Life and death in the tails of the GRW wave function

David Wallace*
Balliol College, University of Oxford
(Dated: July 17, 2014)

It seems to be widely assumed that the only effect of the Ghirardi-Rimini-Weber ('GRW') dynamical collapse mechanism on the 'tails' of the wavefunction (that is, the components of superpositions on which the collapse is *not* centred) is to reduce their weight. In consequence it seems to be generally accepted that the tails behave exactly as do the various branches in the Everett interpretation except for their much lower weight.

These assumptions are demonstrably inaccurate: the collapse mechanism has substantial and detectable effects within the tails. The relevance of this misconception for the dynamical-collapse theories is debatable, though.

INTRODUCTION: THE PROBLEM OF TAILS

The GRW dynamical-collapse theory, and its more sophisticated descendants,[20] set out to solve the measurement problem in perhaps the most direct way possible: by modifying the normal unitary dynamics so as to replace the ill-defined 'projection-postulate' with a genuine dynamical process which with very high probability collapses macroscopic superpositions onto macroscopically definite states.

The 'problem of tails' [1] arises because (it is claimed) the GRW collapse mechanism fails to produce states which actually are macroscopically definite.

In more detail: recall that the fundamental assumption of the GRW model[21] is that each particle (say, the ith particle) has a very small random chance per unit time of collapsing via the process

$$\psi(x_1, \ldots, x_N) \longrightarrow \mathrm{e}^{-(x_i-x_0)^2/2a^2}\psi(x_1, \ldots, x_N) \qquad (1)$$

(I omit normalisation). The 'collapse centre' x_0 is determined randomly, with its probability of being in a region around some x being equal to the 'standard' probability of a position measurement finding the particle at x.

Since a Gaussian vanishes nowhere, obviously the collapse mechanism cannot localise the wavefunction in any region of configuration space that it was not already localised in. If a 'macroscopically definite' state is supposed to be localised in the region of configuration space corresponding to our classical notion of the location of that state, then we have a problem.[22]

The problem actually comes in two flavours (here I follow [16]). The first might be called the problem of 'bare tails', and can be stated for a system consisting of a single particle: a particle in a Gaussian state is not strictly located in any finite spatial region at all, and so (it is argued) cannot be regarded as describing a localised classical particle, no matter how narrow the Gaussian is. The problem of bare tails has received extensive discussion in the literature recently (see, for instance [3, 6, 7, 9, 11–13]).

I will not be concerned much with the bare-tail problem in this paper. I will make one comment, though: the bare-tail problem is not really anything to do with the GRW theory, but is a natural consequence of unitary Schrödinger dynamics. Wave-packets with compact support cannot be created (they require infinite potential wells); if they were to be created, they would spread out instantaneously.[23] As such, if bare tails are a problem then they are a problem for any version of quantum mechanics that takes the wave-function as representing macroscopic ontology (such as the Everett interpretation).[24]

The problem of 'structured tails', by contrast, is explicitly a problem restricted to dynamical collapse theories. Recall that in a Schrödinger-cat situation, with a state like

$$\frac{1}{\sqrt{2}}\left(|\text{alive cat}\rangle + |\text{dead cat}\rangle\right) \qquad (2)$$

ideally what we want dynamical collapse to do is to deliver (half the time, anyway) the state

$$|\text{alive cat}\rangle\,. \qquad (3)$$

But the collapse mechanism doesn't actually do this. Within 10^{-11} seconds or so one of the nucleons[25] in the cat will undergo collapse, with a 50% chance of concentrating almost all its amplitude onto being in the living cat. Since it is entangled with the remaining nucleons in the cat, this will yield a state something like

$$\alpha\,|\text{alive cat}\rangle + \beta\,|\text{dead cat}\rangle \qquad (4)$$

where $|\alpha|^2 \gg |\beta|^2$ but $\beta \neq 0$.

