:** How are we to account for the occurrence, or at least our perception, of unique measurement outcomes, given that the state of the quantum system is a reduced density matrix that is diagonal in the preferred states?
3. **Subsystems:** Why does the observer discard a certain set of degrees of freedom, called the environment?

I have stated these questions in somewhat more general terms than in chapter 3, in order not to be too restrictive in my assessment of these matters in the following discussion. An answer to these questions should nevertheless to some extent take the discussion of chapter 3 into account. Some further clarification is therefore in order:

1. The three selection criteria (A), (B) and (C) (see section 3.2 in de previous chapter) are equivalent only under one or more additional assumptions (locality and additional constraints on the dynamics). The choice for a particular selection criterion should respect the constraints imposed by these assumptions. For instance, there are important differences between quantum measurements and the emergence of classicality that one should keep in mind – I am thinking especially about the number of subsystems. (See section 3.3.1.)

2. Within the individual interpretation (which is the one considered in the present chapter), the achievement of decoherence is usually taken to be the fact that environment-induced decoherence gives us a well-defined set of definite facts ("emergence of classical definiteness", cf. section 1.2.3). I have criticised this claim in the previous chapter, but even if one accepts that decoherence would be able to account for definiteness of the events we perceive, it still cannot account for the (uniqueness of the) events themselves. This is considered to be the most compelling reason to embed decoherence in an Everett-like interpretation. Further, such an interpretation should have something meaningful to say about imperfect decohered density matrices (due to finite-time effects or non-ideal measurements).

3. In an interpretation of quantum mechanics based on decoherence, the concept of "environment" should acquire a natural and unambiguous meaning. Most importantly, it should give an account of how to define the "subsystems" in an ontology that is based on the assumption of an universal wave vector.

The problem of Outcomes is obviously the most pressing, and as such it has always been the main, if not the only, motivation for adding a many-worlds interpretation to the theory of decoherence. The many-worlds point of view is a natural choice, since the problem with the decohered density matrix is that all "events" are still simultaneously represented in the state; and the many-worlds point of view in fact *denies* their uniqueness. (In a sense to be made more precise below.) So as a first step towards a solution one could adopt an MWI-like view that says that all the terms on the diagonal of the density matrix represent actual occurring events, as perceived by different observers in different realities. Yet, there are subtle issues involved which do not always seem to be clearly recognized by those who support such a kind of view. Furthermore, prominent advocates of the decoherence-based MWI (such as Zurek and Wallace discussed here) do not support this "straightforward" application. Instead, they argue that an MWI-like view is in a sense implied by the mechanism of decoherence.

The other main questions I raised did receive some attention in the decoherence literature, but not much, and usually they are not taken up when it comes to an Everettian account of the reduced density matrix. I believe, however, that no interpretation of decoherence can be satisfactory if it does not address the other issues mentioned above.

Further, it must be noted that the relation between decoherence and the many-worlds interpretation is usually considered to be one of mutual support. Of course, if the many-worlds interpretation did not have its own internal problems, there would be little use in invoking decoherence to solve the measurement problem. So adherents of the many-worlds view have for a large part embraced decoherence as a welcome addition –especially as an alleged solution to the notorious preferred basis problem. Although the present discussion is not intended as an investigation of the viability of the many-worlds interpretation (but of decoherence), the issue will receive some attention through my assessment of question 1.

4.2 Zurek's "Existential Interpretation"

As a first attempt to address the questions raised above, I will discuss Zurek's own interpretation of decoherence and its relevance for the measurement problem. This interpretation partly overlaps with the many-worlds view, but I will largely follow Zurek's own presentation of the "existential interpretation" (Zurek, 1998, 2002, 2003) as an interpretation in its own right and postpone a discussion of the many-worlds interpretation to section 4.3. Unfortunately, Zurek's arguments can sometimes appear to be rather confused and even inconsistent, but, nevertheless, I think that they have some interesting features. So what will follow is mostly my personal attempt to distill an argument from Zurek's writings –I might be mistaken in thinking that this is what Zurek is trying to say, but for me this is the clearest way to understand what he is saying.

4.2.1 Zurek's argument

According to Zurek

> "The interpretation which recognizes that decoherence and environment-induced superselection allow for the *existence* of states at the expense of the superposition principle is known as the *existential interpretation.* It accounts for the inability of the observers to "perceive" arbitrary superpositions. The (relatively) objective existence of the records is a precondition for their classical processing and, therefore, for perception" (Zurek, 1998, p.21)

Some clarification is in order.

1. First of all, when Zurek speaks of the "existence of states" he neither refers to "existence" nor to "states" in the usual sense of the word.[1]
2. Rather, taking his cue from the EPR argument, Zurek identifies "existence", or "reality", with an objective aspect of the world that is not influenced by the observer's intervention. I will call this Zurek's *reality criterion*:

---

[1] Although Zurek (1998) emphasizes the operational meaning of "existence", the existential interpretation appears to have a rather obscure ontology. As I will discuss further in section 4.2.2, there is not always a clear distinction between the epistemological and ontological aspects, and it seems that these qualifications are subordinate to physical processes. See also the quotation on p. 109.

> ***Reality criterion:*** *"A defining characteristic of* reality *of a state is the possibility of finding out what it is and yet leaving it unperturbed."* (Zurek, 1998, p. 5).

Note that "reality" is defined in terms of observation, so although objective, the concept is not observer-independent. (See section 4.2.2 below for discussion and alternative formulations of this reality criterion.)

3. Zurek repeatedly says that where he speaks of "states", he actually means "correlations"[2]. As I will explain in more detail below, this is confusing since two of three possible selection criteria for the preferred states, as I paraphrased them in section 3.2, do not refer to correlations between states at all. I will try to respect the distinction between the different selection criteria. (Again, see section 4.2.2 for further discussion).

Note that the "reality criterion" in point (2) should *not* to be understood as a selection criterion for the *preferred* (pointer) states, since the latter are usually determined by the interaction with the environment. Rather, I think Zurek's argument is supposed to be something like the following: 1) define a "criterion of reality" and 2) a selection criterion for the preferred states and 3) show that the preferred states are "real", i.e. satisfy the criterion of reality. With the addition of a few argumentative steps that will be considered below, one could interpret this kind of argument as a justification of the epistemic priority of the preferred states (i.e. that the preferred states form the interpretation basis, cf. section 3.3.1).

A suggestion that this is the kind of argument Zurek has in mind is in (Zurek, 2007). On p. 24, he says:

> "To exist, a state must, at the very least, persist or evolve predictably in spite of the immersion of the system in its environment. Predictability is the key to einselection. Objective existence requires more: it should be possible to find out a state without perturbing it -without threatening its existence."

Thus, Zurek draws a distinction between "existence" (referring to dynamical preference) and "objective existence" (referring to stability under measurement or observation)[3]. Here, the term "objectivity" refers to the program of Quantum Darwinism that I will discuss below, but which is not really relevant at the moment. Zurek does not make precise what is meant by "finding out" and "perturbation", but, in any case, if one rejects the orthodox point of view "observation" cannot be an interpretational primitive (at least not if also identified with a physical effect). Therefore, I think that here, it refers to a specified measurement interaction. In the same vein, I think that "perturbing" should *not* be interpreted as a consequence of the collapse postulate. (For a suggestion of the correct interpretation, see section 4.2.2 below.) Furthermore, Zurek mentions that a state must "evolve predictably"

---

[2] For instance, on page 14 of (Zurek, 1998): "we shall use a shorthand, talking about states, while the real story is played out at the level of multipartite *correlations*."

[3] Note that Zurek does not say what happens to the states that do not "persist" (i.e. superpositions of the states that do): apparently, they do not "exist", but I take this to mean that they should be interpreted in terms of what does exist, i.e. the preferred states. Also note that, according to this definition, the Universe cannot exist.

when interacting with the environment. Despite the confusing reference to "existence" here, I assume that this means that Zurek adopts selection criterion A, which I repeat here for completeness[4]:

> ***Selection criterion A:*** *The preferred states of $\mathcal{A}$ are those that do not get entangled with the environment under the interaction $\hat{H}_{\mathcal{A}\mathcal{E}}$.*

In conjunction with the locality condition, this selection criterion implies that the preferred states of the apparatus evolve unitarily (Zurek says "predictably"), in spite of the latter being immersed in the environment.

So the question is: why are the preferred states "real" (in the sense of Zurek's criterion)? In other words: *why are the states that do not entangle with the environment, insensitive to measurements or observations?* The first thing to note is that taken by itself, the reality criterion does not say that a quantum state needs to satisfy any additional requirement, such as being subject to environmental interactions, in order to be "real". Under plausible interpretations of Zurek's use of the phrase "perturbation" (either as orthodox collapse or analogously to the selection criterion, see below) it seems to imply that any quantum state can in fact be "real", as long as it is an eigenstate of the measurement interaction. So the reality criterion does not seem to imply that states of quantum systems that are isolated from an environment (i.e. for which no preferred basis in the above sense has been singled out yet) are not "real", but it does seem to imply that their status as being "real" is context-dependent. (However, neither Zurek nor the other decoherence theorists spell out what the consequences of his "reality" or "existence" criterion for the interpretation of the states of isolated quantum systems are.)

But what does this imply for open (effectively classical) systems that interact with an environment? In what sense are the states that are preferred in the sense of the selection criterion more "real" than other quantum states? Note that the selection criterion *alone* will not do: when the environment is traced out, the resulting state of $\mathcal{A}$ will be a mixed state diagonal in the basis of pointer states –but in no way does that imply that the state will remain "unperturbed" when measured in a *different* basis (neither in the orthodox sense, nor in terms of a coupling with the state of some external system). Of course, the mixed state will remain invariant when measured in the basis in which it is diagonal, but that is not different from the case in which the system does not interact with an environment at all.

However, to Zurek it is precisely the *possibility* of "non-perturbing" measurement which makes the einselected (preferred) states more "real". (See the reality criterion above!) The interaction with the environment has the fortunate consequence that the observer can always obtain the information required to perform a non-disturbing measurement in the einselected basis. For, Zurek argues:

> "The observer can, for example, measure properties of the Hamiltonian which generates evolution of the system and of the environment. Einselection determines that pointer states will appear on the diagonal of the

---

[4] It must be stressed that Zurek is unclear about the selection criterion he adopts. At various places (including his latest papers) he hints that the "uncontrollable disturbance" by measurements is to be understood in the orthodox sense of a collapse. Zurek (2007) even adopts a criterion of non-perturbing measurement, which I can make no sense of at all. It is therefore questionable whether it is Zurek's argument I am paraphrasing here. In any case, I think it is the clearest.

> density matrix of the system. Hence, the observer can know beforehand what (limited) set of observables can be measured with impunity. He will be able to select measurement observables that are already monitored by the environment. Using a set of observables co-diagonal in the Hilbert space of the system with the einselected states he can then perform a non-demolition measurement to find out what is the state without perturbing it. A somewhat indirect strategy which also works involves monitoring the environment and using a *fraction* of its state to infer the state of the system." (Zurek, 1998, p.5)

Zurek does not explain how the observer can "measure properties of the Hamiltonian"; in fact this idea is not mentioned in his later writings. But the second suggestion –that the information about the preferred pointer observable is copied to the environment– is central to the idea of Quantum Darwinism, which occupies a prominent position in Zurek's more recent publications (Olliver et al., 2004; Blume-Kohout and Zurek, 2005; Zurek, 2003, 2007). The main results of this programme are: 1) that the dynamics responsible for decoherence can copy information about the system to the environmental degrees of freedom, 2) that only information about the preferred pointer observable can be stored *redundantly* in the environment, and 3) that the size of the fragment of the environment intercepted by the observer does not matter for the amount of information about the system acquired this way (except in the "complete information limit" where phase relations become manifest again or in limiting case where the observer simply does not observe enough to obtain any useful information about the system).

Zurek thus invokes Quantum Darwinism in order to establish the "objective existence" of the einselected states. As said, I read this argument as an attempt towards a justification of the claim that the preferred basis is the interpretation basis. A similar idea is expressed, although somewhat differently, by Zurek (2007) when he says that

> "Relative states [i.e. correlations between system and observer] can exist objectively, providing that the observer will only measure observables that commute with the pre-existing mixed state of the system (e.g. in the wake of decoherence, its pointer observable). *But why should the observers measure only pointer observables?*" (pp. 14-15, my emphasis).

According to Zurek, "Quantum Darwinism provides a simple and natural explanation" (*Ibid.*). Since, he argues, we acquire information about the external world indirectly, by intercepting photons from the environment, the results of Quantum Darwinism show that "perception of classical reality seems *inevitable* for observers who –like us– rely on the second-hand information, on the correlations acquired indirectly, from the environment"(Zurek, 1998, p.15, my emphasis). Thus, Zurek's point is that not only do we, human observers, always have the *possibility* to "find out" what the state of a system is without perturbing it (since we know what the preferred basis is); in fact, we *cannot do otherwise* but to make our observations, the human analogue of measurements, in the basis of preferred states.

Exactly how this mechanism is supposed to work will be the subject of a critical analysis in the next section. But taking for granted that Quantum Darwinism would indeed be capable of answering Zurek's question (why the observers measure only pointer states), all this does not yet tell how to account for our perception of unique and definite measurement outcomes, i.e. the transition from a diagonal density matrix to one of the "alternatives" that the states on the diagonal represent. In order to address this issue, Zurek (1998) generalizes the decoherence model and the above argument to apply to the memory of the observer. Admitting that "such "anthropic" attributes of the "observership" as "perception", "awareness" or "consciousness" [...] at present, cannot be modelled with a desirable degree of rigour" (p. 17), he nevertheless believes that the solution to the problem of (apparent) collapse will be found in the specific physical features of the information processing in the observer's brain. The first thing Zurek notes, is that

> "the ability to process information concerning states of objects external to memory (for, say, the purpose of prediction) is then based on the stable correlations between the record bits of memory and the state of the object. [...] For the reliability of memories, is it absolutely crucial that this correlation be immune to further correlations with the environment, i.e. to decoherence"(p. 17).

So he applies the decoherence model of measurement from his (1981) paper to the observer; just as a measurement interaction *must* couple (arbitrary) system states to the (einselected) preferred states of the apparatus if these correlations are to be immune to decoherence (see 3.2.1), so *must* the observer's memory "be stored in robust record states"(p. 19) if his experiences are to remain correlated with the state of the external system that is being perceived.

However, observers are not simple measurement apparatuses. In fact; "there is [...] one feature distinguishing observers from the rest of the universe: They are aware of the content of their memory" (Zurek, 2003, p. 762)[5]. Because "the observer is what he knows" (*Ibid.*), i.e. the information stored in his brain is part of the physical state of the observer, Zurek believes that this ability of introspection leads, in effect, to the appearance of the collapse of the wave function:

> "The perception of the unique measurement outcome is caused by the "redefinition" of the state of the observer, by the change of the observer's identity [...] This resetting of the observer's identity is ultimately the infamous "collapse" (Zurek, 1993, p. 311).

I think the idea behind this statement is the following: given a preferred basis $\{|a_i\rangle\}$, and the fact that the state of $\mathcal{A}$ is a mixture diagonal in this basis, one concludes that $\mathcal{A}$ is in some sense "definite" for the observable defined by the states $\{|a_i\rangle\}$, i.e. it is claimed that the states on the diagonal provide definite and mutually exclusive alternatives. Zurek recognizes that the observer's brain, where the perception of the state of $\mathcal{A}$ actually takes place, is also subject to decoherence. In other words; the state of the observer's brain also represents a set of mutually exclusive and definite *experiences*. Since the observer's self-consciousness is contained within

---

[5] Although Zurek makes explicit mention of the observer's ability of introspection, it is unclear whether this really plays an essential role in his argument. Zurek emphasizes that "decoherence treats the observer as any other macroscopic quantum system" (2007, p.17) and so it seems that he does not claim that it is this special (unphysical) ability of introspection that would result in a real collapse of the wave function in the sense of, for instance, von Neumann.

these alternatives, he cannot be aware of this multiplicity[6]. This idea is not new (an earlier formulation, without decoherence, is at least already implicit the writings of Everett – see section 4.3.1), but for Zurek the mechanism of decoherence is an essential part of his argument; it is the disappearing of interference terms that allows us to interpret the state of $\mathcal{O}$ as a set of mutually exclusive experiences.

All in all, it seems that Zurek does not want to introduce any (explicit) additional interpretational axioms to account for the (apparent) collapse. As he puts it in his latest review paper: "We have been led by quantum formalism to our conclusions, but these conclusions are largely beyond dispute. Our "existential interpretation"' is in that sense not an interpretation – it simply points out the consequences of quantum formalism and some rudimentary assumptions" (Zurek, 2007, p.25)[7]. It remains to be seen whether he really succeeds; a critical analysis of his argument and the nature and tenability of the "rudimentary assumptions" will be the topic of the next section.

4.2.2 Commentary

Zurek's "existential interpretation" leaves quite a few questions open. In particular, it does not address the questions raised at the beginning of this chapter in a straightforward manner, but I will try to shed some light on its possible relevance for to these issues. Since Zurek's wording can be rather ambiguous, I will provide a number of different readings of his argument. Yet, in my opinion, in each of those readings his argument suffers from serious problems that I think are fatal.

The role of the observer

The first thing to note is that the third question that I posed in section 4.1, i.e. what constitutes the "subsystems", or what it is that makes the environment "irrelevant", is not addressed by Zurek's argument; Zurek (1998) clearly states that he considers this to be an "open problem" (p.2, see also his remark in my section 3.3.4). It is noteworthy that, although previously Zurek (1993) has suggested that a procedure similar to the predictability sieve may help define subsystems in the overall Hilbert space, in his latest review paper Zurek seems to have given up on this program. There, he adds an axiom (0) to the standard postulates of quantum mechanics that reads:

(0) *The Universe consists of systems* (Zurek, 2007, p.3)

I am not sure whether this postulate is as "uncontroversial" as Zurek claims it to be, but it will serve us well as a working hypothesis[8].

With this in mind, I will now review what Zurek's argument has to offer for the first two questions ("preferred basis" and "outcomes") above. Although Zurek does

---

[6] Zurek's (1993) paper is the place where he makes his view the most explicit, but definitely not the only one. For instance, in (Zurek, 2003), he makes a similar remark: "The collapse is a consequence of einselection and the one-to-one correspondence between the state of the observer's memory and of the information encoded in it" (p. 710).

[7] It must be noted, however, that elsewhere in this particular paper Zurek explicitly acknowledges the need for such an additional interpretative framework. In particular, he adopts Everett's "relative states" formalism, which I will introduce in the next section.

[8] See section 3.3.4 for a discussion of the problems associated with this issue.

not clearly indicate what an "interpretation" based on decoherence is supposed to achieve, I read the "existential interpretation" with a special focus on these two questions. In particular, I interpret his "reality criterion" as an attempt to justify the special epistemic role of the dynamically preferred basis (in the sense that this basis should be taken as the interpretation basis.) A more critical look at this criterion would therefore be appropriate.

First of all, I think that the "reality" criterion (at least in the formulation adopted here) is not really satisfying. Zurek's definition does not seem to bear much relation to the usual understanding of "reality". In particular, it is not clear what his argument means for the ontology of the micro realm. As I pointed out in the previous section, according to this criterion, the "reality" of a quantum state is context-dependent, i.e. determined by the measurement interaction. Zurek acknowledges this, and calls the quantum states of isolated systems therefore not "real" but "epistemic". His claim is that this epistemic status acquires an ontological meaning only when an interaction with the environment is involved. (For instance (Zurek, 2002, p.22): "Quantum state vectors can be real, but only when the superposition principle [. . .] is "turned off" by einselection. [. . .] Hence, the ontological features of state vectors – objective existence of the einselected states – is acquired through the epistemological "information transfer".) Zurek may be satisfied with this identification of epistemology and ontology, but I am not; I take it merely to mean that all ontological claims in Zurek's interpretation eventually refer to an epistemic preference. In fact, what his reality criterion at the start of the previous section seems to say, is that the true "reality" of isolated quantum states is not *knowable* because knowing implies disturbance; but it does not say that isolated quantum systems do not *possess* an intrinsic reality. (After all, one might say that something should possess some putative physical status if it is to be disturbed by a *physical* interaction.) In that case, it is completely obscure to me what the ontology of Zurek's interpretation is.

