Novel relationships between superoscillations, weak values, and modular variables

Jeff Tollaksen
Center for Quantum Studies, Department of Computational and Data Sciences, and Department of Physics, College of Science, George Mason University, Fairfax, VA 22030
E-mail: jtollaks@gmu.edu

Abstract.
We present several novel, unexpected relationships between superoscillations, weak values and modular variables. For example, we show how an uncertain phase, which characterizes the process of projecting a particle onto a superoscillatory region, can create the high-momentum associated with the super-oscillation. If an uncertain phase can localize the particle, then a definite phase can also localize it. This introduction of a relative phase corresponds to a non-local exchange of modular variables. We also present a new way to measure the nonlocality in the equation of motion for modular variables by using weak measurements.

1. Introduction
Superoscillations [8] are functions which oscillate with an arbitrarily high frequency \( \alpha \), even though they can be understood as superpositions of low frequencies, \( |k| < 1 \), seemingly a violation of the Fourier theorem:

\[
\sum_{|k|<1} c_k e^{ikx} \to e^{i\alpha x}
\]  

(1.1)

Popescu and Rohrlich [31] used such states to model a photon:

A large box is prepared, containing a photon in a state of high momentum. For a certain interval, a small window on one side of the box opens. The box may emit a photon. The energy of this photon will be the same as its energy in the box, up to a corrections due to truncation of the emitted wave train. However these corrections can be made arbitrarily small by sufficiently increasing the time the window is open. So far, there are no surprises. Now we repeat the experiment, but with the photon in a superposition of a number of low-momentum states. Suppose that the resulting wave function has a part that varies rapidly and, in a region around the window, mimics the previous high-momentum state. If this region is large enough, relativistic causality forbids information about the rest of the wave function from reaching the open window. Thus, if a photon comes out of the box, it must be highly energetic exactly as in the first experiment. A high-energy photon could thus be released even though the state inside the box had only low-energy components!

Where does the photon take the extra energy from? One is tempted to trace the missing energy to the trigger mechanism that briefly opens the window; surely, the action of
the trigger supplies energy to the photon. However, this answer is incorrect. When the box contained a high-energy photon, the trigger supplied negligible energy. Causality then implies that the same must be true for the second experiment.

One of the main accomplishments in this article is to provide a new answer to “Where does the photon take the extra energy from?” In addition, because these regions of superoscillations are created at the expense of having the function grow exponentially in other regions, it would be natural to conclude that the superoscillations would be quickly “over-taken” by tails coming from the exponential regions and would thus be short-lived. However, it has been shown that superoscillations are remarkably robust [9] and can last for a surprisingly long time. This has therefore led to practical proposals for applications of superoscillations to situations which were previously probed by evanescent waves (e.g. in the superresolution of very fine features). An evanescent wave occurs, e.g., when an EM wave enters a “forbidden” region. By way of example, suppose an EM wave enters a medium in the y-direction in which case \( k_y^2 \) can become negative and then \( k_y^2 \) can become larger than \( \omega^2 \) (using \( \omega^2 = k_x^2 + k_y^2 \), \( c = 1 \)) with the consequence that the amplitude decreases exponentially (in the y direction) inside the media. However, suppose at \( y = 0 \) we superpose different waves with the correct coefficients \( c_k \) (to be defined below), \( \sum_{|k| < 1} c_k e^{ik_y x} \) (where \( -\omega > k_y > \omega \), setting \( c = 1 \)), to obtain a superoscillation \( e^{i\alpha x} \) where \( |\alpha| \) can be arbitrarily larger than \( |\omega| \). Since \( y \) is analogous to time [9], the superoscillation standing wave can penetrate much deeper into the media than evanescent waves can (and for an arbitrary period of time).

Much has yet to be studied about the physical nature of superoscillations. In this paper, we uncover several new relationships between the physical creation of the high-momenta associated with the superoscillations, weak values, and modular variables which have been used to model topological effects [10–13]. E.g. we show a relationship between the exchange of modular variables and the creation of superfourier components. In addition we show how eccentric weak values, namely weak values that are outside the eigenvalue spectrum, are examples of superoscillations. Finally, we demonstrate how to measure the nonlocality in the equation of motion for modular variables by using weak measurements [35].