Arguably, this doesn't do us much good. The dead-cat part of the state may have very low weight but it's still just as much part of the state. And in the GRW theory, there is no *conceptual* connection between mod-squared amplitude and 'probability' or 'actuality' or anything: the connection is supposed to be purely dynamical, manifesting via the collapse process. So it seems that we still have macroscopic superpositions, and that dynamical collapse has not after all solved the measurement problem.

The problem can be sharpened by comparing dynamical-collapse theories to the Everett interpretation (here I follow [8]). Modern versions of the Everett interpretation do not introduce 'worlds' or 'minds' as extra terms in the formalism: rather, they make use of dynamical decoherence to show that the unitarily-evolving wave-function is a superposition of essentially-independent quasi-classical worlds. In my preferred form of the interpretation (see [19] or [17] for details) the 'worlds' are to be understood as structures or patterns in the underlying quantum state: decoherence, suppressing as it does the interference between quasiclassically definite states in a superposition, guarantees that multiple such patterns evolve almost independently. Applying this to a state like (2) tells us that we have a world with a live cat and another with a dead cat; applying it to a state like (4) tells us exactly the same.

Dynamical-collapse theories have the same ontology as the Everett interpretation; they differ only in dynamics. As such, it seems that (4) must be interpreted as a many-worlds state, with the dead-cat 'tail' being just as real as the much higher-weight live cat component.

How different are the dynamics? It seems to be generally assumed[26] that they are very similar indeed: the only effect of the collapse mechanism is to damp the amplitudes of all branches but one, but the branches themselves continue to evolve normally. Call this the assumption of *quasi-Everettian dynamics*, or QED. Under unitary dynamics (2) evolves into something like (schematically)

$$\frac{1}{\sqrt{2}}\,|\text{Newspapers report `cat lives!'}\rangle$$

$$+\frac{1}{\sqrt{2}}\,|\text{Newspapers report `cat dies!'}\rangle\,; \qquad (5)$$

if QED were true, then, (4) would evolve into

$$\alpha'\,|\text{Newspapers report `cat lives!'}\rangle$$

$$+\beta'\,|\text{Newspapers report `cat dies!'}\rangle \qquad (6)$$

where again $1 > |\alpha'|^2 \gg |\beta'|^2 > 0$.

If QED were true, it would in my view pose a very serious problem for dynamical collapse theories: a problem very similar, in fact, to the 'empty-wave' or 'Everett-in-denial' problem for the de Broglie-Bohm theory (see [5] for a presentation of this problem). The low weight of the 'tail' would be empirically undetectable by anyone in it, and the collapse mechanism would be epiphenomenal.

QED is not true, however: the collapse mechanism has dramatic dynamical consequences for the tail, as we shall see.

LIFE IN THE TAILS

There is a widespread misconception[27] about the effects of the GRW collapse mechanism, which goes as follows: if a state is initially a superposition of states localised around points x and y, then the post-collapse state is again such a superposition, just with one of the localisation peaks greatly magnified in comparison with the other.

This isn't true. If the collapse happens around x, say, then it actually has two effects: the amplitude of the peak around x is greatly increased relative to the peak around y, and the centre of the peak around y is displaced significantly towards x.

It is easy to show this directly: suppose that the initial wave-function is one dimensional and proportional to

$$\psi(x) = \mathrm{e}^{-x^2/2w^2} + \mathrm{e}^{-(x-x_0)^2/2w^2}; \qquad (7)$$

that is, suppose it is an equally weighted sum of two Gaussians. The effect of collapse (assuming that the collapse peak is at x=0) is to multiply ψ by $\exp(-x^2/2a^2)$; a little algebra gives the result

$$\psi(x) \longrightarrow \psi'(x) = \mathrm{e}^{-x^2/2w'^2} + \mathrm{e}^{-(x-x_0')^2/2w'^2}\mathrm{e}^{-x_0^2/a'^2} \qquad (8)$$