But apart from this somewhat philosophical nitpicking, a perhaps more pressing question is what Zurek actually means by "disturbance" of a state through measurement, assuming that he does not refer to the collapse postulate[9]. When discussing the selection of a preferred basis through environmental interactions, Zurek (2007) gives one of his few hints. He says that "[the eigenstates that form the] pointer basis entangle least with the environment (*and, therefore, are least perturbed by it*)" (p.2, my emphasis)[10]. This is an almost literal statement of selection criterion B. If "no entanglement" is equivalent to "no perturbation", it is tempting to interpret the reality criterion in the same way. This would mean that *a state* $|s_i\rangle$ *of* $\mathcal{S}$ *is "real" if it does not entangle with the state of the apparatus* $\mathcal{A}$ *to which it is coupled by the measurement interaction.* This will be the case if the measurement interaction is diagonal in the basis to which $|s_i\rangle$ belongs. Note however, that it makes sense to interpret a superposition $\sum_i c_i|s_i\rangle$ of "real" states as being "perturbed" by the

---

[9] Often Zurek claims to "explain" collapse, or addresses it through an Everettian approach (see (Zurek, 1998, 2003, 2007) and section 4.3.1 below.) But I must say that I am not convinced that Zurek really rejects the collapse postulate. In (Zurek, 1998, 2007) he repeatedly speaks of "repreparing" states through measurement in a way that seemingly refers to the collapse postulate, although he does not make this explicit.

[10] It must be noted, however, that elsewhere in the same paper, Zurek seems to refer to a simple unitary evolution when speaking of "perturbation". On p. 4 he calls states $|s_k\rangle$ "unperturbed" when they remain identical when "measured" by the environment: $|s_k\rangle|\epsilon_0\rangle \rightarrow |s_k\rangle|\epsilon_k\rangle$. I can make no sense of this in the light of his proposed interpretation, so I will ignore it in what follows.

interaction with $\mathcal{A}$ only if the apparatus states are traced out: the fact that the $\mathcal{SA}$ state $\sum_i c_i|s_i\rangle|a_i\rangle$ is entangled alone does not imply a kind of "perturbation". (Not in my view, at least.) However, since this notion of "perturbation" seems to be the only viable option in the present context, I will nevertheless adopt this interpretation of Zurek's reality criterion to see where it will lead.

A further difficulty with the "reality criterion" as it was formulated in section 4.2.1, is that Zurek repeatedly suggests that it is not the stability of the *states* that is essential, but the stability of the *correlations* between them. For instance:

> "[S]tability of the correlations between the states of the system monitored by their environment and of some other "recording" system (i.e. an apparatus or a memory of an observer) is a criterion of the "reality" of these states. Hence, we shall often talk about *relatively objective existence* of states to emphasize that they are really defined only through their correlations with the states of the other systems [. . .]" (Zurek, 1998, p. 4, emphasis in the original).

This statement is slightly obscure, but it is clarified when we realize that the "reality" criterion does indirectly refer to the correlations between $\mathcal{S}$ and $\mathcal{A}$: if an arbitrary state is expanded in the bases $|s_i\rangle$ and $|a_i\rangle$ that "do not entangle" with each other, the result will be a correlated state (bi-decomposition) $\sum_i c_i|s_i\rangle|a_i\rangle$. The point is that Zurek is referring to the environment, which does not play a role in the reality criterion itself. So let me try to make this somewhat more precise.

Although the selection criteria A, B suffice to select a unique bi-decomposition (because the preferred states are uniquely determined by the dynamics), the crucial point remains that $\mathcal{S}$ and $\mathcal{A}$ are to *remain* correlated in case of interaction with an environment. (See section 3.2.) For ordinary quantum measurements, this means that the $\mathcal{SA}$ interaction must be such, that the $|s_i\rangle$ states couple to the preferred apparatus states, i.e. the states $|a_i\rangle$ that do not entangle with $\mathcal{E}$. The result will be a tri-decomposition $|\Phi_{\mathcal{SAE}}\rangle = \sum_i c_i|s_i\rangle|a_i\rangle|e_i\rangle$[11]. As a consequence, the selection criterion C based on the tri-decomposition, will in this ideal case select the same preferred states as A or B.

So, essentially, this connection between the selection criteria A, B on the one hand, and C on the other, is precisely what one obtains when combining the Zurek's selection criterion with his "reality" criterion. Formally:

$$|s_i\rangle \text{ is "real"} \quad \Longleftrightarrow \quad |s_i\rangle \rightarrow |s_i\rangle|a_i\rangle \text{ for certain apparatus state } |a_i\rangle \tag{4.1}$$
$$|\tilde{a}_i\rangle \text{ is preferred} \quad \Longleftrightarrow \quad |\tilde{a}_i\rangle \rightarrow |\tilde{a}_i\rangle|e_i\rangle \text{ for certain environment state } |e_i\rangle. \tag{4.2}$$

Here, the apparatus state $|a_i\rangle$ in (4.1) and the environment state $|e_i\rangle$ in (4.2) are determined by the interaction. In the context of quantum measurements, the requirement that the $|s_i\rangle$ states must remain "real" as $\mathcal{A}$ interacts with $\mathcal{E}$ implies that the $|a_i\rangle$ states in 4.1 should be identical to the $|\tilde{a}_i\rangle$ states in 4.2. (But note that, actually, this kind of argument only applies to the specific "real" respectively "preferred" $\mathcal{S}$ and $\mathcal{A}$ states; it does not tell what happens to their superpositions unless one postulates that all states be *decomposed* in terms of those that are "real" or "preferred".)

---

[11] Recall that this follows directly from the fact that $\mathcal{S}$ and $\mathcal{A}$ are correlated, and the inclusion of a third system $\mathcal{E}$; no specific assumption about the environment states is needed.

So I would claim that the "reality" criterion, that speaks of "perturbation" under measurements, should be understood as a requirement that the states of the measured system $\mathcal{S}$ do not entangle with those of the measuring apparatus $\mathcal{A}$. I have also shown that in case of measurements, the requirement that the "real" states should persist in spite of the interaction with the environment, is equivalent to the existence of a tri-decomposition of the $\mathcal{SAE}$ state, which is unique. But now, consider Zurek (1993):

> "The most obvious of these [aspects of environment-induced decoherence] is the need for stability of the correlations between the preferred sets of states selected by the processes of decoherence occurring simultaneously in the memory and in the system, and the fact that memory is a physical system subjected to decoherence inasmuch as any other macroscopic [...] open system" (p. 309).

So, Zurek wants to extrapolate the measurement model to the observer perceiving any open quantum system (i.e. acting as an apparatus for the latter). This means that we have to replace the isolated quantum system $\mathcal{S}$ in the above analysis by an open system $\mathcal{A}'$ with associated states $|a'_i\rangle$, and the apparatus $\mathcal{A}$ by an observer (Zurek often speaks of "memory") $\mathcal{O}$ that "measures" (observes) $\mathcal{A}'$. The situation becomes slightly different in this case:

$$\begin{align} |a'_i\rangle \text{ is "real"} \quad &\Longleftrightarrow \quad |a'_i\rangle \rightarrow |a'_i\rangle|o_i\rangle \text{ for observer state } |o_i\rangle \tag{4.3}\\ |a'_i\rangle \text{ is preferred} \quad &\Longleftrightarrow \quad |a'_i\rangle \rightarrow |a'_i\rangle|e_i\rangle \text{ for environment state } |e_i\rangle. \tag{4.4}\end{align}$$

Thus Zurek now intends to prove that the *preferred* states $|a'_i\rangle$ of $\mathcal{A}'$ (which is now the *measured* system) are "real" by virtue of their being stably correlated to the observer states $|o_i\rangle$. As I will now show, this slight change of focus has important consequences which Zurek seems to ignore.

Zurek's intuitive idea (or at least, as I interpret it) is simple. Zurek requires that $\mathcal{A}'$ and O be stably correlated; he justifies this requirement by appealing to a kind of "psycho-physical parallelism" that says that correlations enable us to trust our memory records "for the purpose of prediction", as Zurek puts it[12]. I think this is reasonable[13]. Formally, this requirement translates to a tri-decomposition of the final system-observer-environment state, analogous to the measurement model:

$$|\Psi_{\mathcal{AOE}}\rangle = \sum_i c_i|a_i\rangle|o_i\rangle|e_i\rangle, \tag{4.5}$$

with $|a_i\rangle, |o_i\rangle$ and $|e_i\rangle$ preferred states in the the apparatus, observer and environment subspaces, respectively. The interpretation of the state (4.5) is that the

---

[12] However, Zurek does not always say very clearly what he considers to be an essential requirement for perception. Usually, he talks about "predictability" in a sense that seems to be synonymous with unitary evolution (i.e. the preferred states (according to criterion B) are "predictable"), and suggests that this immediately implies that other states cannot be perceived (presumably meaning that other states obtain their empirical meaning from the preferred states). To me this is not so clear. Elsewhere, however, Zurek has argued that predictability is the analogue of measurement reliability, for "correlations between observers and the to-be-classical systems" (Zurek, 1993, p.290), so I believe that the present argument should be considered as the idea behind Zurek's intuition.

[13] Although it seems that this kind of justification at least presupposes a preferred basis.

observer has observed the observable with the $|a_i\rangle$ as eigenstates[14].

Note that the fact that the environment states become orthogonal is irrelevant to the interpretation of (4.5); what counts is that taking the environment into account secures that this correlated decomposition is *unique*. This is not different from the case of "ordinary" measurements, however, so it will not concern me. (Not at the moment, at least.) Second, Zurek considers the observer to be an *open* quantum system. This causes a problem for the characterization of the 'environment' that I adopted previously (section 3.3.4), namely, that the environment is that part of the universe that is not observed, i.e. does not interact with the observer. Likewise, one may ask what the meaning of "irrelevance" or "ignorance" is, if the observer is part of our model.

A further, and probably most pressing problem specific for the existential interpretation is that introducing the observer and the proposed interpretation does not reach its intended goal; in particular, I believe that the assumption that the final $\mathcal{AOE}$ state is of the form (4.5), and thus that there can be "stable correlations" between $\mathcal{A}'$ and $\mathcal{O}$, proves to be unreasonable. To see why, I summarize the differences between the situation of a quantum measurement and that of an observer registering the state of an open quantum system in table 4.1[15].

These differences have obvious consequences for the results that can be derived and for the interpretation one would like to attach to those results. Two points in particular deserve attention:

1. Implicit in the argument that *perception* leads to a tridecomposition (4.5) is the assumption that one can start with the "pre-selected" preferred states of $\mathcal{A}'$ as the *object* in the measurement interaction. In that case, and in contrast to the *measurement*, both parties (here $\mathcal{A}'$ and $\mathcal{O}$) in the measurement interaction also interact with the environment.

2. In the *measurement* process, a tri-decomposition resulted because certain constraints were placed on the $\mathcal{SA}$ interaction; in particular, that the apparatus part of the interaction Hamiltonian is diagonal in the same basis as the apparatus part of the $\mathcal{AE}$ interaction, and that the measurement interaction stops after a finite time t_m. These assumptions will probably not apply to *perception.*

I will discuss these issues in turn.

To investigate what the consequences are of the first point, one could try to apply the different stages of an ordinary measurement process to this new situation, in which an *open* system $\mathcal{A}'$ has established a correlation with its environment *prior* to its interaction with the observer $\mathcal{O}$:

$$\hat{H}_{\mathcal{A}'\mathcal{E}} \rightarrow \sum_i |a_i'\rangle|e_i\rangle \xrightarrow{\hat{H}_{\mathcal{A}'\mathcal{O}}} \sum_i |a_i'\rangle|e_i\rangle|o_i\rangle \xrightarrow{\hat{H}_{\mathcal{O}\mathcal{E}}} \dots \tag{4.6}$$

---

[14] This is the "qualitative" aspect of observation; one only specifies what observable (set of apparatus states) the observer has observed. In what sense the observer can be said to have "observed" anything specific according to this superposition of different "mental states" is a different question; it is central to the Everettian approach, that I will discuss in section 4.3 below.

[15] In what follows, I will emphasize the phrases *measurement* and *perception* when referring specifically to table 4.1.

Measurement	**Perception**
Measured object: *isolated* quantum system $\mathcal{S}$	Measured object: *open* quantum system $\mathcal{A}'$
Measuring subject: *open* quantum system ("apparatus") $\mathcal{A}$	Measuring subject: *open* quantum system ("observer") $\mathcal{O}$
Selection criterion defines preferred states for $\mathcal{A}$	Selection criterion defines preferred states for $\mathcal{A}'$
Interaction of the form $\hat{H}_{\mathcal{SA}} + \hat{H}_{\mathcal{AE}}$	Total interaction Hamiltonian $\hat{H}_{\mathcal{A}'\mathcal{OE}}$ contains an extra term $\hat{H}_{\mathcal{A}'\mathcal{E}}$ and (probably) is of the form $\hat{H}_{\mathcal{A}'\mathcal{O}} + \hat{H}_{\mathcal{A}'\mathcal{E}} + \hat{H}_{\mathcal{OE}}$.
System-apparatus interaction $\hat{H}_{\mathcal{SA}}$ depends on measurement context ("observer's choice"). Apparatus-environment interaction $\hat{H}_{\mathcal{AE}}$ is fixed.	Total interaction $\hat{H}_{\mathcal{A}'\mathcal{OE}}$ is fixed.
During the measurement, $\hat{H}_{\mathcal{SA}}$ is much stronger than $\hat{H}_{\mathcal{AE}}$. After this brief measurement time t_m, $\hat{H}_{\mathcal{AE}}$ takes over.	Unclear: at least the terms $\hat{H}_{\mathcal{A}'\mathcal{E}}$ and $\hat{H}_{\mathcal{OE}}$ continue to be dominant, and possibly $\hat{H}_{\mathcal{A}'\mathcal{O}}$ as well.

Table 4.1: Measurement vs. perception: physical conditions.

Now, it is not at all obvious what happens when $\mathcal{O}$ also interacts with the environment. Even if one assumes that the total interaction consists of separate terms that interact in succession, one cannot analyze this measurement process of observation in the same straightforward way as one could previously analyze the "ordinary" measurement. Of course, a more refined dynamical picture would be required; but the point is that a proper dynamical analysis would, most likely, *not* result in a tridecomposition like (4.5). Because to achieve that, one has to make the additional assumption that the environment state to which a particular $|a'_i\rangle$ correlates, is identical to the environment state to which the associated $|o_i\rangle$ correlates. That is simply unreasonable. (If this is not immediately clear, consider the model in section 2.2.1: the environment states $|\mathcal{E}_1\rangle$, $|\mathcal{E}_2\rangle$ associated with the two pointer states depend crucially on the coupling constants in the interaction Hamiltonian.) What a more refined picture would show, I suppose, is that the ensuing interaction between $\mathcal{O}$ and $\mathcal{E}$ destroys the correlations between the $|a'_i\rangle$ and $|o_i\rangle$ states, even though the $|a'_i\rangle$ are preferred by the interaction $\hat{H}_{\mathcal{A}'\mathcal{E}}$ according to selection criterion A. The result is the kind of subtle interplay between different aspects of the dynamics that is so often stressed by the decoherence theorists. Clearly, such complicated dynamics does not lend itself to simplified presentations as these (or handwaving arguments such as Zurek's)[16].

---

[16] See for instance my brief exposition of the Zeno effect in chapter 2 (section 2.3.3) for an example of the kind of "subtle interplay" in decoherence models.

So even without establishing a precise result (which would require knowledge of the exact dynamics of the system-observer-environment dynamics) one can reasonably expect that taking both the subject and the object in the measurement interaction to be *open* quantum systems, will lead to additional complications that render the assumption that the final state can be tridecomposed *in the preferred, effectively classical bases defined by an independent criterion in terms of the* $\mathcal{A}'\mathcal{E}$ *interaction*[17] highly implausible.

The second point –and especially its relation to Zurek's argument– is somewhat more complicated. Assuming that the overall interaction separates in three terms, the central question hinges on the role of the $\hat{H}_{\mathcal{A}'\mathcal{O}}$ part of the interaction. In the "ordinary" quantum measurement process, this corresponds to the measurement interaction Hamiltonian $\hat{H}_{\mathcal{SA}}$. In section 3.2, I have argued that, for the latter, the crucial point in Zurek's argument is that a good measurement setup generates an interaction Hamiltonian that correlates the to-be-measured states of the system to precisely those apparatus states that are preferred by the environment interaction. It is essentially the experimental design, or rather the observers *choice* to design it that way, that causes the final $\mathcal{SAE}$ state to have a tri-decomposition that obeys the additional requirements (independence of the preferred basis from the initial state, selecting the desired system and apparatus states). But if one moves on to the more esoteric issue of *perception*, it is not so clear what the status of the corresponding $\hat{H}_{\mathcal{A}'\mathcal{O}}$ interaction is.

Note that the first problem discussed above was about the question whether $\mathcal{A}'$ and $\mathcal{O}$ couple to the same environment states $|e_i\rangle$. The present problem is whether $\mathcal{O}$ couples to those states of $\mathcal{A}'$ that are (dynamically) preferred by the $\hat{H}_{\mathcal{A}'\mathcal{E}}$ interaction, i.e. whether $\mathcal{E}$ and $\mathcal{O}$ couple to the same *set* of $\mathcal{A}'$ states. One could perhaps argue that it is self-evident that $\hat{H}_{\mathcal{A}'\mathcal{O}}$ is diagonal in the same basis for $\mathcal{A}'$ as is $\hat{H}_{\mathcal{A}'\mathcal{E}}$, because $\hat{H}_{\mathcal{A}'\mathcal{O}}$ is a physical interaction like any other, and the decoherentist claims that these are generally a function of position. So both interactions $\hat{H}_{\mathcal{A}'\mathcal{E}}$ and $\hat{H}_{\mathcal{A}'\mathcal{O}}$ would single out the same set of preferred position eigenstates[18].

Unfortunately, this assumption leads to a new problem: if no additional stringent constraints are placed on the form of the total interaction Hamiltonian $\hat{H}_{\mathcal{A}'\mathcal{O}\mathcal{E}}$ and on the initial $\mathcal{A}'\mathcal{O}\mathcal{E}$ state, the final state will usually not allow a tri-decomposition in the preferred states. Apparently, the fact that ideal *measurements* result in a tri-decomposition of states selected by an independent dynamical criterion, is due to some hidden assumptions. As noted in section 3.3.1, I think that the crucial assumption leading to the tridecomposed states in the case of *measurements*, is that the $\hat{H}_{\mathcal{SA}}$ interaction terminates at time t_m, after which the environmental interaction takes over. The interaction time t_m is defined precisely in order to correlate the system states to the preferred apparatus states. (If the measurement interaction were not 'cut off' at this point, the total $\mathcal{SA}$ state would lose its bi-orthogonal

---

[17] I emphasise this because the final $\mathcal{A}'\mathcal{O}\mathcal{E}$ state may be tri-decomposable; as Donald (2004) shows tri-*orthogonal* decompositions are rare, but general tri-decompositions tend to exist. However, in the latter case it is highly unlikely that the unique tri-decomposition will yield the classically definite states of our experience that we so eagerly yearn for. (It works in the measurement model of course, because of the specific dynamical assumptions that apply here, cf. section 3.3.1.)