We shall consider the following wavefunction and argue that the results are general:

\[
\psi(x) = \left\{ \frac{1 - \alpha}{2} e^{-ip_x L/N} + \frac{1 + \alpha}{2} e^{ip_x L/N} \right\}^N
\] (1.2)

By performing a binomial expansion, we can see that this wavefunction is a superposition of waves with small wave numbers \( |k| \leq 1 \) (because \(-1 < (\frac{2n-N}{N}) < 1\):

\[
\psi(x) = \sum_{n=0}^{N} \frac{(1+\alpha)^n(1-\alpha)^{N-n}}{2^n n!(N-n)!} \exp \left\{ \frac{ip_x}{Np_0} \right\} \exp \left\{ -ip_x(N-n) \right\} \\
= \sum_{n=0}^{N} c_n \exp \left\{ \frac{ip_x(2n-N)}{p_0N} \right\} \\
= \sum_{n=0}^{N} c_n \exp \left\{ \frac{ip_x Np_0}{p_0N} \right\}
\] (1.3)

Alternatively, we can also expand for small \( |p_x L| \) and obtain:

\[
\psi(x) = \left\{ \frac{1 - \alpha}{2} (1 - ip_x L/N + ...) + \frac{1 + \alpha}{2} (1 + ip_x L/N + ...) \right\}^N \\
= e^{i\alpha p_x L} \left[ 1 + \frac{(p_x L)^2(\alpha^2 - 1)}{N} + O\left( \frac{1}{N^2} \right) \right]
\] (1.4)
That is, for a small region (which can include several wavelengths $2\pi/\alpha$, depending on how large one chooses $N$, the number of particles in the ensemble), $\psi(x)$ appears to have a very large momentum, since $\alpha$ can be arbitrarily large. The superoscillation, however, is created at the expense of having the function grow exponentially in other regions.

2. How to physically create superoscillations

Superoscillations were originally discovered by studying pre- and post-selection (PPS) in Quantum Mechanics (QM) [35]. Pre- and post-selection was originally probed with the time-symmetric re-formulation of Quantum Mechanics (TSQM, introduced by Aharonov, Bergmann and Lebowitz a/k/a ABL [1]).

The “time-asymmetry” attributed to the standard formulation of Quantum Mechanics (QM) was inherited from a reasonable tendency learned from Classical Mechanics (CM) to predict the future based on initial conditions: once the equations of motion are fixed in CM, then the initial and final conditions are not independent, only one can be fixed arbitrarily. In contrast, as a result of the uncertainty principle, the relationship between initial and final conditions within QM can be one-to-many: two “identical” particles with identical environments can subsequently exhibit different properties under identical measurements. These subsequent identical measurements provide fundamentally new information about the system which could not in principle be obtained from the initial conditions.

QM’s “time-asymmetry” is the assumption that measurements only have consequences after they are performed, i.e. towards the future. Nevertheless, a positive spin was placed on QM’s non-trivial relationship between initial and final conditions by ABL [1] who showed that the new information obtained from measurements was also relevant for the past of every quantum-system and not just the future. This inspired ABL to re-formulate QM in terms of Pre-and-Post-Selected-ensembles. The traditional paradigm for ensembles is to simply prepare systems in a particular state and thereafter subject them to a variety of experiments. These are “pre-selected-only-ensembles.” For pre-and-post-selected-ensembles, we add one more step, a subsequent measurement or post-selection. By collecting only a subset of the outcomes for this later measurement, we see that the “pre-selected-only-ensemble” can be divided into sub-ensembles according to the results of this subsequent “post-selection-measurement.” Because pre-and-post-selected-ensembles are the most refined quantum ensemble, they are of fundamental importance and subsequently led to the two-vector or Time-Symmetric re-formulation of Quantum Mechanics (TSQM) [4, 5]. TSQM provides a complete description of a quantum-system at a given moment by using two-wavefunctions, one evolving from the past towards the future (the one utilized in the standard paradigm) and a second one, evolving from the future towards the past.