where $a'^2 = a^2 + w^2$, $(1/w'^2) = (1/a^2) + (1/w^2)$, and $x_0' = x_0 \times a^2/(a^2+w^2)$. The GRW parameter a is generally taken to be $\sim 10^{-7}m$; on the assumption that the peaks are much narrower than this (i. e. $w \ll a$) this simplifies approximately to

$$\psi'(x) \simeq \mathrm{e}^{-x^2/2w^2} + \mathrm{e}^{-(x-x_0[1-w^2/a^2])^2/2w^2}\mathrm{e}^{-x_0^2/a^2}. \qquad (9)$$

So, as well as being shrunk by a factor $\mathrm{e}^{-x_0^2/a^2}$, the 'tail' peak has also been displaced a fraction (w^2/a^2) of the distance towards the collapse centre.[28]

The quantitative form of the displacement is dependent on the particular (Gaussian) form of the peaks used in $\psi(x)$. The overall conclusion, however, is robust: the effect of collapse on the tail peak is to displace it towards the collapse centre. From the point of view of an observer in the tail of a Schrödinger-cat state like (4), the effect is that a particle is 'kicked' by the collapse. Furthermore, since the tail's amplitude is very small, any subsequent collapses will almost certainly not prefer (that is, be centred on) the tail, so all other particles in the tail whose counterparts in the main part of the wavefunction are spatially separated from them will also be kicked if they are subject to collapse.

The 'kick' has some interesting consequences for the stability of matter in the tails. Suppose a particle undergoing 'kicking' is part of some compound system with centre-of-mass coordinate vector R; the particle's own coordinate vector x can then be written as $x = R + r$. Let ψ be the wave-function of the compound, by default take expectation values with respect to ψ, and choose coordinates so that $\langle R\rangle = 0$. We can assume that ψ may be written as a product of a free-particle wavefunction for the centre of mass and a much more complicated wavefunction for the internal degrees of freedom; for simplicity assume that this internal wavefunction is rotationally invariant.

Now suppose that the collapse is centred on some point x_0 and that the collapse function is $c(x)$ (in the standard GRW model, $c(x) = \exp(x - x_0)^2/2a^2)$, so that the post-collapse wavefunction is

$$\psi'(x, y_1, \ldots y_n) = c(x - x_0)\psi(x, y_1, \ldots, y_n). \tag{10}$$

The expected value of the particle's position relative to the compound centre of mass is then

$$\langle r\rangle_{\psi'} = \frac{\int c(x)^2 r\psi^*\psi}{c(x)^2\psi^*\psi}. \tag{11}$$

If we take a linear approximation to $c^2(x)$: that is, approximate $c^2(x) \simeq c^2(x_0) + x \cdot \nabla c^2(x_0)$, then we get

$$\langle r\rangle_{\psi'} = \frac{c^2(x_0)\langle r\rangle + \nabla_i c^2(x_0)\langle(R^i + r^i)r\rangle}{c^2(x_0) + \nabla c^2(x_0) \cdot \langle(R + r)\rangle} \tag{12}$$

which simplifies, given our symmetry assumptions, to

$$\langle r\rangle_{\psi'} = \frac{2\nabla c(x_0)}{c(x_0)}\langle r^2\rangle \sim \frac{w^2\nabla c(x_0)}{c(x_0)} \tag{13}$$

where w is the characteristic width of the compound system.

That is: the 'kick' displaces a bound particle partly outside its parent compound. If the kick distance is comparable to the actual width w of the compound, this will at least significantly excite it, and perhaps even disrupt it. From the above, a rough-and-ready criterion for excitation is

$$\left|\frac{\nabla c(d)}{c(d)}\right| > \frac{1}{w} \tag{14}$$

where d is the distance from the collapse centre. Using the standard GRW collapse function, this yields

$$d > \frac{a^2}{w} \tag{15}$$

as the criterion for excitation. Since the normal value of a is taken to be $\sim 10^{-7}$ m, the kick will cause atomic excitation when $d \sim 10^{-4}$ m and nuclear excitation when $d \sim 1$ m.