[18] The essential point is that here, there only needs to be *some* environment state that couples to (does not entangle with) each $|a_i\rangle$, i.e. for the present discussion it is irrelevant what these environment states are. However, above I argued that *each* state $|a'_i\rangle$ would have to couple to the *same* environment state $|e_i\rangle$ as the $|o_i\rangle$ that is correlated to $|a_i\rangle$. In order not to make things too confusing, I will ignore this complication.

form and eventually de-separate again. See also section 3.3.1) This separation of measurement and environment interactions (based on their relative strengths at different stages of the measurement process) has the consequence that the $|s_i\rangle$ states remain essentially fixed as the apparatus interacts with the environment[19]. No such assumption applies in the case of *perception.* In particular, it seems reasonable to assume that the $\hat{H}_{\mathcal{A}'\mathcal{E}}$ interaction continues as $\mathcal{O}$ interacts with $\mathcal{A}'$ (or with $\mathcal{E}$).

Measurement	**Perception**
The final state $|\Psi_{\mathcal{SAE}}(t > t_m)\rangle$ has a tri-decomposition $\sum_i c_i|s_i\rangle|a_i\rangle|e_i\rangle$	The final state $|\Psi_{\mathcal{S}'\mathcal{OE}}(t)\rangle$ does *not* necessarily have a tri-decomposition.
The tri-decomposition fixes the preferred system and apparatus states $|s_i\rangle$, $|a_i\rangle$ according to selection criterion C. The preferred apparatus states coincide with the preferred states as selected by (A) or (B). (Similarly, the system states $|s_i\rangle$ in the tri-decomposition are co-diagonal with the system-apparatus interaction.)	In case $|\Psi_{\mathcal{S}'\mathcal{OE}}\rangle$ *does* have a tri-decomposition, the system and apparatus states thus selected will not be related to another selection criterion. In particular, the states $|a'_i\rangle$ in the tri-decomposition will have nothing to do with the preferred states selected in terms of criterion A or B applied to the interaction $\hat{H}_{\mathcal{A}'\mathcal{E}}$.
The preferred apparatus states $|a_i\rangle$ are independent of the initial state $|\Psi_{\mathcal{SAE}}(t_0)\rangle$	The preferred system states as selected by the tri-decomposition (if its exists) *do* depend on the initial state.
After the measurement has terminated, further time evolution leads increasing orthogonality of environment states $|e_i\rangle$, but does not have any effect on the system and apparatus states $|s_i\rangle$, $|a_i\rangle$	Time evolution will in general affect the state $|\Psi_{\mathcal{A}'\mathcal{OE}}(t)\rangle$ in such a way that it may no longer have a tri-decomposition. And if it does, the component states $|a'_i\rangle$, $|o_i\rangle$ will be highly unstable with respect to this evolution.

Table 4.2: Measurement vs. perception: results.

I summarize my findings in table 4.2. From this I conclude that, for various reasons, for the present model of *perception* Zurek's argument fails because

1. the final state will generally not be tri-decomposable and will therefore not represent the "stable correlations" between $\mathcal{A}'$ and $\mathcal{O}$ that Zurek demands, and
2. even if such a tridecomposition (incidentally) would be possible, it will in general not lead to the classical states (for the system or the observer's mental

---

[19] As noted previously (section 2.2.2), Zurek (1998) acknowledges the idealized nature of the measurement model (although he only refers to the possibility that $\mathcal{S}$ has a self-Hamiltonian), but he does not make clear what the consequences are for the uniqueness of the preferred states, the stability of correlations, or his interpretation in general.

records) of our experience.

The concept of "measurement"

Particularly important in the light of Zurek's argument is that simply assuming that both the apparatus-environment interaction and apparatus-observer interaction select the same preferred pointer basis will not lead to the tridecomposition (4.5) and its alleged interpretation. Not straightforwardly, at least. In fact, Zurek even seems to claim that it is not even necessarily so that both the apparatus-observer and apparatus-environment interaction will be diagonal in the same basis for $\mathcal{A}'$, as is clear from his question quoted earlier: *"why should the observers measure only pointer states?"*

There are basically two ways to understand this question, I think. The first is to interpret the phrase "measure" as referring to a physical interaction. The second is to interpret it in orthodox terms, i.e. as a shorthand for specifying an observable to extract probabilities from the (reduced) state of $\mathcal{A}$. Since the orthodox approach seems to clash with the goal of decoherence, the first of these approaches may seem more likely to be what Zurek has in mind. But if we take his question to heart, this kind of view would imply that the interaction term $\hat{H}_{\mathcal{A}'\mathcal{O}}$ reflects a deliberate intention of the observer to observe the preferred states $|a'_i\rangle$. In other words, the observer *chooses* to observe (that is, correlate to his brain states) precisely the states of $\mathcal{A}'$ that are preferred by the environment interaction. Now, I think (I hope!) that everyone will agree with me that this is plain nonsense. Even if it is not, it simply replaces the orthodox concept of "observable" with a no less obscure notion of "intention" that is supposed to be capable of steering the physical interactions in the world[20].

This probably means that for Zurek, contrary to what one should expect, the phrase "measurement" (as performed by an observer, not by an apparatus) does not refer to an interaction Hamiltonian but to a more primitive, indeed orthodox, concept of "measurement". Considering what Zurek says just before he poses his question (see p. 106), it seems that the problem he is alluding to is the following: the post-decoherence reduced density density matrix of $\mathcal{A}'$ is $\sum_i c_i |a'_i\rangle\langle a'_i|$, with the $|a'_i\rangle$ the preferred pointer states. With some effort, we could consider this as an "effectively classical" or "apparent" mixture of pointer states. But for any basis different from $|a'_i\rangle$ the reduced density matrix still contains interference terms and is not an "apparent mixture" of the corresponding basis in any reasonable sense. Now, if Zurek simply wanted to *postulate* that the preferred basis is the interpretation basis, he would not need to pose his question. Instead, he seems to be concerned with the possibility of applying the statistical algorithm in a basis different from the $|a'_i\rangle$[21]. In order to *explain* why an observer can (or should, or must) only "measure" (in an orthodox sense) the pointer observable on the reduced density matrix, Zurek (1998, 2007) introduces the ideas of Quantum Darwinism.

---

[20] I would not want to argue that there are no intentions. I would like to see that they do not play any role in a physical theory, though.

[21] As a matter of fact, Zurek is concerned with the possibility of disturbing the state of $\mathcal{A}'$ by measurement, and thus violating its "objective existence". But I either have to interpret this as a collapse, or in terms of an explicit measurement interaction with a corresponding Hamiltonian $\hat{H}_{\mathcal{A}'\mathcal{O}}$. The first approach implies the orthodox point of view that is presently under discussion, the second I have just discarded on grounds of its untenability.

Quantum Darwinism

The central idea of Quantum Darwinism is that the observer obtains his information about the observed system through the environment, and that the latter stores only information about the preferred pointer observable redundantly. This gives rise to a number of questions. First of all, as noted previously (section 3.3.4), Quantum Darwinism modifies the meaning of the "environment". As a consequence, one can no longer simply say that it is the part that is being disregarded. In fact, the environment becomes the observer's *source* of information about the system that he is interested in (although he need not even interact with the latter). Zurek (2007) suggests that the "observed fragment" of the environment could be included in the state space of the observed system. But this does not help to explain why the "environment" (as a whole) should be traced out to obtain a reduced density matrix of the system (*excluding* the observed fragment of the environment). In any case, the analysis of decoherence adopted previously (above and in section 3.2) clearly no longer applies.

Second, in what sense does the environment contain "information" about the observed system? Zurek quantifies this information in terms of conditional entropy, but this is simply a measure of the degree of correlation between the (open) system $\mathcal{A}$ and the environment $\mathcal{E}$. So, the observer does not infer directly from the state of the environment what the state of the system is; he needs to know what kind of relationship there is between the environment and the system states that underlies this kind of inference. In any case, if this "transfer of information" merely indicates a correlation, then it is hard for me to see what this additional step in the von Neumann chain accomplishes.

In essence, I think there are at least four ways to read Zurek's Quantum Darwinism argument, but I think the one closest to his ideas (at least as expressed in (Zurek, 2003, 2007; Olliver et al., 2004; Blume-Kohout and Zurek, 2005)) is to reject the idea that there is a direct, one-to-one relation between the state of a quantum system and an observation. Rather, "observing a system to be in a state" would come to mean that the observer can operationally verify, by applying repeated measurements and the statistical algorithm, that the system is in a certain quantum state. This approach adds a new dimension to the concept of "observation" in which it differs from the orthodox interpretation. For the latter, we can "observe" individual members of the ensemble, causing a collapse of the state to the eigenstate corresponding to the outcome that has been found. Therefore, in the orthodox formalism one can never verify the state of an individual system. (Not even by repeated measurements, because each measurement collapses the state to another eigenstate of the measured observable.) This is the problem that also worries Zurek: the state of $\mathcal{A}'$ cannot "exist objectively" because arbitrary measurements will inevitably disturb it, and we will never be able to find out what its state was before the measurement[22]. In the orthodox formalism, one can determine the quantum state of the *ensemble*, however, by applying repeated measurements to the members of the ensemble and infer the state from the statistical information thus gathered. Now, according to Quantum Darwinism, we are in the possession of such

---

[22] As I said, Zurek may well refer to the collapse postulate when speaking of "disturbance by measurement". The "apparent collapse" should in any case not be identified with the transition of the superposition to a mixture of pointer states, since this alone will not affect the measurement statistics for the corresponding observable.

an ensemble, also for the system $\mathcal{A}'$: it is in the environment! Each photon in the environment contains a copy of the state of the system, and we can mess around with these photons as we like, because there are plenty available. I think this also means that Zurek considers "observation" of an effectively classical quantum system $\mathcal{A}'$ to be the result of such a repeated "measurement" and a statistical analysis of the entire ensemble, not as a single operational act. So any observation of an open quantum system is actually a shorthand for such a statistical analysis of the corresponding ensemble represented in a fragment of the environment, that allows one to verify what the quantum state corresponding to this ensemble, and therefore of $\mathcal{A}'$, is.

If this interpretation of the Quantum Darwinism argument does not seem obvious, one should keep in mind that the argument probably rests on an orthodox conception of "measurement" or "observation", as I explained. In particular, Quantum Darwinism is therefore *not* a statement about the form of the *interactions* between $\mathcal{A}'$, $\mathcal{O}$ and $\mathcal{E}$. Zurek's question implies that he does not associate the observer's choice of what to "measure" with a *fixed* interaction $\hat{H}_{\mathcal{A}'\mathcal{O}}$. And what difference would it make to replace the $\mathcal{A}'\mathcal{O}$ correlation by $\mathcal{O}\mathcal{E}$? If $\hat{H}_{\mathcal{A}'\mathcal{O}}$ is not fixed, why should $\hat{H}_{\mathcal{O}\mathcal{E}}$ be? In that case, one can simply repeat Zurek's question "but why should the observers measure only pointer states?" for the photon environment[23].

Similarly, I think that Quantum Darwinism also does *not* simply mean that "we discover states from the imprint they make on the environment" (Zurek, 2003, p.759). To me, the argument that we perceive systems to be in an eigenstate of a certain preferred observable because these are the states that are (redundantly) copied to the environment, seems rather pointless; for Zurek, the problem is that the interpretation basis is not fixed, regardless of what kind of state the (open) system is in. But in that case, the simple fact that photon states are (correlated to) preferred states, has no more special interpretational relevance than the preferred states of the system have of themselves. (Which, if I understand Zurek correctly, is none.) So again, I repeat: "but why should the observers measure only pointer states?"

Finally, I also think Zurek is not trying to say that the environment merely supplies us with the information required to *identify* the preferred pointer observable, so that the observer is able to perform a non-perturbing measurement on $\mathcal{A}'$ (and not on a fragment of the environment!). Even though this is what Zurek (1998) suggests when he says that "the observer can know beforehand what (limited) set of observables can be measured with impunity." (See the previous section.) Such an argument may resolve the issue of "objective existence", in the sense that the observer may *know* how to measure the state without "threatening its existence", but I do not think that it is really convincing: the mere fact that the observer can *infer* from the environment what the preferred basis is, does not make it *necessary* for him to perform measurements in this basis on the system.

These considerations lead me to interpret Quantum Darwinism along the lines suggested above. Zurek does not explicitly say that this argument is what he has in

---

[23] Notwithstanding this objection, Quantum Darwinism might very well have interesting implications for the interactions. In particular, the fact that we do not observe the system directly, but correlate our memory to the environment instead, would imply that the term $\hat{H}_{\mathcal{A}'\mathcal{O}}$ in the overall interaction is effectively zero. The consequence would be that many of the problems associated with equation (4.6) disappear. Nevertheless, this point of view does not solve all of the problems because the case of *perception* is still not equivalent to that of ordinary *measurements*.

mind, but it resonates with his idea that Quantum Darwinism leads to "objectivity", by which he means that "the state of the system can be found out by many observers, who will agree about their conclusions" (Zurek, 2007, p.17). If "finding out" refers to a single "measurement" (in an orthodox sense), this statement is simply false; different observers may perhaps agree on the preferred observable (basis) which is copied to the environment, but it does not imply that they will find the same value for this observable. But if "finding out" is a shorthand for a statistical inference of the whole of the ensemble in the observed fragment of the environment, Zurek may have a point if he says that this enables multiple observers (each in possession of his own ensemble) to infer the state of the system to which it is correlated.

I say *may* have a point, but actually I think that this argument, although it may come close to Zurek's intentions, has major flaws. First of all, I find it troubling that the only way I can understand his argument is from a very orthodox point of view that seems to conflict with the intentions of decoherence. But especially, it is very unclear whether it yields the kind of results we are looking for. Recall that Quantum Darwinism is supposed to explain why the observer "measures" only a particular observable on the (open) system $\mathcal{A}$ (in my words; why there is only one particular interpretation basis that determines the empirical content of the state of $\mathcal{A}$). Zurek argues that "[the environment] can reveal only the very same pointer states $\mathcal{E}$ has helped select". What does this mean if $\mathcal{A}$ is not in an eigenstate of the preferred pointer observable? Zurek is strikingly silent about this, but on p. 18 of Zurek (2007) he points out that the observed fragment of the environment will be in a mixed state diagonal in the same basis as the decohered reduced density matrix of $\mathcal{A}$. So how does this help with our problem? Either one can only infer from the environment that the state of $\mathcal{A}$ is mixed, or one needs some *additional* reason to interpret the mixed states in terms of the states on the diagonal. Thus it seems that one cannot invoke Quantum Darwinism for the latter purpose – in fact, it has merely displaced the problem of interpretation from the system of interest to the environment.

Apart from these inherent obscurities about interpretation, it is absolutely unclear to me how Zurek intends to explain the difference between (probabilistic) quantum measurements and ordinary classical perception, given the fact that each experiment always includes an apparatus which interacts with the environment. In fact, this seems to me a major, though neglected, problem for environment-induced decoherence in general: quantum measurements will yield a mixed state diagonal in the pointer basis for the apparatus, but actually, the measurement interaction is completely irrelevant in this respect. (Recall section 3.2.1: the measurement interaction is only relevant in the light of the tridecomposition of the final post-measurement state.) According to environment-induced decoherence, one obtains the same kind of mixture, the same kind of pointer states, for any open system interacting with its environment. So where does the difference between the quantum and the classical come from, if we take the decoherence point of view seriously? Perhaps a more thorough analysis of the role of decoherence in this process could alleviate my concerns, but in his papers Zurek is silent about this – which makes me rather sceptical about this issue.

Introducing Everett

I would emphasise that it is not the task of Quantum Darwinism to address the issue of Outcomes. But the problem of Outcomes presupposes at least that there is a set of well-defined outcomes to choose from − and it is this precondition that Quantum Darwinism, as I see it, fails to address. That being said, one can consider Zurek's answer to the issue of the appearance of collapse as independent from the Quantum Darwinism program.

Briefly, Zurek's idea is that the memory records of the observer are part of his identity, so that we should consider different records (apparatus-observer correlations) as being perceived by different observers. (This idea is close in spirit to Everett's "relative-state" formalism. See section 4.3.1 below.) Unfortunately, however, I don't grant this argument much credibility, for a simple reason; if the "observer is what he knows", he could just as well be an observer having perceived a (decohered) superposition of measurement outcomes, for there is a state of the observer and his memory content corresponding to this situation[24]. This is of course just another version of the preferred basis problem (see section 4.3.2 below for its role in the Everett interpretation); although for Zurek it is essential that the memory states of the observer are subject to decoherence, the mere fact that the observer's "identities" are subject to decoherence does not really help –one does not have a way to interpret this incoherent mixture without assuming what we set out to prove. Specifically, one cannot say that the states on the diagonal really represent "alternatives" before it has been established what the empirical meaning of the reduced density matrix is.

Instead, as Zurek (2007) eventually admits, the "apparent collapse" should be understood in an Everettian sense (see section 4.3.1 below). This means, in particular, that the terms in the superposition (4.5) must be understood as defining separate "branches" or different realities[25]. This interpretational move does indeed address the issue of the unique measurement outcomes, albeit at the price of a rather far-fetched metaphysics that many (including Zurek) will find hard to accept.

Somewhat prematurely perhaps, I would conclude that if Zurek's true intention is to adopt an Everettian approach, his insistence on the stability of correlations indicates that he wants to interpret the terms in the tridecomposition (4.5) as the "branches" that represent separate "observer identities" or even "realities". Although Zurek does not make this claim really explicit, he frequently suggests this kind of approach[26]. In the (unlikely) case that the tri-decompositions would really

---

[24] For a related and more sophisticated argument along the same lines, but directed against the "bare theory" of Albert and Loewer, see (Bub, 1997, section 8.2).

[25] However, Zurek (2007, p.24) says he is reluctant to take this "Many Worlds mythology" at face value. For Zurek, Everett's main messages are the idea that quantum mechanics needs no modifications to account for (the appearance of) classicality, and the "relative states" concept. The latter implies that quantum states only take values (or acquire "properties") *relative* to an observer state. But I believe that this idea was already implicit in Zurek's adaptation of the measurement process to apply to the observer's perception and in his "reality criterion", and we have just shown that this alone does not resolve the issue of collapse. So to me it is still an open question whether Zurek has really added anything new (an explicit interpretational axiom of some sort) in his (2007) paper that addresses my concerns.

[26] For instance, Zurek (1993) clarifies his ideas as follows: "Information processing which leads to "becoming conscious" of different alternatives will lead to states of computers and other potentially conscious systems which differ in their physical configuration, and which, because their state is a presently accessible partial record of their past history, exists in different Universes, or rather, in

describe all physically relevant situations this approach would have obvious merits: 1) the assumption that the preferred basis is the interpretation basis would acquire a natural justification, similar to the more tangible case of ordinary measurements, 2) the tri-decompositional uniqueness theorem would guarantee that the preferred basis is unique, and 3) the argument would not have to rely on the orthogonality of the environment states to endow the quantum state with empirical meaning, and would therefore be insensitive to the charge of imperfect decoherence. But if Zurek's intention is to interpret the decohered density matrix $\rho_{\mathcal{SA}}$ as a collection of "alternatives" (the states on the diagonal), then the problem of approximate decoherence is as pressing as ever (see section 3.3.3).

Unfortunately, however, I think that the foregoing arguments have shown that one should not rely on the existence of a tridecomposition to explain our "classical" experiences, since it will usually not obtain. Other approaches have been suggested, however, that claim to pose a less strict demand on the exactness of decompositions and orthogonality of the environment states. Especially David Wallace has championed this kind of view, which relies more heavily on the many-worlds metaphysics and less explicit on the formalism of decoherence. After outlining the general features of Everettian (or Many Worlds) interpretations of quantum mechanics and the objections that have been raised against it in the next section, I will discuss Wallace's ideas in some detail.