While TSQM is a new conceptual point-of-view that has predicted novel, verified effects which seem impossible according to standard QM, TSQM is in fact a re-formulation of QM. Therefore, experiments cannot prove TSQM over QM (or vice-versa). The motivation to pursue such re-formulations, then, depends on their usefulness. The intention of this article is to answer this by discussing how TSQM fulfils several criterion which any re-formulation of QM should satisfy in order to be useful and interesting:

To be useful and interesting, any re-formulation of QM should meet several criteria such as those met by TSQM:

- TSQM is consistent with all the predictions made by standard QM,
- TSQM brings out features in QM that were missed before: e.g., ABL considered measurement situations between two successive Ideal Measurements (IMs) in which the transition from a state $|\Psi_{in}\rangle$ (pre-selected at a time $t_{in}$) to a state $|\Psi_{fin}\rangle$ (post-selected at a later time $t_{fin}$) is generally disturbed by an intermediate precise measurement. Post-selection reflects a unique aspect of QM in that measurement results are not determined by
equations of motion and initial conditions. A subsequent theoretical development arising out of the ABL work was the introduction of the “Weak Value” (WV) of an observable which was probed by a new type of quantum measurement called the “Weak Measurement” (WM) [4]. WM experiments have been performed and results are in very good agreement with theoretical predictions. One of the unusual properties of WVs (and a manifestation of superoscillations) is that they can lie far outside the eigenvalue spectrum.

- TSQM lead to simplifications in calculations (as occurred with the Feynman re-formulation) and stimulated discoveries in other fields: e.g. ABL influenced work in cosmology (e.g. Gell-Mann and Hartle [17]); in superluminal tunneling (Chiao [16] and Steinberg [18]); in quantum information [49] (e.g. the quantum random walk [19] or cryptography [20, 21]), etc.

- TSQM suggests generalizations of QM that were missed before - e.g. a new solution [22] to the quantum measurement problem.

2.1. Time-symmetric Quantum Mechanics

2.1.1. The Main Idea: TSQM contemplates measurements which occur at the present time \( t \) while the state is known both at \( t_{\text{in}} < t \) (past) and at \( t_{\text{fin}} > t \) (future). More precisely, we start at \( t = t_{\text{in}} \) with a measurement of a nondegenerate operator \( \hat{O}_{\text{in}} \). This yields as one potential outcome the state \( |\Psi_{\text{in}}\rangle \), i.e. we prepared the “pre-selected” state \( |\Psi_{\text{in}}\rangle \). At the later time \( t_{\text{fin}} \), we perform another measurement of a nondegenerate operator \( \hat{O}_{\text{fin}} \) which yields one possible outcome: the post-selected state \( |\Psi_{\text{fin}}\rangle \). At an intermediate time \( t \in [t_{\text{in}}, t_{\text{fin}}] \), we measure a non-degenerate observable \( A \) (for simplicity), with eigenvectors \( \{|a_j\rangle\} \). We wish to determine the conditional probability of \( a_j \), given that we have both boundary conditions, \( |\Psi_{\text{in}}\rangle \) and \( |\Psi_{\text{fin}}\rangle \).

1 To answer this, we use the time displacement operator: 
\[
U_{t_{\text{in}} \rightarrow t} = \exp\{-iH(t - t_{\text{in}})\}
\]
where \( H \) is the Hamiltonian for the free system. For simplicity, we assume \( H \) is time independent and set \( h = 1 \). The standard theory of collapse states that the system collapses into an eigenstate \( |a_j\rangle \) after the measurement at \( t \) with an amplitude \( \langle a_j|U_{t_{\text{in}} \rightarrow t}|\Psi_{\text{in}}\rangle \). The amplitude for our series of events is