To see the practical effects of this, suppose that we prepare a "Schrödinger cat" state: a macroscopic object of say 10^{27} atoms in a superposition of two locations a couple of metres apart. (This isn't terribly difficult: Schrödinger's original method will do fine.) Within 10^{-14} seconds (assuming a GRW collapse frequency of 10^{-16}s), the first collapse will occur and the amplitude of (say) the left-hand part of the superposition will be drastically reduced relative to the right hand side.

The macroscopic object contains $\sim 10^{28}$ nucleons and $\sim 10^{28}$ electrons, and $\sim 10^{12}$ of these will undergo dynamical collapse per second. This will have a completely negligible effect on the left-hand term in the superposition. The nucleons and electrons in the right-hand side, however, will be kicked towards the left-hand side. The kick to the electrons is not particularly important; the kick to the nucleons, on the other hand, will kick each nucleon clean outside the nucleus. Due to the short range of the nuclear force, this will cause that nucleon to be ejected from the nucleus entirely.

The structure of the nucleus is complicated and not that well understood quantitatively, but in qualitative terms this will lead to (at least)

1. Gamma radiation as the remnants of the nucleus settle down from their current highly excited state into the ground state appropriate to the new number of protons and neutrons in the nucleus.

2. Possible beta or alpha radiation or electron capture, since the remnant nucleus is probably unstable.

3. If the ejected particle is a neutron, then beta radiation as it decays.

For instance, a collapse hit on a neutron in the nucleus of a carbon atom kicks the tail component of that nucleus into a highly excited state of carbon-11, which (assuming it is not so excited as to break up altogether) rapidly emits gamma radiation as it relaxes into the ground state and then decays via electron capture to boron-11, emitting $\sim$ 2MeV; the neutron decays in $\sim$ 10 minutes into a proton-electron-neutrino pair and emits $\sim$ 1MeV in doing so.

To summarise: if objects in the tails of the wavefunction are displaced by about a metre from the location of their counterpart in the main part of the wavefunction, they become radioactive, with a mean lifetime equal to the GRW collapse rate — that is, $\sim 10^{16}$s, or about 100 million years. Estimating the decay rate for smaller separations is much more technically difficult, but if we crudely assume a Gaussian wave-packet for the nucleon, the decay rate for small separations is proportional to the square of the separation. Hence (for instance) a 1-centimetre separation between the component of the superposition in the tail and its counterpart leads to decay rates suppressed by a factor of 10^4 relative to the one-metre rate.

DEATH IN THE TAILS

The decay energies quoted in the previous section — $\sim$ 3MeV — are fairly typical of the energies produced by dynamical-collapse-triggered decay; as such, a kilogram of matter in the tail which is displaced by more than a metre or so from its high-weight counterpart will emit energy at a rate of (10^{-16} decays per second per nucleon $\times \sim 10^{27}$ nucleons per kg $\times \sim$ 1MeV per decay)$=\sim 10^{11}$ MeV per second, or about 10^{-8} watts. (This should be taken as a lower limit as it makes no allowance for gamma radiation when excited states de-excite.) A large fraction of this energy will be in the form of highly ionizing beta radiation caused by neutron decay.

This level of radiation will be harmful to living creatures in the tails. Precise calculations seem inappropriate given the very rough nature of the estimates used so far. Note, however, that if a living being (say, Schrödinger's unfortunate cat) were to be displaced by more than a metre or so from its high-weight counterpart, and were to absorb all of the ionizing radiation emitted by its own radioactive components (a reasonable order-of-magnitude approximation given the relatively high cross- section of beta radiation with matter), it would receive a radiation dose of $\sim$ 100 rem per year. This compares very unfavourably to the Environmental Protection Agency's recommended safe (human) dosage of 100 millirem per year; the EPA quotes 400 rem as the threshold fatal dose. And this is only the radiation exposure from the being's own body, and does not take account of likely radiation from surrounding matter.