4.3 Everett's relative states, many minds and many worlds

I will start with a brief exposition of the formalism and ideas that lie at the heart of the different versions of what is usually called the many-worlds interpretation (MWI), already mentioned briefly in chapter 1. Although Everett formulated these ideas already in 1957, they have practically gone unnoticed for many years, until from the 1970's onwards the need for an objective interpretation of quantum mechanics became urgent with the rise of quantum cosmology. As a consequence of this increased attention, various difficulties with the interpretation have been identified that I will briefly discuss.

4.3.1 The central concepts

The phrase "relative-state interpretation" is a collective term for a variety of philosophical stances, each of which takes Everett's original proposal as its point of departure, but differing widely in the way his formalism is developed into a full-fledged interpretation of quantum theory. Broadly speaking, three views can be distinguished: the "relational" approaches, many-minds, and many-worlds interpretations. A relational approach is perhaps closest to what Everett himself had in mind (although there still is no agreement on this issue). In this kind of interpretation, one tries to get away with a minimum of additional interpretative rules and to stick to the formalism of quantum mechanics (without the collapse postulate) as it is. Besides Everett's own proposal, the "bare theory" of Albert and

---

different branches of the same universal state vector"(p. 310).

Loewer is another example of such a minimal interpretation. Since the bare theory (and related proposals) are generally considered to be rather unsuccessful in their attempts to make sense of quantum mechanics without state vector reduction[27], I will not consider them here. Neither will I discuss many-minds interpretations in detail, because the differences between these and the many-worlds interpretations can be very subtle but are largely irrelevant for my discussion[28].

Although Everett did not succeed in his original aim of providing a minimal interpretation of quantum mechanics without state vector reduction, his proposal is generally reckoned an "heroic attempt" and his formalism still constitutes the backbone of the many philosophically more sophisticated interpretations it has motivated. The basic idea is as follows:

Assuming that quantum mechanics is a universal theory, Everett postulates the existence of a "universal state vector". He retains the whole of the mathematical machinery of quantum mechanics –including the eigenstate-eigenvector link– but drops the collapse postulate on grounds of its apparent inconsistency: the special axiomatic role it requires for the 'observer' seems to be incompatible with the presumed universality of the quantum mechanical laws. Besides, if an observer external to the system is essential to account for "facts", how are we ever going to apply quantum mechanics to the universe as a whole? (In fact, it was precisely the latter question which led to the revival of Everett's interpretation in the light of quantum cosmology.)

To overcome these difficulties, Everett invokes the notion of a "relative state": although the standard quantum formalism predicts that at the end of a measurement process the system-apparatus (or system-observer) couple ends up in an entangled state, for a *given* state of the system there is a (relative) state of the apparatus or observer to which it is correlated. Hence, Everett concludes:

> "Thus, each element of the resulting superposition describes an observer who perceived a definite and generally different result, and to whom it appears that the system-object state has been transformed into the corresponding eigenstate. In this sense the usual assertions of [the collapse postulate] appear to hold on a subjective level to each observer described by an element of the superposition." (Everett, 1973, p.10)

Central in Everett's exposition therefore is the idea that one does not need to account for definite *facts*, but only for the definite *appearance of facts* to the observer. (Who is now treated on equal footing with any other kind of quantum object.) In the words of Everett "We are [...] led to the novel situation in which the formal theory is objectively continuous and causal, while subjectively discontinuous and probabilistic"(Everett, 1973, p.9).

To see how all this works formally, consider again the measurement of spin in a

---

[27] See the for instance the criticism of Bub (1997).

[28] Simply put, many-worlds interpretations postulate a "splitting" of the actual world, whereas many-minds interpretations require only that the "mind" of the observer splits, in order to explain why only one of these alternatives is experienced. See Barrett (2003) for a brief discussion of these and other stances one may take with respect to Everett's proposal, including the bare theory. Barrett also considers the "consistent histories" approach of Gell-Mann and Hartle as a relative-state interpretation, but their mutual relation is subtle and since the consistent-histories approach does not invoke Everett's formalism, I will only mention it in passing. (See the discussion about Wallace below.)

Stern-Gerlach device as discussed in the previous chapter. (Instead of an apparatus we now consider an observer who registers the result of measurement.) An observer $\mathcal{O}$ who sees the spin $\mathcal{S}$ with orthogonal states $|\uparrow\rangle$, $|\downarrow\rangle$ has an experience represented by the states $|O_\uparrow\rangle$ or $|O_\downarrow\rangle$, corresponding to the state of the spin. Starting out with the spin in a superposition state $|\phi_S\rangle = c_1|\uparrow\rangle + c_2|\downarrow\rangle$ leads to an observer in a superposition of "brain states":

$$|\phi_{\mathcal{SO}}\rangle = c_1|\uparrow\rangle|O_\uparrow\rangle + c_2|\downarrow\rangle|O_\downarrow\rangle. \tag{4.7}$$

According to the eigenstate-eigenvalue link of orthodox quantum mechanics, this state does not assign a definite experience to the observer. Everett retains the orthodox e-e link in the sense that values are only ascribed to eigenstates of spin and observer (or apparatus) observables, but he adopts an additional interpretative rule which Everett (1957) calls "the fundamental relativity of states": although the formalism does not predict any definite *value* for a measurement outcome, the observer registers a definite value *relative* to a particular value of the spin. (And vice versa.)

Everett shows that such 'relative states' can be defined for any state of a composite system (e.g. also in the case of non-ideal measurements). As already noted in chapter 1, an arbitrary state on the system-observer state space $\mathcal{H}_1 \otimes \mathcal{H}_2$ can be written as

$$|\Psi\rangle = \sum_{ij} c_{ij}|\psi_i\rangle|\phi_j\rangle, \tag{4.8}$$

for an arbitrary choice of orthogonal bases $\{|\psi_i\rangle\}$, $\{|\phi_j\rangle\}$ on $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. For any $|\psi_k\rangle$ on $\mathcal{H}_1$ one can define a relative state $|\phi^k_{rel}\rangle$ on $\mathcal{H}_2$ by taking:

$$|\phi^k_{rel}\rangle = N\sum_j c_{jk}|\phi_j\rangle, \tag{4.9}$$

(in which N is a normalisation constant) such that

$$|\Psi\rangle = \sum_k |\psi_k\rangle|\phi^k_{rel}\rangle. \tag{4.10}$$

Note that this definition is basis-independent and that the relative states need not be orthogonal. (If both sets of basis vectors are orthogonal, as in the case of an ideal measurement, the relative-state expansion equals the Schmidt-decomposition.) With this definition the relative-state formulation yields the same statistical predictions as the orthodox formalism[29]: conditional upon the result of a measurement on $\mathcal{H}_1$, say $|\psi_k\rangle$, the expectation value of a subsequent measurement of any quantity $\hat{A}$ on $\mathcal{H}_2$ is given by the relative state $|\phi^k_{rel}\rangle$:

$$\langle\Psi|\hat{P}_{\phi_k}\hat{A}|\Psi\rangle = \langle\phi^k_{rel}|\hat{A}|\phi^k_{rel}\rangle \tag{4.11}$$

4.3.2 Why the Everett interpretation does not work

The strategy of Everett faces serious objections. The most obvious one is that it is not clear what this formalism is supposed to achieve. The notion of "relative

---

[29] That is, if one assumes the Born rule (i.e. interprets the coefficients as probabilities for measurement outcomes). This assumption is problematic however, as will be explained in some detail below.

state" does not explain how a measurement device records a definite value, nor how an observer acquires definite experiences. After all, measurement or memory records are only determinate *relative* to a certain determinate value of the spin, and since there is nothing in Everett's formalism to account for the latter in the light of the superposition (4.7), it is unclear what this concept of relative states is supposed to accomplish. In order to address this issue, several strands of interpretations have emanated from Everett's formalism. Perhaps the best known is the "Many Worlds" interpretation (MWI) popularised by Bryce DeWitt, John Wheeler and David Deutsch, among others. Following similar suggestions of Everett, deWitt (1973b,a) suggests to take each term in the superposition (4.7) to represent an *actual* world in which both the spin and the apparatus' pointer take a definite value. Note that by taking this idea seriously one is committed to a radical metaphysics: not only do the spin and apparatus or observer somehow "split" (again, in a sense to be made precise) into different-valued copies, but with them also the rest of the universe, and not only at the instant of measurement, but in fact each time a superposition state, of some preferred observable, appears *somewhere* in the universe.

It must be noted that modern versions of the MWI, which will be my subject in what follows, are a bit more cautious in their terminology and metaphysics. Everett's relativism is still at the forefront of most interpretations based on his formalism, but to make this philosophically concise, modern Everettians adopt the concept of "branch" which can be regarded as a kind of parameter relative to which observables take values (or objects have properties). This parameter can be understood in a way similar to the concept of time (Butterfield, 2001). All the branches together comprise the single Universe (also called Multiverse by some) in which all values (or measurement outcomes) are realised. Definiteness (dictated by the e-e link) is only secured at the single-branch level[30].

But even if its extravagant ontology does not keep one from adopting the many worlds view, there are also concrete technical difficulties the theory has to face. In the first place, and this is a problem I will be primarily concerned with in the rest of the chapter, many-worlds interpretations suffer from the preferred-basis problem. Since the relative states need not be orthogonal, *any* composite state has an infinite number of arbitrary expansions of the form (4.10). Everett defines the relative states in terms of primitive notions as "observable" and "measurement", but if one wants to avoid such *ad hoc* criteria then it is not clear what determines the interpretation basis[31]. And if one emphasises the definiteness of our experiences, in contrast to the simpler case of apparatus' pointers, it is not at all obvious what such a preferred quantity should be. Consequently, discussions in the current literature go far beyond the usual formalism of quantum mechanics. Some claim that the problem involves "quantum mechanics, ontological questions concerning the philosophy of mind, and epistemological questions concerning the nature of our best physical theories"(Barrett, 2003, p.8). These authors (of which Wallace will be the

---

[30] Although MWI's based on decoherence may loosen this requirement of definiteness somewhat – see below.

[31] Note that the relative state is not just the state of a single observer, but actually represents the whole of the 'rest of the universe'. The preferred basis thus not only has to account for determinate experiences of the observer (or values of the pointer), but also for all other observers, apparatuses, objects (or whatever is supposed to acquire definite values or properties) in the universe. In the case of N observers or apparatuses, one should factorise the total wave function into $N+1$ components and pick a basis that ascribes a definite value to each in each branch.

main subject in the discussion below) also invoke more sophisticated arguments, often based on decoherence, to justify their choice of preferred basis, or to argue that such a strict division is not essential. In the same vein, they defend a notion of an 'indefinite' macrorealm. Allowing such indefiniteness, they need not require that the preferred basis is a simultaneous eigenstate of all 'macroscopic' objects (having determinate properties) in the universe. (See also Butterfield (2001).) Having a physical mechanism like decoherence defining the branches also affords a way to understand why there is no "splitting of the wave function" for genuine quantum systems, or in other words; why the superposition principle ever made sense in the first place. I will consider the preferred basis problem in the MWI and the role of decoherence in more detail when I discuss the views of Wallace and Zurek below[32].

Second, the formalism does not provide a criterion to trace the branches in time as the universal state evolves; they are only defined at a given instant. The time evolution of the universal state is reflected in the coefficients of the component states (the "weights" of the branches), whereas the component states themselves are time-independent. Thus, there seems to be no room for dynamical evolution at the level of components (each of which representing some 'actuality') (Wallace, 2002). Apart from this, there is a metaphysical concern of how one should understand the notion of transtemporal identity of objects in an Everettian universe. Modern Everettians (like Wallace) hold that there is exactly one object associated with each branch[33]. This raises the question as to what happens to the objects within the branches when the branches split. There are various ways to formulate a concept of transtemporal identity in the Everett interpretation, but none seems to be satisfying. Particularly problematic is the point that for any exact choice of branching, objects should be understood as a collective feature of many branches. (See the discussion of Wallace's views below and in Butterfield (2001).)

Last but not least, as a consequence of the previously mentioned problem, the Everettian strategy suffers from a serious problem with statistics. The point is that it is unclear how Everett's formalism is supposed to relate to the usual probabilistic predictions of quantum mechanics. In the MWI formalism, the Born coefficients appear as "weights" attached to the branches defined by the components of the wave function. It seems that - without adding some additional interpretative rule to the formalism - these weights have nothing to do with the relative frequency of measurement outcomes. In fact, the whole concept of probability is problematic in an MWI: if everything will happen, how can it be that, nevertheless, everything happens with a certain (non-trivial) probability? Viewed from within a particular branch, each term is either realised as a measurement outcome or it is not, and thus should be assigned probability 1 or 0. To make sense of non-trivial probabilities, one should have some notion of transtemporal identity for observers across branches.

---

[32] It must be noted that if one is willing to adopt a more prominent role for the observer's experience in an Everettian framework, the preferred basis problem can, to a certain degree, be circumvented. This is the strategy of the many-minds interpretations. These interpretations secure the definiteness of our experiences, and of our experiences *only*, by appealing to a strong mind-body dualism. The preferred quantity is some quantity defined on brains, but it is argued that this does not represent a *physical* fact. Rather, it determines the relation between the physical world and mental processes. (See Barrett (2003) and Butterfield (2001) for a brief, respectively very brief, discussion.)

[33] One may drop the assumption that each branch is "inhabited" by an object, but that would be against the spirit of Everett's proposal and leave us with the good-old indeterminism of orthodox quantum mechanics. Alternatively, one may ascribe to each branch a whole population of objects, but this is a view which has now largely been abandoned. (Cf. Butterfield (2001)).

Moreover, even if there was a way to define temporally extended worlds (existing across the points of branching) the Everett interpretation seems to imply that any sequence of measurement outcomes is realised in some world. Thus, if one is to associate the weights to relative frequencies, one must be sure not to be living in one of the 'maverick' worlds which produces only non-typical sequences. (Sure, these maverick worlds have measure zero – but to interpret this as meaning that they have vanishing probability of occurring would be begging the question.)[34]

In spite of these difficulties, the MWI has gained substantial attention and support, from physicists but most of all from philosophers. Consequently, the true believers have spent so much time and effort on making sense of the many-worlds concept, that it may seem that more ingenious arguments in its favor have been developed than for any other interpretation. These arguments often invoke a decoherence-based solution to the preferred basis problem. On the other hand, decoherence theoreticists often appeal to a many-worlds or many-minds metaphysics in order to formulate a suitable interpretation of the decohered density matrix that would be capable of solving the measurement problem.[35] The result is that both decoherence and relative-state interpretations, in particular the MWI, have now come to dominate the debate on the interpretation of quantum theory to such an extent that it led Bub (1997) to call both jointly the 'new orthodoxy'. I cannot possibly cover all aspects of the current debate here, but in the following section I will concentrate on David Wallace as an author who defends the MWI by appealing to the results of decoherence.

4.4 A modern Everettian account of classicality

4.4.1 Wallace and emergent reality

In two papers, Wallace (2002, 2003) has developed a particular version of the MWI in which decoherence as well as a "structural" view on macro-ontology play a crucial role. In a recent paper, he characterises this view as follows:

> "It remains to consider whether multiplicity does indeed emerge from realist reading of the quantum state, and if so how. The 1990's saw an emerging consensus on this issue [. . .]: the multiplicity is a consequence of decoherence. That is, the structure of "branching worlds" suggested by the Everett interpretation is to be identified with the branching structure induced by the decoherence process. And since the decoherence-defined branching structure is comprised of quasi-classical histories, it would follow that Everett branches too are quasi-classical. It is important to be clear on the nature of this "identification". It cannot be taken as an additional axiom (else we would be back to the Many-Exact-Worlds theory); rather, it must somehow be forced on us by a realist interpretation of the quantum state" (Wallace, 2007, p.32).

---

[34] I cannot go into this issue here, but an extensive critique of early attempts to ground a statistical interpretation of the wave function in an MWI is Kent (1990), and a recent review from a modern perspective can be found in Wallace (2007).

[35] In contrast to Zurek, Zeh has always been very explicit about his Everettian inclinations, cf. (Zeh, 1971; Joos et al., 2003; Camilleri, 2008)

For the careful reader, this quotation contains three significant statements about interpretation: 1) that the interpretation basis of the quantum state is the dynamically preferred decoherence basis, 2) that the components of the wavefunction, when expressed in this basis, should be interpreted as simultaneously existing "worlds", and 3) that these interpretations are not introduced *by definition*, but follow naturally from the physics and a realist interpretation of the quantum state (which, in this case, is an individual interpretation.)

These features make Wallace's account a particular interesting one to consider in the light of the criticisms of the previous chapter, since it has close connections with the ideas often expressed by the decoherence theorists (an example being Zurek's existential interpretation discussed in the first part of the present chapter); namely, that quantum mechanics does not require an (explicit) interpretation to make sense of our experiences. In what follows I will give a condensed presentation of Wallace's argument that deals with the first two claims (in respect of the third) above in turn. A more critical analysis, relating his account to the three questions of the start of this chapter, will be presented in the next subsection.

The first issue is the preferred basis problem. As I explained, this is one of the major obstacles for an Everett-like interpretation of quantum mechanics. How does Wallace address it? Essentially, by declaring the question irrelevant:

> "Defenders of an Everettian viewpoint face a dilemma: just how seriously are we to take these worlds [i.e. branches]? On the one hand, if we were to take them literally and build them into our formalism then we would face the preferred-basis problem at its worst. Decoherence would be no use to us, for we would need an *exact* world-defining principle and not some pragmatic criterion. [...] On the other hand, if we banish these worlds from our formalism we must answer the criticism that our theory is just uninterpreted mathematics, and explain how sense can be made of the universal state.[...] Everettians can avoid this dilemma by a compromise: they may legitimately and meaningfully use the terminology of many worlds without being required to represent these worlds in their formalism."(Wallace, 2002, p.638)

By abandoning the idea that the branches should have an exact formal representation in the theory, Wallace can drop the requirement of a preferred decomposition: the universal wave vector can, in principle, be decomposed in any way whatsoever. As a consequence, Wallace's interpretation comes with a multi-layered ontology (part of which it shares with the more traditional Everettian approach): "on top" there is a single Universe (or Multiverse), which is very remote from our intuitive, classical understanding of the word. It can be described by a single "universal wave function" (one usually assumes that the universe is in a pure state) and thus represents the collection of all possibilities the laws of quantum mechanics ascribe to this wave function[36]. This wave function can be decomposed in various ways. Each possible decomposition defines a different division of the Universe into "systems" (i.e. a specific tensor product structure) and what quantities take definite

---

[36] So this is not to say that the Universe can be anything, since only *one* state in the universal Hilbert space can be the Universe's *actual* state. (On top of this, the universe is of course subject to strict metaphysical constraints, such as that it obeys the laws of quantum mechanics. See Butterfield (2001) for a brief comparison of the Everettian terminology and the concept of "possible worlds" in modal semantics.)

values on those parts. Such a choice of branching thus defines the "aspects of the multiverse". Following the terminology of Butterfield (2001), one may call the instantaneous state of affairs in the Universe from the point of view of such a particular branching a "world"[37]. Within such a world, there is a multitude of "definite macrorealms", corresponding to the different branches. Whereas a specific choice of branching defines what quantities are definite in a given world (say, the position of a pointer), a choice of branch determines what definite value the quantity actually takes (what position the pointer is in).

How are we to make sense of this rather grotesque metaphysics? Wallace claims that any decomposition of the universal wave vector is allowed, but most of these choices will be useless: as Butterfield (2001) describes it "they do not make any empirical facts, whether microscopic or macroscopic, clear or even comprehensible: physicists would call such resolutions 'unphysical' or 'a bad choice of variables'." Wallace's claim is that one may appeal to pragmatic arguments to choose a particular decomposition of the wave function: this choice is, as such, arbitrary. To this end, he draws an analogy with general relativity: just like any foliation of general relativistic spacetime destroys some of the overall symmetry, but nevertheless yields enough information for a complete picture of the world, so does any choice of preferred basis.

The basic entity in Wallace's ontology is the universal wave function. Our down-to-earth, single-branch ontology is merely a "theoretical construct", a way to get a conceptual grasp on the connection between the abstract formalism of Everettian quantum mechanics and our experiences. As Wallace argues:

> "To interpret a mathematically formulated theory as physics we need some way of linking the mathematics to the physical world. At the least, this requires that we locate *ourselves* somewhere in the theory, or else that we have some idea of how the objects and concepts of our everyday world are mapped into the mathematics. We are able to accept that, in the light of new theories, these objects and concepts may turn out to be not exactly what they seem –and indeed some may be completely illusory– but our theory needs at least to make contact with them in order for us to do physics."(2002, p. 639)

Thus, an abstract theory like general relativity or Everettian quantum mechanics need not coincide with our experiences, but it should *encompass* them –perhaps in some approximate sense. And this may well be achieved by purely pragmatic considerations. In fact, Wallace argues, it would be a mistake to believe that the world of our experiences should be considered as a bare fact without anthropocentric connotations; there are no theory-neutral observations, since in order to handle our perceptions conceptually we have to translate them in terms of the "everyday theory", even though this leaves us with a picture which is actually imprecise and slightly distorted. To regard our ordinary perceptions as arising from a pragmatic perspective on the Everettian universe is just an extrapolation of this idea. Wallace argues that the situation for the MWI is not that much different from general relativity: both theories allow us to make sense of our experiences, by acknowledging the anthropocentric element that they are *our* experiences, and not a fundamental

---

[37] But note that this terminology is ambiguous; e.g. Wallace sometimes also speaks of "worlds" when he means branches.

aspect of the theory.

The main motivation for Wallace to abandon the idea of a strictly defined preferred basis is that any such choice will have undesirable physical consequences. After all, quantum mechanics makes sense and the interference between eigenstates *does* have certain well-confirmed empirical consequences, some of which are essential for the very existence of the objects that constitute our everyday reality. For instance, choosing position as the once-and-for-all preferred quantity would lead to the destruction of superpositions of position eigenstates (the terms "go different ways"), which would be disastrous for the chemical bonds that hold molecules together. Similar objections hold for other quantities, such as energy (Wallace, 2002; Butterfield, 2001). The basis-selection rule should therefore be dependent on the the physical constitution of the considered system –and is therefore defined only locally[38].

By now it should be clear how decoherence enters the MWI in Wallace's formulation: it is precisely the decoherence basis which yields the *epistemically* preferable decomposition of the wave function. Recall Wallace's analogy with relativistic spacetime: there, one chooses a particular foliation that is the most convenient one, but which is highly non-unique and which is defined only locally. Some phenomena may require a foliation which does not coincide with our common-sense notion of the "here and now" to make sense – just like typical quantum events (like interference) require a different choice of basis. A different choice of branching is theoretically possible – one is not ontologically committed to the branches themselves – but this will usually be highly unpractical. As Butterfield (2001) puts it: "the system inhabiting a world is not a familiar object –unless the world is defined by one of the (vaguely delimited!) class of 'decoherence bases' ".

Defining the preferred basis being by means of decoherence comes with a certain amount of inexactness, in various guises: because the interference terms never completely vanish (when are they "sufficiently small"?), because decoherence removes interferences only locally, may not select a unique preferred basis, and in general will select some coarse-grained quantity (e.g. the center of mass of a particle) and therefore not specify a basis.

However, none of these points is considered to be a serious drawback, mainly *because* the branching, and with it the objects of our daily experience (that are identified with these branches), are anthropocentric constructs. Wallace (2003) paraphrases his argument as follows:

> "[T]he conceptual criticisms of the decoherence-based approach [...] are based upon a false dichotomy – that either the macroscopic world is written directly in the formalism or it is simply an illusion." (p. 88)

Instead, Wallace argues, the objects of our experience are not one-to-one identifiable with the mathematical entities of the formalism. This specific view on macro-ontology enables Wallace to dismiss the objections associated with the inherent "inexactness" of a decoherence-based approach. His argument against the

---

[38] Here and in what follows, the phrase "local" should be interpreted in a broad sense. It should be contrasted with the idea of a "global" branching, which defines a preferred basis for the total universal wave function. A "local" branching applies to the wave function corresponding to the system of interest only. In general, therefore, the branching will be observer-dependent, since it is the observer who determines what the system of interest is.

"inexactness fallacy" depends on the second claim; that the decoherence "branches" represent multiplicity, and specifically, that this multiplicity is not an additional interpretational axiom. So before considering how this particular view on macro-ontology helps to overcome the difficulties associated with the "inexactness" of decoherence, I first turn to Wallace's argument in support of the second claim. That claim is in fact similar to the one I criticised in section 1.2.1: that decoherence implies definiteness i.e. that it yields a set of well-defined events. (Consider the similarity between what I have called the problem of "emergence of classical definiteness" and what Wallace calls "the problem of indefiniteness in quantum mechanics: the problem that the theory, without changes to its formalism, seems to predict that macroscopic quantities have no definite values" (Wallace, 2003, p.87).)

According to Wallace, macro-objects cannot be defined exactly –Wallace prefers to consider them as "structures" or "patterns" that emerge from the low-level quantum physics– and therefore it would be misguided to apply many-worlds talk to the objects themselves: what "branches" is the wave function, not the objects of our experience (which are structures within the wave function). Wallace therefore concludes that:

> "It makes sense to consider a superposition *of* patterns, but it is just meaningless to speak of a given pattern as being *in* a superposition." (2003, p. 98)

Applying this to a worn-out example, it means that a linear superposition of two wave functions corresponding to "dead cat" and "live cat" respectively, is another quantum state –but not of a cat. (See the discussion below for further clarification of this point.)

However, this is the point where one should take into account the results of decoherence. Decoherence removes interference between different components of the wave function. Wallace claims that structures can be identified in each of the components, and thus one can say that there is an object present in each of the branches (with corresponding properties, such as being dead or alive). Thus, "multiplicity occurs at the level of structure"(2003, p.17). This account is supposed to fill the interpretational gap between having a decohering wave function and obtaining two simultaneously existing copies of an object with contradictory properties. Essential in this account is thus that one has an ontology of *quantum states*, not of objects. So, the branches of the wave function do not represent contradictory properties of a single object, but rather different objects with contradictory properties in each branch[39].

Returning to the "inexactness" problem introduced decoherence; how is this accounted for by the above identification of structures in the quantum state with the objects of ou experience? First of all, since Wallace considers the choice for a preferred basis as a mere pragmatic requirement, this choice is only *locally* relevant and the fact that decoherence merely displaces the interference terms into the environment is not considered to be a problem. Second, Wallace acknowledges that

---

[39] See Butterfield (2001) for a more detailed argument and a defence of this view. Butterfield also notes that "we must in any case accept state-dependent (and so dynamics-dependent) definitions of physical systems" for a different reason; namely, a truly fundamental theory should not be based on an ontology of particles, but of fields. And in quantum field theory, the number of particles need not even be fixed, but is dependent on what the state is.

decoherence takes some time, but again, at the level of the macroscopic structural features of the quantum state this is irrelevant: "During the decoherence period the wave function is best regarded as some sort of 'quantum soup' which does not lend itself to a classical description – but since the decoherence timescale τ_D is incredibly short compared to any timescale relevant at the cat level of description, this need not worry us." Because of his structuralist, pragmatic view on macro-objects, Wallace can claim that a precise answer to the question *when* the state represents an object with particular properties (say, a dead cat), is irrelevant. The fact that no precise criterion can be given as to what makes a particular object exist, should not undermine its status as being real "any more than the absence of a precise point where a valley stops and a mountain begins should undermine the status of the mountain as real" (2003, p. 95).

Another aspect of inexactness (see also Butterfield (2001)) is that the theory does not specify a basis: quantities on macroscopic objects are coarse-grained, e.g. the position of the centre-of-mass of a pointer is non-maximal since it does not specify the positions of the individual particles. On Wallace's account this, again, is no problem since the branches are defined anthropocentrically and emerge at the level of macroscopic objects. A complete specification of a macroscopic system in terms of quantum mechanics is possible, but one will have to accept that there are no preferred bases, no definite properties and no persistence in time at this level:

> "At the lowest level, we have a description in terms of *Everett worlds*, i.e. by means of a fine-grained basis. It is at this level that we are giving a *complete* description of the state; nonetheless these worlds are instantaneous entities with no traceable histories, and there is a high degree of arbitrariness about the choice of basis. The arbitrariness is not complete however, since to be practically useful this basis must be a fine-graining of some *decoherence basis*, which will in turn be given by some choice of consistent history space. At this level of the description we can talk usefully of histories, describing a branching [...] set of worlds [macrorealms]. This is the level at which we obtain a useful classical limit." (Wallace, 2002, p. 651)

How does Wallace explain the persistence in time of macroscopic objects, then? As the previous quote suggests, he invokes the notion of consistent histories. The consistent histories formalism has connections with both the theory of decoherence and the Everett interpretation. Briefly, the idea is that one may define a "history" in the quantum formalism as a time-ordered sequence of projectors. A "consistent history space" is a collection of histories, such that at each instant of time, the projectors constitute a basis and the instantaneous events they define satisfy the classical probability calculus. Now there are many possible choices of consistent history spaces, but, says Wallace:

> "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. [...] [T]he point is (in part) 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. So for describing events in our vicinity, at least, there is an overwhelmingly preferred choice. As

> with the pragmatically preferred reference frames of relativistic spacetimes, the preference is only approximate and really only extends to our spatial vicinity: if we wish to pick a truly global, fine-grained basis then there will be considerable arbitrariness" (Wallace, 2002, p. 649)

Again, decoherence provides a criterion to define a pragmatic concept of continuance (by picking out a particular class of consistent history spaces). Wallace argues that as long as wave-packets do not significantly overlap, histories can be defined and it is unproblematic to regard individual worlds as persisting in time. Decoherence is supposed to secure that (on a coarse-grained level) macroscopic objects at different times can be identified. In more traditional MWI's, the notion of transtemporal identity is problematic because, among other things, the universal wave function branches each time an object is in a superposition state somewhere in the universe. Defining the preferred basis anthropocentrically circumvents this problem, and therefore also makes it more sensible to speak of worlds persisting in time. For if the branching basis is defined only *locally*, and on a coarse-grained level, it is reasonable to speak of histories, of transtemporal identity of objects, although "it will fail in some circumstances [...] or when looked at too closely" (Wallace, 2002).

In the next subsection I will examine whether Wallace's three central claims are tenable. First I will give my personal reconstruction of his argument and criticise what I consider to be the weak spots. After that, I will relate his claims to the discussion of the previous chapter and analyse to what extent they provide an answer to the questions that I posed at the start of this chapter,

4.4.2 The MWI made plausible?

The first problem Wallace sets out to tackle is the preferred basis problem (the general version). Essentially, his claim is that one should not pin ones ontology down on a particular branching, but consider the universal wave function as the fundamental entity. Wallace adopts an analogy with general relativity to argue that we should accept the idea of an universal wave function in the same way as we have (more or less) accepted the idea of a four-dimensional spacetime. The kind of reality it depicts is highly abstract, but it connects to our experiences when we adopt "a certain perspective" within the formalism. In GR, a choice of foliation represents such a perspective , whereas in QM it is a choice of basis.

I think that this certainly is an interesting way to think of quantum mechanics, but, regrettably, it does not seem to me to be more than that. If any choice is allowed, and we are asked to simply pick the basis that yields the most comprehensible picture, then this argument seems to add little to what we already have. Quantum mechanics has always been perfectly meaningful when one adopts a particular "perspective"; this is essentially what specifying an observable amounts to. In fact, both the consistent-histories programme[40] and modal interpreters such as Bub (1997) have quite successfully attempted to develop this idea formally as well as philosophically. One does not need decoherence for that. But what decoherence adds to the argument is that it defines a preferred basis unambiguously (as op-

---

[40] This approach is discussed by for instance Landsman (2007, section 7.2) and Schlosshauer (2004, sect. 4.G) in some detail.

posed to an arbitrary decomposition of the state vector in terms of relative states) and that this basis will correspond, more or less, with the "effectively classical" appearance of ordinary objects (both in a kinematical and in a dynamical sense, cf. section 2.3 in chapter 2). Thus, to recover our familiar reality, one should take the decoherence basis. I can agree with that. But I would say that with this prescription one *does* commit oneself to an explicit interpretational rule. (After all, the question of interpretation *is* about how to recover familiar reality from the formalism.)

Further, the claim itself that a particulary useful, effectively classical, basis exists does not tempt me to believe that all branches of the wave function defined that way represent actual "realities". In Wallace's his argument, however, these claims seem to go hand in hand (both are supposed to follow from the mechanism of decoherence). Nonetheless, Wallace's focus on the ontology of macro-objects seems to be an attempt toward a justification of the "multiple realities" idea. I take the core of Wallace's argument to be that, simply, one should not mess with one's ontology. Recall that he says that a superposition of "patterns" is not the same as "patterns" being "in" a superposition. According to Wallace, the latter phrase is "meaningless". Now, I do not believe that the statement is meaningless in itself, but perhaps Wallace means that it refers to some non-existent entity. Let me elaborate a bit:

The things that superpose are wave functions. If we indicate those with square brackets and put its physical referent (i.e. the observable consequences the theory attaches to the wave function) between the brackets, and if we omit the terminology of "patterns", the two statements Wallace is referring to take the following form:

1: "There is a superposition of [object A with property p] and [object A with property q]".

2: "Object A is in a superposition of [property p] and [property q]".

Now it is not immediately obvious that either of the two propositions its meaningless, but it is clear that they cannot correspond to one and the same ontology. In the first, the wave function is ontologically primary, and manifests itself as the object together with its observable properties –Wallace's idea of a "pattern". In the second, the object is not represented in the wave function but is ontologically prior to it; the wave function is the bearer of the object's properties. The latter statement is not meaningless by itself (although we do not know what physical meaning to attach to a superposition of well-defined physical properties), but it *is* meaningless with respect to Wallace's particular structural view on higher-order ontology. Because according to such a view, macro-objects are not ontological primitives, i.e. they are not the kinds of things wave functions are "assigned to". (As Wallace puts it: "cats themselves are not the sorts of things which can be in superpositions" (2003, p.98).) Note that Wallace claims that the second statement *is* meaningful for elementary particles since we can (if we disregard relativistic complications) regard them as ontologically primitive, in order to comply with standard usage of the wave function in ordinary, microscopic quantum physics, where there is a one-to-one correspondence between the objects and the subspaces associated with them. ("Any superposition of electron states is another electron state" (p. 98).)

According to Wallace, it makes no sense to think of cats in this way: "a state of a cat is actually a member of a Hilbert space containing states representing all possible macroscopic objects made out of the cat's atomic constituents" (p.98), and so the Hilbert space to which a particular cat state belongs, includes all kinds of things which are in fact not cats (or other well-defined objects) at all. Thus, one should keep a strict distinction between the "system" as an actual physical configuration of elementary particles on the one hand, and an "object" (what Wallace sees as the structures, or "patterns" emerging from the wave function), on the other. Macro *systems* are described by a state vector in a Hilbert space which is a product of the Hilbert spaces of all the micro systems that constitute it: $\mathcal{H}^n = \bigotimes_{i=1,\ldots n} \mathcal{H}_i$. A macro *object* does not correspond to such a Hilbert space. In fact, the number n of micro systems that constitute an object need not even be exactly defined. But there are certain (temporally extended) configurations of particles in certain states that one would associate with a certain object, i.e. particular states in $\mathcal{H}^n$ (ignoring possible fluctuations in the number of particles/elementary systems of which the object is composed). Clearly, a superposition of such states is another state of the same system, but usually not of the same object.

So far so good. Actually, I do think that Wallace has a point here; but at best it points us to a problem, not a solution. For what are we to make of the formalism if we cannot associate wave functions to the objects we are talking about in a straightforward way? Decoherence plays a crucial role in this respect; without decoherence, his "structuralist" analysis merely says that at the end of the measurement process there will be an awkward superposition that we do not understand –but that is not different from what the measurement problem is usually taken to mean. The idea that the problem is in the object which is indefinite rather than its properties does not seem to make much of a difference. But Wallace jumps from the preceding argument to the conclusion that decoherence gives rise to "the sort of branching structure which allows the existence of effectively non-interacting near-copies of a given process" (p.102) –a jump that is in my opinion way too quick. For what does it help us to replace "superposition" in the statement (1) above, with "decoherent superposition"? Wallace claims that this decoherent structure allows us to regard superpositions as "functionally distinct copies" of one and the same object, and that "this multiplication of patterns happens naturally within the existing formalism, and does not need to be added explicitly to the formalism" (Wallace, 2003, p.99). But to me this just seems a *non sequitur*. Of course, one might just *postulate* that decoherence leads to alternatives (perhaps in different realities), but that does not seem to be where Wallace is at –and in that case there would be no real merit of Wallace's approach over the more straightforward direct application of a MWI metaphysics to the results of decoherence anyway.

Perhaps one could even say that Wallace, by taking recourse to a structuralist view on macro objects, is undermining his own argument: his application of decoherence in not as straightforward as usual, in any case, given the fact that decoherence at least seems to presuppose a basic ontology of objects (cf. section 3.3.4). The central problem that the standard decoherence literature aims to address is how these objects acquire definite properties, not how the objects themselves emerge from an abstract universal wave vector.

Summing up, it is unclear to me what Wallace's structural view actually contributes to the overall argument. About the Schrödinger Cat example, Wallace says that "[to] predict what happens if we have a superposition of decaying and not decaying [...] we are taking advantage of the patterns present in the two branches of the wave function. [...] in each of the branches there is a 'cat' pattern, whose salience as a real thing is secured by its crucial explanatory and predictive role. Therefore, by Dennett's criterion [that says that "the existence of a pattern as a real thing depends on the usefulness of theories which admit that pattern into their ontology" (p. 92)] there is a cat present in both branches after measurement" (2003, p. 96). So, the argument for multiplicity hinges on the idea that the concept of a "cat", as a macro *object*, is not reducible to a description in terms of the micro-ontology (such a description would do no justice to the "cat character" of the of this macro *system*). But then it is unclear what his argument implies for, say, the Stern-Gerlach experiment, for it seems that one can give a perfectly exhaustive description of the relevant parts of the latter in terms of the micro physics.

Although Wallace's point that the usual representation of macro-objects in quantum mechanics is oversimplified, and the distinction between properties and objects as empirical referents of the wave function that he draws may well be relevant to the measurement problem, I do not think that it, as it stands, has much to offer for the questions raised at the start of this chapter, as I will now argue.

Evaluation of the questions

The three questions of which I claimed that they should in one way or another be addressed by a decoherence-based interpretation were Preferred Basis (how the interpretation basis is selected), Outcomes (how to explain our unique experiences) and Subsystems (how to separate the system of interest from the environment).

In order to appreciate Wallace's answer to the first of those three questions, it is important to distinguish two aspects of the problem; on the one hand how the preferred basis is defined, on the other why this (dynamically) preferred basis is the interpretation basis. With respect to the second point, Wallace adopts the pragmatic stance that I have already discussed. I find this acceptable, but in contrast to Wallace, I do consider the choice for the decoherence basis as the interpretation basis to be an explicit interpretational rule. (A reasonable one, but still.)

Wallace defines the decoherence basis as follows:

> "Decoherence specifies [...] a collection of projectors, and [...] any exact basis which [...] is a fine-graining of this collection, will consist of states which are stable against the decoherence process. By this, we mean that [...] the off-block-diagonal matrix elements of the operator will decay on a vanishingly short timescale compared to the timescale on which the block-diagonal matrix elements evolve" (2002, p.648).