What consequences does this observation have for the measurement problem? Recall that the dynamical collapse program hoped to establish

> **With overwhelmingly high probability, agents will observe quantum statistics very close to the averages predicted by quantum mechanics.**

However, this is threatened by the problem of structured tails. If QED had held, we would instead have had

> **With overwhelmingly high probability, the overwhelmingly high-weight branch will be one in which agents will observe quantum statistics very close to the averages predicted by quantum mechanics; however, the weight of a branch is not detectable by any agents, including those who are in very low-weight, anomalous branches.**

The failure of QED analysed in this section and the last leads to an intermediate result:

> **With overwhelmingly high probability, agents will either observe quantum statistics very close to the averages predicted by quantum mechanics, or in due course die of radiation sickness.**

Strictly speaking, I suppose that this intermediate result rescues dynamical-collapse theories from the problem of structured tails and ensures that they do, after all, solve the measurement problem. It explains why the scientific community has so far observed statistical results in accord with quantum mechanics (via the anthropic fact that worlds in which violations were observed are now radioactive deserts). And it explains why it is rational to act as if the predictions of quantum mechanics were true (because in those worlds where they turn out false, we're all doomed anyway).

However, from a purely sociological viewpoint I suspect that this will not be deemed adequate by the foundations community. If so, then the problem of structured tails is real even if its usual description (via QED) is false. The only recourse that I can see for dynamical-collapse theorists is then to modify the form of their collapse function. If the collapse function is taken to have *compact support* (differing from an exponential at distances of, say, $\sim 10a$ from the collapse centre) then the dynamical effect on the main part of the wavefunction will be completely negligible but the tails will be erased entirely.

Letting the collapse function have compact support is useless from the point of view of the problem of bare tails, since the wave-function will instantaneously evolve to have non-compact support. However, it is the *structure* of the tails, and not their mere existence, that causes the problem. Compact support would solve the problem very straightforwardly[29] and would spare our low-weight counterparts in the tails from their otherwise grim fate.

* Electronic address: david.wallace@philosophy.ox.ac.uk

[1] Albert, D. Z. and B. Loewer (1990). Wanted Dead or Alive: Two Attempts to Solve Schrödinger's Paradox. In A. Fine, M. Forbes, and L. Wessels (Eds.), *Proceedings of the 1990 Biennial Meeting of the Philosophy of Science Association*, Volume 1, pp. 277–285. East Lansing, Michigan: Philosophy of Science Association.

[2] Allori, V., S. Goldstein, R. Tumulka, and N. Zanghi (2008). On the common structure of Bohmian mechanics and the Ghirardi-Rimini-Weber theory. *British Journal for the Philosophy of Science 59*, 353–389.

[3] Bassi, A. and G. Ghirardi (1999). More about Dynamical Reduction and the Enumeration Principle. *British Journal for the Philosophy of Science 50*, 719–734.

[4] Bassi, A. and G. Ghirardi (2003). Dynamical reduction models. *Physics Reports 379*, 257.

[5] Brown, H. R. and D. Wallace (2005). Solving the measurement problem: de Broglie-Bohm loses out to Everett. *Foundations of Physics 35*, 517–540.

[6] Clifton, R. and B. Monton (1999). Losing your Marbles in Wavefunction Collapse Theories. *British Journal for the Philosophy of Science 50*, 697–717.

[7] Clifton, R. and B. Monton (2000). Counting Marbles

with 'Accessible' Mass Density: a reply to Bassiand Ghirardi. *British Journal for the Philosophy of Science 51*, 155–164.

[8] Cordero, A. (1999). Are GRW tails as bad as they say? *Philosophy of Science 66*, S59–S71.