Crucial here, and an important difference with Zurek, is that Wallace considers the *diagonalising* effect of decoherence as the defining aspect of the decoherence basis. Effectively, this requires the additional assumption that the environment states will quickly approach orthogonality. Thus, Wallace's selection criterion is equivalent to B, under an additional assumption about the $\mathcal{AE}$-dynamics which is satisfied for the known models (but perhaps not necessarily in reality). Unlike Zurek, Wallace does not emphasize the role of correlations; it therefore seems that he need not rely on the existence of a tri-decomposition of the universal state[41].

---

[41] So compare Wallace's views with (Zurek, 1993, p.309): "The key question in discussing the

However, like Zurek, Wallace introduces observers into the picture as "information processing systems". (See the quotation on p. 131.) Although he does not say clearly what he means by this statement, I assume that he means "information processing" and the reference to "storage of memories" in the same sense as Zurek; namely, as establishing stable correlations. If this is what Wallace is indirectly referring to, the criticism of section 4.2.2 applies here equally well: we have little reason to believe that realistic dynamics will result in tri-decomposable states, and therefore this simple model of perception and the epistemic relevance attached to it do not seem physically plausible.

However, in my opinion there is something else problematic about Wallace's account. Consider what he says in (Wallace, 2003, p.99): "we can and do remain neutral about the how this [quantum] state is itself to be interpreted, since all we need from it are its structural properties, such as: what its representation is in the eigenbasis of a given operator." Thus, in fact, Wallace admits that he needs to specify an operator to arrive at these "structural properties". This brings to mind the discussion in the previous chapter: the mixed states do not represent an "apparent mixture of pointer states" unless one chooses to represent it in that particular basis (which has to be estblished independently). Apparently, Wallace takes diagonality of the reduced density matrix as defining the emergent structural properties that are supposed to correspond to the macro ontology. But it seems that he has trouble making up his mind since he acknowledges that for finite times there will be interference terms present; this indicates that he has another selection criterion in mind, independent of diagonality.

Apparently, this leads Wallace to think that it necessary to defend his Everettian stance against the charge of "inexactness", about which he says that "[the objection is that] the decoherence process is only approximate: [...] the interference between terms, though very small, is not zero" (*Ibid.*, p. 90). This refers to the problem of approximate decoherence as was discussed in section 3.3.3: the modified e-e link (that assigns definite values to the diagonal terms of a (reduced) density matrix) does not assign any definite "properties" to the almost-but-not-completely diagonal density matrix $\rho_{\mathcal{SA}}(t)$ for finite t. However, one may doubt whether there really is an objection of inexactness that Wallace has to counter. If decoherence is invoked to solve the preferred-basis problem in the MWI, all that counts is that the relevant selection criterion yields a *uniquely* defined preferred basis. And this is the case for all the selection criteria A-C. (In contrast, diagonality alone is not sufficient to fix a decomposition for the instantaneous state.) In fact, the selection criteria do not even refer to the orthogonality of the environment states. On the other hand, as I said, for Wallace the disappearance of the interference terms due to the latter is a crucial element of his argument that multiplicity emerges "from within the formalism". (Whereas the traditional Everettian approach simply assumes such multiplicity.) So, although there is perhaps not a *general* "inexactness problem" for the MWI in the above sense, authors like Wallace, who ground their interpretation in the *diagonality* (in a certain fixed basis) of the density matrix, will in particular have to explain how the non-ideal nature of decoherence interaction fits into his overall conceptual scheme.

---

emergence of classical observables is not the set of states which are on the diagonal after everything else (including the memory [...]) is traced out, but, rather, the set of states which can be faithfully recorded by the memory."

One of the reasons for Wallace's particular approach to higher-order ontology is precisely to argue that there actually is no problem of inexactness in the Everettian approach; he calls such a demand for exactness a "fallacy" because macro-objects, the kinds of things for which the interpretational problems of quantum mechanics become genuine, are simply not exactly definable. So he concludes that

> "We can tolerate some small amount of imprecision in the macroworld: a slightly noisy pattern is still the same pattern. Hence we do not need to worry that decoherence does not give *totally* non-interfering branches, just very nearly non-interfering ones" (2003, p. 102)

Now, I agree that, as a physical system, macro-objects are "noisy" (Wallace's tiger may loose its hair and still be the same tiger). But I do not see why one kind of inexactness should be quite like the other; the presence of interference terms in the (local) quantum state does *prima facie* have little to do with Wallace's structuralist views on macro-ontology. Without some further explanation of what the connection between the two is supposed to be, Wallace's argument seems to be little more than a (rather superficial) rhetorical move. Alas, the problem is not solved by declaring both macro-objects and the results of decoherence "imprecise" and to note that they agree. If one does not have a quantum state picture of macro objects, then the empirical meaning of the interference terms is even more obscure than in orthodox quantum mechanics. In fact, one cannot even claim that the state with the tiny interference terms does in fact represent the object of interest.

Although, I believe, it does not really help with the problem of approximate decoherence, Wallace at least acknowledges that the "subsystems" that play such a prominent role in the decoherence argument, are usually not exactly defined. Does this have consequences for his assessment of the third problem, viz. the problematic status of these subsystems and the associated concept of "environment" in the decoherence argument? Despite his extensive discussion of the structuralist features of macro-ontology, Wallace does not directly address this issue. Concerning the definition of subsystems, I would say that, again, this is for Wallace merely a pragmatic issue. The factorization of the global Hilbert space is just as arbitrary as the choice of a preferred basis on a particular factorization. Only, given the dynamics of the universal state (the total interaction Hamiltonian) it may turn out that there are particular factorizations that are dynamically preferred in the sense that they give rise to effectively dynamically independent subspaces (i.e. the subspaces that are intrinsically contained in the form of the interaction Hamiltonian, cf. section 3.3.4), and on these subspaces we can then define the preferred states. This kind of view would fit in nicely with the approach proposed in section 3.3.4.

However, to me it is not so clear whether Wallace associates systems with subspaces, or with patterns within states defined on these subspaces. If the latter is the case, it seems that Wallace has a problem with the notion of the "environment". Namely, the patterns emerge within the wave function from the decoherence dynamics –Wallace associates the "two cat states" with the diagonal terms of the reduced density matrix. But this understanding requires that the irrelevant degrees of freedom are traced out: without the tracing procedure, there *are* simply no "patterns" to be found in the universal wave function. In other words; in order to obtain patterns in the wave function corresponding to well-defined classical systems with well-defined properties, one has to know what to count as "relevant" and what not –but intuitively, this rests on a pre-established concept of such well-defined "systems". With his emphasis on the "patterns", Wallace's account does therefore not seem to shed much light on the standing problem. It would help if Wallace would define the subsystems –the boundaries between what is and what is not to be ignored– with reference to the Hilbert space structure. Unfortunately, he does not devote much discussion to this important issue. (He claims that the consistent histories formalism he is using "abstracts this requirement [i.e. vanishing of interference] and frees it from direct reference to subsystems and the environment" (Wallace, 2002, p.648)–but I would say that this reference either is still there but has only become less explicit, or has been replaced by a not very insightful consistency condition.)

What remains to be addressed is the problem of Outcomes. Wallace's answer to this question is obviously Everettian. A crucial difference with the standard Everettian approach, however, is that Wallace wants us to believe that, thanks to decoherence, the replacement of indefiniteness by multiplicity "happens naturally within the existing formalism", and that this "essentially solves the problem of indefiniteness" (Wallace, 2003, p.99). Now, I doubt that this "essentially solves the problem", but it depends on what one takes this multiplicity to mean: at a glance multiplicity seems no less worse than indefiniteness. What really solves the problem (if you want), is the way this multiplicity is interpreted; strangely enough, Wallace has not much to say about this issue, but I take that to mean that it is here where he embraces the real implications of his Everettian stance: the multiple copies exist relative to multiple observers that are (again, thanks to decoherence) non-interacting, and supposed to exist in different "realities". Although it has little to do with decoherence or a "pattern" view on macro-ontology, it is precisely *this* (implicit) move that solves the problem of Outcomes.

4.5 Concluding remarks

In the previous chapters I have repeatedly remarked that a decoherence approach can only address the problem of Outcomes by supplementing the theory with some interpretational framework, and that the usual strategy is to adopt an Everettian stance in this respect. The two authors I have criticised in this chapter, Zurek and Wallace, do so in a way that may be less straightforward than anticipated. As I see it, these authors are reluctant to introduce any explicit additional postulates; especially Wallace seems to think that the MWI can be presented as a natural consequence of a realist interpretation of quantum mechanics (combined with a structural view on higher-order ontology). Now I am not convinced by their arguments, so a good way to start would be to see where an Everettian account would lead us if we *were* to introduce it at the interpretational level.

With respect to environment-induced decoherence, basically two different Everettian approaches can be distinguished. These correspond to the two essential consequences of environment-induced decoherence (section 3.2.1), the selection of a preferred basis and the (approximate) diagonality of the reduced density matrix in this basis. I paraphrase these as follows:

1. *Decoherence-relative states:* Analogous to the measurement process, one introduces an observer $\mathcal{O}$ that correlates his brain state to the state of the

observed object $\mathcal{A}$. The $\mathcal{OA}$-couple also interacts with an environment, and the result is a tri-decomposed state, which defines an unique branching (thanks to the tri-decompositional uniqueness theorem). These branches are interpreted in an Everettian sense as representing multiple, simultaneous but non-interacting observer's experiences.

2. *Diagonality-branching:* The different "outcomes" represented by the states on the diagonal of the decohered reduced density matrix are interpreted in an MWI-like fashion; they all occur simultaneously, but in different "realities".

I have argued that Zurek's argument, when stripped from all references to "reality" and "predictability", eventually boils down to the first of these approaches. Next, I argued that the crucial assumption that the final $\mathcal{OAE}$-state generally be tridecomposable and in particular that this decomposition behaves neatly, is untenable. Thus, one cannot invoke a criterion of "stable correlations" between observed an observer and the tri-decompositional uniqueness theorem to arrive at a uniquely defined branching which would solve the basis ambiguity inherent in the Everett interpretation. So I conclude that this idea will not work. In my opinion, the idea of Quantum Darwinism does not add much to the argument.

The second approach seems to be close to Wallace's views (but also to those of Zeh, (Joos et al., 2003)). Ignoring his specific ideas about the subject for the moment, it seems that it could prove to be somewhat more fruitful than the first since it does not require any specific assumptions about the dynamics except that the environment states will quickly approach orthogonality (which is claimed to be an (almost) universal characteristic of environmental interactions). Nor does it require a division into three subsystems (although, of course, the $\mathcal{SA}$-split is still essential). Note that it does require a special role for observation, however, since the diagonality of the density matrix relies on the locality assumption (3.2.2), although this role differs from that of the first approach above (which does not rely on locality, by the way).

However, also the second approach faces serious difficulties. The most prominent is the problem of approximate decoherence. In section 3.3.3 I argued that the interference ought to disappear *completely* if diagonality is to have any particular interpretational consequences, since any kind of interpretation that can dispose of them when they are small can also dispose of them when they are large. The situation for an MWI-like interpretation is not much better, it seems to me. Furthermore, there is the question of how the preferred basis (in this case, the branching) is defined. It seems that for this specific approach, the branches are generated by the states on the diagonal of the density matrix. As I argued in section 3.2 of the previous chapter, decoherence models are such that the dynamically preferred states will eventually diagonalize the reduced density matrix, but this does not imply that the two criteria are equivalent, in particular not for finite t. *Defining* the branching basis in terms of diagonality avoids the problem of approximate decoherence, but runs into trouble in case of degeneracy, as was shown in section 3.3.3. Moreover, if one ties the emergence of a "classical" ontology (or at least the definite appearance of a more abstract underlying quantum ontology) to the diagonality of the reduced density matrix, there seems to be a serious problem with the justification of the assumption of Locality, as I discussed in connection to Wallace's arguments (section 4.4.2).

Some of these questions are in a way addressed by Wallace, who recognizes their relevance for a decoherence-based MWI. Wallace tries to cope with the problem of approximate decoherence and the fact that the moment of branching (occurrence of measurement outcomes) is ill-defined by means of his structuralist view on macro-ontology. I am not convinced by his arguments, however, and I have tried to make clear why. Nevertheless, my criticism of Wallace seems to be founded on a fundamental disagreement rather than on solid logical or mathematical arguments. That does not necessarily counts in favor of Wallace's ideas; I think that it merely indicates that Wallace's interpretation fails make contact with the technical results it draws on. It may be compelling rhetoric, but in the end his argument remains floating in mid air.

I would conclude that there is a sense in which environment-induced decoherence may solve the preferred basis problem for the Everettian approaches. But tri-decompositional uniqueness will not do the trick, and I believe that also Wallace's attempt to recover the branching, together with its inherent "multiplicity" unterpretation, as a naturally emerging structural property of the universal wave function faces serious difficulties. The sense in which environment-induced decoherence may be of help, is that if one wants a kind of precise *rule* to define a preferred basis, one could appeal to the interaction Hamiltonian that defines a preferred basis in terms of the selection criteria A or B. Of course, the problem with this approach still is that a line must be drawn between the system of interest and the environment. (Recall that this is irrespective of the issue of locality, i.e. whether one should *ignore* the environment. See section 3.3.4.) I have tried to sketch an account of "subsystems" from a "bare bones" ontology consisting of the universal wave vector and the interaction Hamiltonian in section 3.3.4, but so far this did not seem very promising. That means that, as things stand, there seems to be no way to get round the prominent role for the observer in environment-induced decoherence. This would have been a major disappointment for Everett himself, who initially intended his relative-state interpretation as an "observer-free" reformulation of orthodox quantum mechanics.

Considering the fact that environment-induced decoherence does not really help with any of the other problems encountered by the MWI either[42], I find this kind of interpretation rather unattractive. It certainly is not the definitive answer to the kind of questions that have plagued quantum mechanics for more than half a century. So when Zurek announces in the abstract of his latest review paper that "In conjunction with Everett's relative state account of the apparent collapse these advances [i.e. emergence of pointer states, envariance and Quantum Darwinism] illuminate [the] relation of quantum theory to the classical domain of our experience. They [...] justify our confidence in quantum mechanics as [the] ultimate theory that needs no modifications to account for the emergence of the classical" (Zurek, 2007, p.1), he merely seems to have fallen prey to the kind of enthusiasm and self-confidence that is typical for the decoherence literature. In my opinion, however, this confidence is misguided.

---

[42] Zurek has claimed that one can derive the Born rule from the decoherence-related concept of "envariance", but I chose not to discuss this topic. See (Zurek, 1998, 2003, 2007).

Conclusion

> *"Today there seem to be no phenomena which contradict quantum theory – perhaps with the sole exception that there are definite ("classical") phenomena at all!"* - E. Joos, introduction to Joos et al. (2003).

Evaluation: Environment-induced decoherence and the measurement problem

I believe that I have gathered enough material to address the main question I posed in the introduction:

> *What exactly does environment-induced decoherence contribute to a solution of the measurement problem?*

From the decoherence literature, one might easily get the impression that there is a simple answer to this question: environment-induced decoherence means that "interaction with the environment will typically single out a preferred set of states [that] remain untouched in spite of the environment, while their superpositions lose phase coherence and decohere" (Zurek, 2003, p.717). Thus, it is claimed, decoherence not only solves the Preferred Basis problem, it also causes arbitrary states to appear as effectively classical (i.e. incoherent) mixtures of these states. I hope to have made clear that this claim is oversimplified.

Preferred Basis and Interference

To handle this issue with some more care, recall the discussion at the end of chapter 1. There, I concluded that decoherence does not solve the measurement problem in the sense as it is usually understood (at least by the more realist-minded), namely as the Problem of Outcomes: where do facts (or measurement outcomes) come from, given the fact that the e-e link does not assign any definite property to a quantum state that is in a superposition of eigenstates (of the corresponding observable)? However, environment-induced decoherence aims to address a weaker formulation of the measurement problem, or rather two specific aspects thereof; namely, the Preferred Basis problem and the Problem of Interference. An solution to both, or even just one, of these is often regarded to be sufficient to explain at least the *appearance* of facts. This led me to formulate two somewhat more specific questions:

> *What exactly is the (conceptual) relevance of addressing the problems of Interference and the Preferred Basis?*
>
> *Does environment-induced decoherence indeed solve these problems?*

With the results of the other three chapters in mind, these questions need some reconsideration.

First of all, the relevance of the Preferred Basis and the Interference problem depend on the specific kind of interpretation of the wave function (ensemble or individual) one adopts. The decoherence literature is pretty unclear about this, but considering the emphasis on the emergent "classical" features of individual quantum systems due to decoherence, it seems safe to assume an individual interpretation approach. In that case, decoherence theorists hope to explain that systems appear to be "definite" with respect to the properties represented by the states on the diagonal of the decohered density matrix. I have repeatedly stated that decoherence does not imply definiteness, and I stick to that (recall section 1.2.1). But in the light of the Everett-interpretation discussed in chapter 4 this criticism takes a different form. What I have called the "diagonality-branching" approach purports to interpret the decohered density matrix in terms of well-defined measurement outcomes occurring simultaneously in different "realities". However, it is not so obvious to me that the Everett-kind of approach *requires* diagonality as a necessary condition in order to work. Rather, I would say that any unambiguous and empirically adequate selection criterion for a preferred basis would do. But decoherence provides such an answer (in terms of either one of the selection criteria A, B or C of section 3.2.2) regardless of the diagonality of the reduced density matrix. In fact, were one to insist on diagonality, then finite-time effects may cause additional troubles for the Everettian (sections 3.3.3 and 4.4.2). (With regards to Interference, the arrow of explanation in fact seems to point the other way; we obtain a (reduced) density matrix diagonal in the basis of preferred states, but what are we to make of it? The ignorance interpretation now being abandoned, it seems that an Everettian account is the most obvious move for decoherence theorists like Zurek and Zeh.)

Second, if the decoherence theorists would endorse an ensemble interpretation, then it is not clear to me what one achieves by solving *both* the Preferred Basis and the Interference problem, as I argued before (cf. sections 1.1.1 and 1.3). As I explained in section 1.1.1, there are two ways to understand why one would need to obtain a diagonal density matrix at the end of the measurement process: from a "quantum" point of view that identifies this state with an "apparent mixture" of *quantum* states, and as a indication for the effective *classicality* of the state in terms of a certain privileged observable (represented by its eigenstates on the diagonal). However, for the latter, merely defining a Preferred Basis or a preferred observable (by means of a dynamical selection criterion perhaps, see section 3.3.1), seems to be sufficient: the interference terms have become superfluous since their presence makes no difference to the probabilities assigned to the states on the diagonal. (The fact that the non-diagonal terms have now been deprived of empirical meaning may perhaps be the reason that one would want to get rid of them.) On the other hand, the identification of the decohered reduced density matrix with an "apparent mixture" of quantum states rests on their statistical equivalence when *all* observables are taken into account, i.e. this conflicts with the idea of a preferred observable. This kind of conflict is also at the heart of the "approximate decoherence" objection (section 3.3.3).

Considering these objections, I believe that the relevance of the problem of Interference is actually rather marginal. (That is, as long as decoherence is not embedded in some interpretational framework that assigns specific relevance to diagonality, such as the "diagonality-branching" Everettian approach.) That environment-induced decoherence might be able to select a preferred basis is in my opinion of greater importance, although the (in my view obvious) consequence that the concept of observables needs to be reconsidered along the lines discussed in sections 3.3.1 and 3.4 requires more thorough philosophical underpinning. (Again, it might be more fruitful to combine this achievement of environment-induced decoherence with some kind of non-orthodox interpretation, like the MWI or a modal interpretation.) In any case, I do not think that environment-induced decoherence itself provides a consistent interpretation of a quantum state as long as no additional postulates are introduced. In section 3.3.1 I have already explained this in some detail, but let me briefly recall:

- First of all, a choice of basis is not physical: the dynamics does not "force out" a certain decomposition of the (reduced) state on $\mathcal{H}_{\mathcal{S}} \otimes \mathcal{H}_{\mathcal{A}}$.
- Nevertheless, a (specific class of) interactions between two systems show a preference for certain states of these systems, in terms of the selection criteria A-C, as explained in section 3.2.1.
- This dynamical selection of a preferred basis does not *destroy* states that are not preferred in this dynamical sense. But one may postulate an explicit rule (for instance by reconsidering the meaning of observables) that frames the empirical content of arbitrary states in terms of these observables.
- Decoherence models show that the dynamically preferred states will generally be (close to) position eigenstates. Thus, the kind of postlate suggested above would qualify "position" as the *a priori* definite quantity of our experiences. This could be considered as a partial explanation of the emergence of classicality.
- However, a different kind of approach is possible that does not need to postulate a new interpretational rule to explain our experiences, but instead takes a more pragmatic stance to resolve an ambiguity in the measurement model. This ambiguity is what I have named the "decomposition" version of the Preferred Basis problem. Briefly put: if one insists on *correlations* between system and apparatus states, one can invoke the tridecompositional uniqueness theorem to fix a unique preferred *decomposition* of the post-measurement state.
- In Zurek's "existential interpretation" (section 4.2.1) an analogy with the measurement situation is invoked in order to apply the same kind of argument to explain our perception of "classically definite states" in general. (In other words; this argument is supposed to explain why the dynamically preferred states determine the empirical content of arbitrary states.)
- Apart from the conceptual objections, this approach will not work in general because tridecomposable states are rare and/or very unstable against small

> perturbations. That the approach seems to work for simple measurement models is because of specific assumptions about the dynamics which are actually quite unrealistic (see also my evaluation of Zurek's interpretation in section 4.2.2).

Turning to the second question; "does environment-induced decoherence solve the Preferred Basis problem?", I would therefore say "yes and no". "Yes", in the sense that taking the interaction Hamiltonians into account may provide a *context-related but state-independent* criterion for selecting an observable or basis, "no" in the sense that this preferred basis *by itself* does not yield the (qualitative) empirical content of the quantum state, and "no" in the sense that the specific assumptions that yield a unique tridecomposition in terms of the states selected by the dynamical selection criteria, will not be satisfied for realistic situations.

Second, in spite of its questionable relevance, one would perhaps like to know whether environment-induced decoherence solves the problem of Interference. This is a more technical issue that I have not examined in detail, but I am willing to follow the literature in the claim that all models of environmental interactions show rapid convergence of the reduced density matrix to an expression diagonal in a basis which can be classified as in a sense "classical" (e.g. coherent states), that fluctuations are small, recurrence of coherence ("recoherence") is rare, and that more complicated and exceptional dynamics will lead to other aspects of "effectively classical" behavior or otherwise unexplained quantum effects. (Like for instance chirality or the Zeno effect, cf. section 2.3.3.) But my objections against the claim that decoherence addresses Interference are conceptual, rather than technical, in nature. First of all, the dynamics by itself does not cause the reduced density matrix to "look diagonal" (i.e. like an "apparent mixture" of these preferred states). Decoherence merely says that the dynamically preferred basis will (approximately) diagonalize the reduced density matrix (cf. section 3.2). Furthermore, if one insists on *exact* diagonality, in such a pre-determined preferred basis, then environment-induced decoherence does not achieve this. I have discussed the kind of difficulties this gives rise to in section 3.3.3.

Towards an account of facts

So far I have focussed on the two questions that environment-induced decoherence explicitly claims to solve, viz. the Preferred Basis problem and the problem of Interference. Although an answer to these questions is in a sense relevant (although in my opinion not so essential as is usually claimed), it does not pertain to the question that is usually considered to be at the heart of the measurement problem, namely the problem of Outcomes: how to account for the occurrence (of perception) of unique and definite "events" (or "facts", or "measurement outcomes") within the quantum formalism. So I would also like to briefly consider the following question:

> *What does environment-induced decoherence have to offer for a solution to the problem of Outcomes?*

I have argued (section 1.2.1) that a solution to this problem is not implied by a solution to Interference. I have also argued (section 3.3.1) that the idea that the selection of a preferred basis at least leads to the *appearance* of outcomes (even though the theory cannot account for the occurrence of a particular one) is seriously misguided. Furthermore, an ignorance interpretation is well-known to be untenable (section 1.2.2). The only sense in which environment-induced decoherence, unsupplemented, may support an interpretation in terms of "facts" is by adopting an ensemble interpretation, which takes the occurrence of these facts more or less for granted (section 3.3.2).

Nevertheless, all this does not yet imply that environment-induced decoherence would not be relevant in this respect when embedded in a non-orthodox interpretation. In chapter 4 I have investigated how environment-induced decoherence supports two interpretations of the Everettian kind: Zurek's "existential interpretation" and Wallace's pragmatic-structuralist interpretation of the universal wave vector. I believe that Zurek's proposal essentially rests on the tridecompositional uniqueness theorem, which, I argued, is not likely to apply in this case (section 4.2.2). Wallace's account I find more difficult to grasp, but as I have understood him he proposes that the diagonality of the reduced density matrix essentially governs the branching. This means that either he *defines* diagonality as the mechanism behind our experiences, or that he determines the preferred basis by different means and accepts the claim of the decoherence theorists that in this basis the state will rapidly converge towards diagonality. (But as discussed in section 3.3.3, the first approach faces the problem of degeneracy, and the second the problem of imperfect decoherence.)

Disregarding the fact that I am not convinced by Wallace's arguments (4.4.2), the idea itself that one should interpret the "events" on the diagonal in an MWI-like fashion seems to fare slightly better than Zurek's proposal. But as I said in section 3.4, environment-induced decoherence may address the preferred basis problem for the MWI in a more straightforward way by appealing to the interactions directly. Personally, I do not believe, however, that the MWI is the right way to address the interpretational difficulties of quantum mechanics. Considering the other, rather serious, problems this approach is facing (section 4.3.2) for which decoherence does not seem to be of much help, I am hesitant to accept the decoherence-MWI interpretation as final word on the subject. Especially because all of the selection criteria (A, B and C in section 3.2.1) depend on a "split" into subsystems that seems hard to justify from an MWI point of view – even if the criterion does not depend on a locality constraint, such as criterion A. (I have made an attempt to justify this split in section 3.3.4, but I do not consider that to be successful.)

Getting rid of the observer

The problem of defining the subsystems raises another issue that I would like to spend a few more words on, namely the role of the observer. Let me put this in terms of the following questions:

> *Does the decoherence approach require a primitive notion of "observation" in order to give an empirical interpretation of the quantum formalism?*
>
> *If so; is this a problem?*

Returning to the discussion of the measurement problem in chapter 1 and the Introduction, we see that the first question reflects an often repeated objection against the orthodox and Copenhagen interpretation of quantum mechanics. In the Copenhagen interpretation, it is the physicist who determines what propositions about a physical system should be considered "meaningful", and in the orthodox interpretation[43] the act of observation even induces a physical change of the quantum state. In both cases, one cannot conceive of quantum mechanics as depicting an observer-independent reality.

For a realist, this is somewhat disturbing, but the fact that the observed measurement outcome is in part created by the act of observation itself, need not be a complete drawback. If one could understand this unavoidable perturbation by the measurement as acting upon something *real*, possessing well-defined properties before *and* after the measurement, I see no reason to reject this intervention of the observer from a realist point of view. However, the problem of course is that orthodox quantum mechanics needs the observer in order to establish something actual; before the collapse of the wave function, the standard interpretation does not assign any meaning to a superposition of eigenstates[44].

So far the metaphysics; but in section 1.1.1 I formulated a second objection against the orthodox interpretation that ought to worry physicists and philosophers alike. The point is that the collapse postulate is supposed to be a physical effect that, nevertheless, cannot be accounted for by the theory and that is supposed to take place at a vaguely defined moment caused by a vaguely defined, and manifestly *un*physical act of "observation", or "measurement". It is this *ad hoc* manoeuvre that most interpretations of quantum mechanics, quite rightfully, want to abolish. In section 3.1, I argued that in this respect, environment-induced decoherence succeeds. Although it requires a prominent role for (non-)observation, this is merely so because classicality is assumed to be merely apparent.

I have two things to add to this. First of all, when formulating the argument that environment-induced decoherence survives the objections against the collapse postulate, I assumed that the theory of decoherence can explain the classical appearances all by itself. But I have just concluded that this is not the case, and that the theory needs to be embedded in, for instance, an Everettian framework. In a sense, this does not matter; modern versions of the MWI do not regard classicality as an absolute independently existing aspect of reality, but as an artefact of our information-processing machinery. But in another sense, it seems to me that the MWI cannot dispel the ghost of the collapse completely –but this depends on the way one incorporates decoherence into the story. In the line of Zurek's argument, one could include the observer into the wave function, and then the appearance of definite facts is due to the observer becoming correlated to the superposition state. This requires no active branching, but is simply part of the interpretation of the formalism. But if one starts with the diagonal density matrix, and aims to interpret the states on the diagonal as existing simultaneously in different "realities", or branches of the universal wave function, then it is not so clear when the

---

[43] See section 1.1.1 for the difference between the two.

[44] One could perhaps say that the wave function nevertheless refers to a real entity, but that the latter does not possess any "well-defined property" when it is not in an eigenstate of a corresponding observable. But I think that making sense of an entity without well-defined properties either requires obscure metaphysics (which the adherents of the orthodox interpretation would most likely want to avoid) or is not that much different from having no real entity at all.

"facts" appear. Wallace argues that it is incorrect to demand a precisely defined moment of branching, but this argument rests on his assumption that multiplicity "happens naturally within the existing formalism", i.e. that the branching is a natural consequence of the decoherence dynamics. But I am not so convinced of this.

Second, environment-induced decoherence is of course not immune to the first objection formulated above. Although it does not require an active collapse, the theory *does* require the observer to make sense of the formalism in the first place. Simply put, classical definiteness emerges as the observer decides to focus his attention on a certain subsystem (or equivalently, to ignore the environment). But the point is not that prior to this act of (non-)observation, reality is "not classical" in any sense; the point is that the theory is unable to make *any* statements about what kind of reality this is. This is not merely a concern for realists; as I pointed out in section 3.3.4, I think it is simply obscure to say that one needs to "observe" a subsystem in order to arrive at the kind of comprehensible, well-defined classical picture in which those subsystems play an ontological role (or any role whatsoever) of their own.

The feasibility of "no-interpretation" interpretations

Thus, as I see it, at the heart of the problem is that a decoherence-based interpretation does not want to be an interpretation, but would like to be a natural consequence of the quantum formalism as it is generally understood. (Indeed, both Zurek and Wallace take a similar stance with respect to their Everettian views.) As I argued in the conclusion of chapter 1 (section 1.3), this kind of approach seems to be self-defeating. Because such a "no-interpretation" interpretation is essentially an orthodox interpretation. The problem is not just that it refuses to make statements about the kind of reality behind the phenomena (cf. my criticism above). On top of that, there seems to be a consistency problem; the decoherence theorist *cannot avoid* making statements about the empirical content of the theory during the process by which he aims to arrive at this empirical content. And that empirical content is generally framed in orthodox terms. To be somewhat more concrete: the orthodox interpretation interprets the quantum formalism in terms of probabilities for measurement outcomes. The goal of the decoherence argument, ideally, is to arrive at a formal expression that indeed represents such a set of well-defined measurement outcomes, in the light of the e-e link. The idea is that the quantum state can only be rightfully interpreted in orthodox terms as a set of measurement outcomes if it is diagonal in the corresponding basis – and environment-induced decoherence aims to show that diagonality will in fact be the *general* state of affairs, in order to secure consistency. However, this claim is only approximately true (both the generality and the diagonality). I argued in section 3.3.3 that in order to justify the empirical validity of this approximate result, the decoherence theorist will eventually have to take recourse to the orthodox interpretation, *prior* to having established the result that is supposed to validate the latter.

But the problem goes beyond that of approximate decoherence. It is especially relevant in order to appreciate the meaning and relevance of the experimental confirmation of decoherence effects. As I already noted in section 1.3, this empirical confirmation is to be expected if the decoherence theorists have done a good job.

After all, decoherence is not supposed to show a violation of the predictions of quantum theory, and the latter has already been confirmed to be correct to an extraordinary extent. (Contrast this situation with, for instance, the possibility of an experiment to decide between standard quantum mechanics or GRW theory.) But what do these experiments show? They yield certain statistics that are in line with the predictions of the decoherence model –that is, predictions extracted from the model by applying the orthodox statistical algorithm to it, as is common practice in quantum mechanics[45]. (Indeed, we do not have any other method at our disposal.) So in particular, one is *supposing* that the model should be interpreted in terms of measurement outcomes for a certain specified observable (or several observables) that are indeed obtained in the experiment. But one cannot (of course not!) observe how these measurement outcomes "come into being". And wasn't that what the decoherence argument was supposed to show? From this I conclude that decoherence may be good physics –it was not my purpose here to criticize that part of the story. But this physics cannot support the philosophy of the decoherence theorists.

What environment-induced decoherence has to offer

All this makes it rather difficult to give a clear-cut answer to my main question. What does decoherence contribute to a solution of the measurement problem? I would say that it all depends (of course) on what one takes to be the problem. I have just spent a couple of pages in trying to disentangle these various perspectives. So let me just close by stating what I think the problem is that environment-induced decoherence might be able to help with.

I think that for environment-induced decoherence to be of use for the interpretation of quantum mechanics, one would have to tone down his ambitions. This means giving up the pursuit of a no-nonsense realist interpretation of the formalism, in particular the pursuit of a solution to the problem of Outcomes. I say that this program –the idea that one could recover familiar reality from the quantum state concept alone– must be given up because of the difficulties I see with the "no-interpretation" idea underlying decoherence, and the role the decoherence formalism requires for the observer. In particular, I think that one should give up altogether the attempt to recover –without adding any extra metaphysics– "classical" reality from a formalism of which the ontological meaning is not even remotely understood. This is partly because of the philosophical problems I foresee with such an approach, and partly because I believe that there is no reasonable sense in which decoherence can be claimed to lead to definiteness.

In fact, for me it is tempting to take an anti-realist view of quantum mechanics. That does not mean that I see instrumentalism as the goal of science – but it means that I find it hard to conceive of quantum mechanics as anything else than a very useful tool for predicting (probabilities for) empirical data. There have been various attempts at a realist re-formulation of the theory and I believe these kind of attempts are worthwhile. But such a realist interpretation of the empirical data need not be cast in quantum-mechanical terms. Accordingly, I would suggest to consider the relevance of decoherence from the point of view of an ensemble

---

[45] For a particularly illuminating discussion of the often-cited decoherence experiment of Brune *et al.*, see Placek (1998).

interpretation[46].

I think that anti-realism comes in degrees, however. One of the most unpleasant aspects of orthodox quantum mechanics is that it is unable to save the phenomena – unless one specifies what phenomena are to be saved (even though only in a probabilistic sense). Experiments yield empirical data belonging to a specified class (a physical quantity); so for instance, to take a simple example, the orientation of a Stern-Gerlach experiment determines what spin component is measured. The particular physical setup is a physical condition *required* to obtain data for a particular quantity. The formalism, however, is indifferent in this respect; the state itself, possibly combined with a detailed description of the measurement process, does not supply these empirical data unless one specifies what the relevant observable is. Since, moreover, there is no clear-cut unified picture behind these data, it seems that there exists a gap between the phenomena predicted and the formalism that predicts them.

It is in the latter sense that I think environment-induced decoherence may make a modest contribution. But this contribution is remote from the initial intentions of the decoherence theorist. Since a substantial part of my criticism is directed at the "no-interpretation" or "physics only" approaches to the measurement problem, I would conclude that what quantum mechanics needs is a clear and empirically adequate interpretation of the formalism, not a physically more sophisticated description of the measurement process. From that point of view, as I see it, the essential tenet of decoherence is the following: *include the interaction Hamiltonian as an extra ingredient in the interpretation.* This provides the desired link between the physical context of an empirical question and the (probabilistic) answer provided by the formalism. Decoherence may do so in various ways, but the easiest and probably most successful way is to define a preferred basis in terms of what I have called the "no-entanglement" selection criterion (A).

I prefer (A) as a selection criterion in the first place because it does not require an observer in the sense of a locality condition (although it does require a, perhaps artificial, split into subsystems). The "purity" criterion that follows from this locality condition may seem to allow a stronger physical justification, but I think this is mistaken when it comes to defining the interpretational axioms. Pragmatically, tridecompositional uniqueness may seem to do the trick, but as I have explained this is unlikely to yield empirically adequate results. So what we get is that the empirical content of the quantum state is fixed directly by reference to the Hamiltonian, without more ado.

Despite this rather optimistic closing remarks, I think that a lot of work remains to be done if we are to understand the implications of the decoherence program for the foundations of quantum mechanics. In the Outlook I will briefly sketch some of these open problems. All in all, my work has consisted mainly in raising questions. Questions about the interpretation of the results of the decoherence program, but

---

[46] This is not to say that an ensemble interpretation is the only viable option. But I just mean to contrast this with decoherence in the light of the individual interpretation. An alternative would be to consider a modal interpretation of the quantum state, in which probabilities are associated with propensities - with events which may or may not be the case - of a single quantum system, instead of distributions within an ensemble. Because of its rejection of the e-e link, the modal approach does not fall prey to my criticism of the individual interpretations. The modal approach is also anti-realist in the sense that it leaves the actualization of possible events unexplained. (Although for instance Lombardi and Castagnino (2008) hold a different view.)

also questions about the meaning and relevance of the kind of foundational difficulties this theory aims to address. Clearly, that task is easier fulfilled than providing the answers to the philosophical problems that have surrounded quantum physics since the days of its conception. In any case, I do not have those answers. But I do hope to have made clear that there are no compelling reasons either to think that the answers can or should be found in decoherence theory.

Summary

I close with a pointwise summary of my main conclusions:

1. Aspects of the measurement problem (Chapter 1):
    (a) **Outcomes:** to account for the occurrence of *unique* events (facts) in quantum mechanics.
    (b) **Definiteness:** to obtain a set of well-defined alternatives (for Outcomes).
    (c) **Preferred Basis:** to specify a physical quantity that determines the empirical content of the quantum state.
2. Aspects of the Preferred Basis Problem (Chapter 1):
    (a) **General** (associated with the problem of emergence): to give an unambiguous and objective criterion that fixes the interpretation basis of a quantum state $|\Psi\rangle \in \mathcal{H}_{\mathcal{A}}$.
    (b) **Decomposition** (associated with the ideal measurement scheme): to fix a unique bi-decomposition of a state $|\Psi\rangle \in \mathcal{H}_{\mathcal{S}} \otimes \mathcal{H}_{\mathcal{A}}$.
3. Interpretations of the quantum state (Chapter 1):
    (a) **Ensemble:** Quantum state describes an ensemble of identically prepared physical systems (which are not to be described quantum mechanically).
    (b) **Individual:** Quantum state describes an individual quantum system.

    The questions of Outcomes and/or Definiteness only pertain to the Individual interpretations. In the light of an Ensemble interpretation, only the Preferred Basis problem is relevant.
4. Environment-induced decoherence: can only account for a Preferred Basis with the help of an additional new postulate. This is *prior* to the question of how to account for Outcomes or Definiteness. Two approaches possible (section 3.2):
    (a) **Correlations:** Require a *tri-decomposition* of the quantum state (selection criterion C). This requirement is naturally motivated from the context of a perfect measurement that includes environmental interaction. The tri-decompositional uniqueness theorem implies that this decomposition is uniquely defined, *if* it exists.
    (b) **Observable:** Postulate that the dynamically preferred states (in terms of selection criterion A, section 3.2.2) define the interpretation basis of the (reduced) quantum state.