[9] Ghirardi, G. and A. Bassi (1999). Do Dynamical Reduction Models Imply That Arithmetic Does Not Apply to OrdinaryMacroscopic Objects? *British Journal for the Philosophy of Science 50*, 705–720.

[10] Ghirardi, G., A. Rimini, and T. Weber (1986). Unified Dynamics for Micro and Macro Systems. *Physical Review D 34*, 470–491.

[11] Lewis, P. J. (1997). Quantum Mechanics, Orthogonality, and Counting. *British Journal for the Philosophy of Science 48*, 313–328.

[12] Lewis, P. J. (2003). Counting marbles: Reply to critics. *British Journal for the Philosophy of Science 54*, 165–170.

[13] Monton, B. (2004). The problem of ontology for spontaneous collapse theories. *Studies in the History and Philosophy of Modern Physics 35*, 407–421.

[14] Pearle, P. (1989). Combining Stochastic Dynamical State-Vector Reduction with Spontaneous Localization. *Physical Review A 39*(5), 2277–2289.

[15] Wallace, D. (2006). In defence of naiveté: The conceptual status of lagrangian quantum field theory. *Synthese 151*, 33–80.

[16] Wallace, D. (2008). The interpretation of quantum mechanics. In D. Rickles (Ed.), *The Ashgate Companion to Contemporary Philosophy of Physics*, pp. 197–261. Burlington, VT: Ashgate. Available online at http://arxiv.org/ under the title "The Quantum Measurement Problem: State of Play".

[17] Wallace, D. (2010). Decoherence and ontology: Or: How I learned to stop worrying and love FAPP. In S. Saunders, J. Barrett, A. Kent, and D. Wallace (Eds.), *Many Worlds? Everett, Quantum Theory, and Reality*, pp. 53–72. Oxford: Oxford University Press.

[18] Wallace, D. (2011). Decoherence and its role in the modern measurement problem. Forthcoming in *Philosophical Transactions of the Royal Society A*; available online at http://arxiv.org/abs/1111.2187.

[19] Wallace, D. (2012). *The Emergent Multiverse: Quantum Theory according to the Everett Interpretation.* Oxford: Oxford University Press.

[20] The GRW theory was originally proposed in [10], and was significantly revised by [14]. For a comprehensive review, see [4]. For philosophical discussion, see section 6 of [16] and references therein.

[21] The collapse mechanism proposed by GRW has since been superseded in technical work by more sophisticated variants (primarily so as to address the problem of identical particles) but to the best of my knowledge nothing conceptually fundamental depends on this change, so for convenience I will work with the basic GRW model.

[22] The problem can be avoided via the "primitive ontology" proposal in [2], which supplements the wavefunction in GRW with additional properties intended to represent the spatially localised entities; I will not be concerned with this strategy here. (In their terms, I am concerned only with GRW_0.)

[23] *Instantaneously*, not just at lightspeed: this isn't an artefact of non-relativistic physics. See [15] and references therein for more on this topic.

[24] I don't think they are a problem, in fact: rather, they demonstrate that the eigenvector-eigenvalue link is a hopeless way to understand the ontology of realistic quantum systems . . . but this is not the topic of the present paper. (See my discussion in [18].)

[25] I am assuming the modern 'mass-density' version of the dynamical-collapse theory, in which electron collapse is negligible compared to nucleon collapse; nothing at all hangs on this, though.

[26] See [8] for an explicit statement, but I have frequently heard it said in conversation and correspondence.

[27] See, e. g. , [1].

[28] This should not actually be surprising. The effect of the collapse on the tail peak is to multiply together two Gaussians with centres a distance x_0 apart: one would expect the resultant function to have a peak somewhere between the two original peaks.

[29] Whether it would make it easier or harder to generalise dynamical-collapse theories to relativistic quantum mechanics, I do not know; harder, at a guess, since compact-support functions are less straightforward to handle than exponentials.