    The first approach may be extended to situations more general than the (ideal) measurement, by motivating it in terms of stable correlations between an observer and an observed system. ("Existential interpretation", section 4.2) However, the general existence and usefulness of the tri-decomposition basis is questionable. The second approach is less explenatory powerful, but likely to yield a well-defined and empirically adequate interpretation basis. Because of its axiomatic status and therefore lack of proper motivation the second approach does not seem to go well with a purely realist interpretation, however.

5. Obtaining Definiteness: requires, but does not follow directly from, Preferred Basis. Requires a further additional postulate that the terms in the quantum state defined by the Preferred Basis represent alternatives (facts/ events/ measurement outcomes). (Compare the "emergent multiplicity" argument of Wallace, section 4.4.)
6. Obtaining Outcomes: requires, but does not follow directly from, Definiteness. Requires interpretation of alternatives in terms of unique experiences. For instance MWI-like view (section 4.5).
7. Subjective elements: the Locality condition (ignoring the environment) is irrelevant for a decoherence-based interpretation. (Decoherence only yields the Preferred Basis.) Any decoherence-based approach *does* require a "split" into subsystems, however (section 3.3.4), but this split may perhaps also be defined dynamically, in terms of the interaction Hamiltonian.

Outlook

Although my aim was to provide a discussion of the relevance of environment-induced decoherence for the measurement problem that goes somewhat beyond the usual literature, it is of course subject to limitations. So I wish to close by mentioning a number of the topics that I have not been able to discuss but that nevertheless seem to be particularly relevant.

I have purposely abstained from mathematical detail. Nonetheless, I believe that a clear mathematical proof makes a more convincing argument than a thousand words. Although largely conceptual in nature, part of the arguments that I presented could become stronger (or perhaps shown to be mistaken!) if supported by a solid mathematical argument. This concerns in particular the following issues:

- In section 3.2 I used an insightful, but rather limited simplified presentation of the decoherence argument to support my conclusions. It would be interesting to consider the selection criteria (i.e. the no-entanglement and purity conditions) from a more general mathematical point of view, i.e. for more general Hamiltonians and for different factorizations of the overall Hilbert space.

- Similarly, it would be interesting to investigate what the exact conditions on the dynamics and initial states are that yield a final state that can be tri-decomposed.

- I claimed that the limiting orthogonality of the environment states is a dynamical effect that is independent of the dynamical conditions that lead to specific preferred states. Decoherence models in the literature generally show this kind of orthogonality, which in turn is associated with the concept of dynamically induced superselection rules. The mathematical conditions leading to the latter are rather well-known[47] but I do not know whether there are any results that point at a deeper connection between these conditions for limiting orthogonality and the existence, nature and dynamical behavior of a preferred basis in terms of the selection criterion (A).

Furthermore, there are a couple of points of mathematical concern that I have not really discussed, but that ought to be considered in more detail:

---

[47] See for instance the contribution of Kupsch in Joos et al. (2003), who states that the crucial condition for limiting orthogonality is that the model has a semibounded Hamiltonian, and whether and what superselection sectors are induced is completely determined by the system-environment Hamiltonian.

- Environment-induced decoherence is generally associated with a kind of "dynamical superselection rule", but this identification rests on a rather sloppy use of the mathematical terminology. More work would have to be done to examine to what extent the results of decoherence can be fitted into the mathematical formalism (cf. section 2.2.3).

- Related to the latter issue is the question of a connection between environment-induced decoherence and the spin-measurement model of Hepp (1972). The latter is usually not considered as a decoherence model since it does not invoke an external environment, but relies on an algebraic model in terms of quasilocal observables instead. In the light of the discussion in section 3.3.4, such an approach could nevertheless perhaps be considered as a kind of coarse-graining approach, i.e. a model of decoherence by an internal environment. It would be interesting to investigate to what extent the general decoherence models and Hepp's approach overlap, since the latter exemplifies what a thorough mathematical derivation of approximate superselection rules within a (more or less realistic) model of measurement could look like[48].

- For reasons of illustrative transparency and mathematical convenience, I have framed the discussion in a setting of finite-dimensional Hilbert spaces. Specifically, I have throughout assumed that the observables have a discrete spectrum. The fact that position has no exact eigenstates, for instance, has been neglected in my discussion (and many other discussions!) of the measurement problem. I suspect that taking this into account may cause additional difficulties; the first one I foresee is that my insistence on an exact eigenstate-eigenvalue link (and consequently, perhaps the points of criticism derived from it) may be quite unrealistic.

Besides these mathematical refinements, a point that needs more consideration, both mathematically and conceptually, is the idea of defining the observables dynamically (sect. 3.3.1 and above). It would be interesting to see whether a consistent re-formulation of the quantum formalism starting from the states and the Hamiltonians as its basic ingredients is possible, and whether the resulting formalism turns out to be operationally equivalent to the orthodox approach in all respects. (If so, one could consider this dynamical background as a kind of explanation of why the orthodox approach is successful – although also such a dynamical reconstruction still remains mute as to the ontological meaning of the formalism.) Although conceptually different, the modal interpretation of Lombardi and Castagnino (2008) could be considered as an indication of what such a program might look like.

A further prominent interpretational issue that I hardly touched is the status of the Born rule in environment-induced decoherence. Zurek has made some efforts to derive the Born rule from the concept of "envariance" that is based on symmetry considerations of states entangled with the environment (Zurek, 1998, 2003, 2007). However, it is not clear to me what exactly the relevance of this attempt at derivation is. I have criticised the application of the Born rule as part of my objection

---

[48] Kiefer notes that "superselection sectors induced by the environment differ qualitatively from the reduction of the algebra of observables as considered by Hepp" (Joos et al., 2003, p.341). This remark pertains specifically to the decoherence time scale, which is typically larger for the Hepp model.

against the “no interpretation-interpretations” idea. But that criticism focuses on the underlying orthodox assumption that the empirical content of the wave function consists of definite measurement outcomes. Would decoherence be able to establish the latter, then the numerical validity of the Born rule follows naturally from Gleason’s theorem (that is, if one associates these measurement outcomes with one-dimensional projectors, i.e. adopts the e-e link.) However, it seems to me that Zurek does not question the assumption of Outcomes (or at least definiteness) in his derivation of the Born rule. For instance, Zurek (2007) appears to identify the (decohered) components of the $\mathcal{SA}$-state with “outcomes” or “facts” straight away (and proceeds to derive the numerical equivalence of his “envariantly” defined probabilities with Born’s rule from a principle of equal likelihood). If my understanding is correct, I do not see what his argument is supposed to deliver. In any case, I have the impression that Zurek’s derivation suffers from other difficulties as well that would be interesting to examine in more detail.

At the end of chapter 2, I have discussed the claim that environment-induced decoherence would explain the emergence of classicality (in the sense of section 1.2.3). I concluded that this claim seems to be rather far-fetched, but since I decided to focus on the measurement problem, I concede that perhaps I did not give the subject enough attention to warrant this kind of scepticism. In any case, I have not been able to find any compelling evidence in the literature that the decoherence program has really obtained crucial results in this direction, but perhaps I did not look hard enough. The question is interesting in its own right, also when the measurement problem is not at stake, so my criticism of the decoherence in the light of the latter should not be taken as being applicable in all respects to the general question of the quantum-to-classical transition.

Another issue that can be considered separately from the measurement problem, and that I have therefore largely ignored, is the connection between environment-induced decoherence and the emergence of irreversibility in thermodynamical processes. Environment-induced decoherence is considered to be a practically irreversible process, since the correlations between the system and the (much larger) environment can practically not be undone once they have been established. And so Zeh, for instance, claims that “Decoherence by “continuous measurement” [by the environment] seems to form the most fundamental irreversible process in Nature. It applies even where thermodynamical concepts do *not* (such as for individual molecules [...]), or when any exchange of heat is entirely negligible” (Joos et al., 2003, p. 16). On the other hand, it seems that decoherence does not yield an answer at all to the central question of how time-asymmetric processes can emerge from the time-symmetric fundamental law. Indeed, the argument that entanglement with the environment can hardly be undone, and that this is the source of irreversibility, presumes a particular initial state for the apparatus; to wit, one that is pure. But if one is to take decoherence seriously, in particular the claim that environmental interactions are abundant, then one should conclude that the state of the apparatus will practically always be mixed. (Recall that the decoherence argument does not show that the environmental interactions will cause the apparatus to be in one of the einselected states; at most it shows (e.g. by embedding it in an Everett-interpretation) that the apparatus *appears* to be in such a state).

I have criticized the claims of experimental testability for “no-interpretation” interpretations on grounds of the orthodox presuppositions that underlie the empirical interpretation of these experiments. This is a rather general point, but otherwise I have largely neglected the experimental aspects of the decoherence program. A more detailed investigation of this purported empirical evidence would therefore be appropriate (although I do not expect that it would invalidate my main point.)

Finally, throughout I have confined this discussion to the non-relativistic domain. This was for obvious reasons; the measurement problem is in the first place a problem of non-relativistic quantum mechanics, and a relativistic treatment will only cause further complications and is unlikely to be very illuminating. Nevertheless, if an answer to this problem can be found, it should allow a natural relativistic extension. This does not seem to be the case for environment-induced decoherence. I am not sure whether this is particularly problematic, but one may argue that since the standard formulation of the theory crucially depends on a locality principle, a relativistic formulation would in fact be indispensable to apply that idea rigorously.

These are just a small sample of issues that I did not discuss, mainly for reasons of limited time and scope. I suppose that more gaps in my argument remain. But I also believe that the same is true for the theory of environment-induced decoherence. As I see it, to uncover these flaws would have to be the primary task of philosophers interested in the decoherence program. Not only is a critical evaluation essential to properly judge the theory on its foundational merits, it will eventually also lead to further progress.

Bibliography

Amman, A. (1991). Chirality, a superselection rule generated by the molecular environment? *Journal of Mathematical Chemistry*, 6:1–15.

Anglin, J., Paz, J., and Zurek, W. (1996). Deconstructing decoherence. *arXiv: quant-ph/9611045v1.*

Bacciagaluppi, G. (2000). Delocalized properties in the modal interpretation of a continuous model of decoherence. *Foundations of Physics*, 30(9):1431–1443.

Bacciagaluppi, G. (2005). The role of decoherence in quantum mechanics. *The Stanford Encyclopedia of Philosophy: http://plato.stanford.edu/archives/sum2005 /entries/qm-decoherence/.*

Bacciagaluppi, G. and Hemmo, M. (1994). Making sense of approximate decoherence. In *PSA: proceedings of the Biennal meeting of the Philosophy of Science Association*, volume 1, pages 345–354.

Bacciagaluppi, G. and Hemmo, M. (1996). Modal interpretations, decoherence, and measurements. *Studies in History and Philosophy of Modern Physics*, 27:239 – 277.

Barrett, J. (2003). Everett's relative-state formulation of quantum mechanics. In *The Stanford Encyclopedia of Philosophy: http://plato.stanford.edu/archives /spr2003/entries/qm-everett/.*

Bassi, A. and Ghirardi, G. (2000). A general argument against the universal validity of the superposition principle. *Physics letters A*, 275:373–381.

Batterman, R. (2007). Intertheory relations in physics. In *The Stanford Encyclopedia of Philosophy: http://plato.stanford.edu/archives/fall2007/entries/physics-interrelate/.*

Bell, J. S. (1990). Against 'measurement'. *Physics World*, aug:33–40.

Blume-Kohout, R. and Zurek, W. (2005). Quantum darwinism: Entanglement, branches, and the emergent classicality of redundantly stored quantum information. *arXiv: quant-ph/0505031v1.*

Brown, H. (1986). The insolubility proof of the quantum measurement problem. *Foundations of Physics*, 16(9):857–870.

Bub, J. (1997). *Interpreting the quantum world.* Cambridge University Press.

Butterfield, J. (2001). Some worlds of quantum theory. *arXiv:quant-ph/0105052v1.*

Butterfield, J. and Earman, J., editors (2007). *Handbook of the Philosophy of Physics*, volume 2A of *Handbook of the Philosophy of Science*. North-Holland.

Camilleri, K. (2008). The history of decoherence: the dynamics of entanglement and interpretation. lecture given at the HQ2 Second International Conference on the History of Quantum Physics, Utrecht 2008.

Dalvit, D., Dziarmaga, J., and Zurek, W. H. (2005). Predictability sieve, pointer states, and the classicality of quantum trajectories. *Physical Review A*, 72(062101):1–10.

Daneri, A., Loinger, A., and Prosperi, G. (1962). Quantum theory of measurement and ergodicity conditions. *Nuclear physics*, 33:297–319.

deWitt, B. (1973a). The many-universes interpretation of quantum mechanics. In DeWitt and Graham (1973).

deWitt, B. (1973b). Quantum mechanics and reality. In DeWitt and Graham (1973).

DeWitt, B. S. and Graham, N., editors (1973). *The many-worlds interpretation of quantum mechanics*. Princeton University Press.

Dickson, M. (2007). Non-relativistic quantum theory. In Butterfield and Earman (2007).

Donald, M. J. (2004). Instability, isolation, and the tridecompositional uniqueness theorem. *ArXiv: quant-ph/0412013v1.*

Elby, A. (1994). The 'decoherence' approach to the measurement problem in quantum mechanics. In *PSA: proceedings of the Biennal meeting of the Philosophy of Science Association*, pages 355–365.

Elby, A. and Bub, J. (1994). Triorthogonal uniqueness theorem and its relevance to the interpretation of quantum mechanics. *Physical Review A*, 29:4213–4216.

Everett, H. (1973). The theory of the universal wave function. In DeWitt and Graham (1973).

Faye, J. (2002). Copenhagen interpretation of quantum mechanics. In *The Stanford Encyclopedia of Philosophy: http://plato.stanford.edu/archives/sum2002/entries /qm-copenhagen/.*

Fine, A. (1996). *The Shaky Game: Einstein, Realism, and the Quantum Theory.* The University of Chicago Press, second edition.

Goldstein, S. (2006). Bohmian mechanics. In *The Stanford Encyclopedia of Philosophy: http://plato.stanford.edu/archives/sum2006/entries/qm-bohm/.*

Hackermüller, L., Hornberger, K., Brezger, B., Zeilinger, A., and Arndt, M. (2004). Decoherence of matter waves by thermal emission of radiation. *Nature*, 427:711–714.

Hepp, K. (1972). Quantum theory of measurement and macroscopic observabes. *Helvetica Physica Acta*, 45:237–248.

Joos, E. (2005). Dynamical consequences of strong entanglement. In *Séminaire Poincaré*, number 2, pages 1–12. Available online at arXiv: quant-ph/0512079v1.

Joos, E. and Zeh, H. (1985). The emergence of classical properties through interaction with the environment. *Zeitschrift für Physik B*, 59:223–243.

Joos, E., Zeh, H., Kiefer, C., Giulini, D., Kupsch, J., and Stamatescu, I.-O. (2003). *Decoherence and the appearance of a classical world in quantum theory.* Springer-verlag Berlin, second edition.

Kent, A. (1990). Against many-worlds interpretations. *International Journal of Modern Physics A*, 5:1745–1762. Online (1997) version (with foreword) on arXiv: gr-qc/9703089v1.

Landsman, N. P. (1995). Observation and superselection in quantum mechanics. *Studies in History and Philosophy of Modern Physics*, 26:45–73.

Landsman, N. P. (2007). Between classical and quantum. In Butterfield and Earman (2007), pages 417–533. Also available online at *arXiv: quant-ph/0506082 v2.*

Leggett, A. J. (2002). Testing the limits of quantum mechanics: motivation, state of play, prospects. *Journal of Physics: Condensed Matter*, 14:415–451.

Lombardi, O. and Castagnino, M. (2008). A modal-hamiltonian interpretation of quantum mechanics. *Studies in History and Philosophy of Modern Physics*, 39:380–443.

Maudlin, T. (1995). Three measurement problems. *Topoi*, 14:7–15.

Mermin, N. D. (1993). Hidden variables and the two theorems of John Bell. *Reviews of Modern Physics*, 65:803–815.

Ochs, W. (1981). Some comments on the concept of state in quantum mechanics. *Erkenntnis*, 16:339–356.

Olliver, H., Poulin, D., and Zurek, W. H. (2004). Environment as a witness: selective proliferation of information and emergence of objectivity. *arXiv: quant-ph/0408125v1.*

Omnès, R. (1992). Consistent interpretations of quantum mechanics. *Reviews of Modern Physics*, 64:339–382.

Omnès, R. (2003). Decoherence, irreversibility and the selection by decoherence of quantum states with definite probabilities. *arXiv: quant-ph/0304100v1.*

Pessoa, Jr., O. (1998). Can the decoherence approach help to solve the measurement problem? *Synthese*, 113:323–364.

Placek, T. (1998). Observing a Schrödinger cat. *Foundations of Physics Letters*, 11(3):209–231.

Redhead, M. (1987). *Incompleteness, Nonlocality and realism.* Oxford University Press.

Ruetsche, L. (1995). Measurement error and the Albert-Loewer problem. *Foundations of Physics Letters*, 8(4):327–344.

Schlosshauer, M. (2004). Decoherence, the measurement problem, and interpretations of quantum mechanics. *arXiv: quant-ph/0312059v2.*

Stamp, P. (2006). The decoherence puzzle. *Studies in History and Philosophy of Modern Physics*, 37:467–497.

van Fraassen, B. (1991). *Quantum Mechanics: An Empiricist View.* Clarendon Press, Oxford.

von Neumann, J. (1955). *Mathematical foundations of quantum mechanics.* Princeton University Press. English translation from the 1935 German edition.

Wallace, D. (2002). Worlds in the Everett interpretation. *Studies in History and Philosophy of Modern Physics*, 33:637–661.

Wallace, D. (2003). Everett and structure. *Studies in History and Philosophy of Modern Physics*, 34:87–105.

Wallace, D. (2007). The quantum measurement problem: State of play. *PhilSci archive: 00003420.* Draft version, forthcoming in the Ashgate Companion to the New Philosophy of Physics.

Zeh, H. (1971). On the irreversibility of time and observation in quantum theory. In d'Espagnat, B., editor, *Foundations of Quantum Mechanics.* Academic Press New York.

Zeh, H. (1973). On the interpretation of measurement in quantum theory. *Foundations of Physics*, 1:69–76.

Zurek, W. H. (1981). Pointer basis of quantum apparatus: Into what mixture does the wave packet collapse? *Physical Review D*, 24:1516–1525.

Zurek, W. H. (1982). Environment-induced superselection rules. *Physical Review D*, 26:1862–1880.

Zurek, W. H. (1993). Preferred states, predictability, classicality and the environment-induced decoherence. *Progress of Theoretical Physics*, 89:281–312.

Zurek, W. H. (1998). Decoherence, einselection, and the existential interpretation (the rough guide). *Philosophical Transactions of the Royal Society of London*, 356:1793–1821. Page numbers refer to online version at arXiv: quant-ph/9805065v1.

Zurek, W. H. (2002). Decoherence and the transition from quantum to classical – revisited. *Los Alamos Science*, 27:86–109. Modified version of paper originally published in Physics Today (1991). Available online at http://www.fas.org/sgp/othergov/doe/lanl/pubs/number27.htm.

Zurek, W. H. (2003). Decoherence, einselection, and the quantum origins of the classical. *Reviews of Modern Physics*, 75:715–775.

Zurek, W. H. (2007). Relative states and the environment: Einselection, envariance, quantum darwinism, and the existential interpretation. *arXiv: quant-ph/0707.2832v1.*