\[
\alpha_j \equiv \langle \Psi_{\text{fin}}|U_{t_{\text{in}} \rightarrow t_{\text{fin}}}|a_j\rangle \langle a_j|U_{t_{\text{in}} \rightarrow t}|\Psi_{\text{in}}\rangle
\]

which is illustrated in figure 1.a. This means that the conditional probability to measure \( a_j \) given \( |\Psi_{\text{in}}\rangle \) is pre-selected and \( |\Psi_{\text{fin}}\rangle \) will be post-selected is given by the ABL formula [1] [2]:

\[
Pr(a_j, t|\Psi_{\text{in}}, t_{\text{in}}; \Psi_{\text{fin}}, t_{\text{fin}}) = \frac{\langle \Psi_{\text{fin}}|U_{t_{\text{in}} \rightarrow t_{\text{fin}}}|a_j\rangle \langle a_j|U_{t_{\text{in}} \rightarrow t}|\Psi_{\text{in}}\rangle^2}{\sum_{n} \langle \Psi_{\text{fin}}|U_{t_{\text{in}} \rightarrow t_{\text{fin}}}|a_n\rangle \langle a_n|U_{t_{\text{in}} \rightarrow t}|\Psi_{\text{in}}\rangle^2}
\]

(2.5)

As a first step toward understanding the underlying time-symmetry in the ABL formula, we consider the time-reverse of the numerator of eq. 2.6 and figure 1.a. First we apply \( U_{t_{\text{fin}} \rightarrow t_{\text{in}}} \) on \( \langle \Psi_{\text{fin}}|U_{t_{\text{in}} \rightarrow t_{\text{fin}}} = \langle U_{t_{\text{fin}} \rightarrow t_{\text{in}}}|\Psi_{\text{fin}} \rangle \) by using the well-known

1 Such an arrangement has long been considered in actual experiments: consider a bubble chamber scattering experiment. The incoming particle, \( |\Psi_{\text{in}}\rangle \), interacts with a target and then evolves into various outgoing states, \( |\Psi_{\text{fin}}\rangle \). Typically, photographs are not taken for every target-interaction, but only for certain ones that were triggered by subsequently interacting with detectors. In CM, there is (in principle) a one-to-one mapping between incoming states and outgoing states, whereas in QM, it is one-to-many. By selecting a single outcome for the post-selection-measurement, we define the pre-and-post-selected-ensemble that has no classical analog.

2 ABL is intuitive: \(|\langle a_j|U_{t_{\text{in}} \rightarrow t}|\Psi_{\text{in}}\rangle|^2 \) is the probability to obtain \( |a_j\rangle \) having started with \( |\Psi_{\text{in}}\rangle \). If \( |a_j\rangle \) was obtained, then the system collapsed to \( |a_j\rangle \) and \(|\langle \Psi_{\text{fin}}|U_{t_{\text{fin}} \rightarrow t_{\text{in}}}|a_j\rangle|^2 \) is then the probability to obtain \( |\Psi_{\text{fin}}\rangle \). The probability to obtain \( |a_j\rangle \) and \( |\Psi_{\text{fin}}\rangle \) then is \(|\alpha_j|^2 \). This is not yet the conditional probability since the post-s election may yield outcomes other than \( |\Psi_{\text{fin}}\rangle \). The probability to obtain \( |\Psi_{\text{fin}}\rangle \) is \( \sum_{j} |\alpha_j|^2 = |\langle \Psi_{\text{fin}}|\Psi_{\text{fin}}\rangle|^2 < 1 \). The question being investigated concerning probabilities of \( a_j \) at \( t \) assumes we are successful in obtaining the post-selection and therefore requires the denominator in eq. 2.6, \( \sum_{j} |\alpha_j|^2 \), which is a re-normalization to obtain a proper probability.
QM symmetry $U^\dagger_{t-t_{fin}} = \{e^{-iH(t_{fin}-t)}\}^\dagger = e^{iH(t_{fin}-t)} = e^{-iH(t-t_{fin})} = U_{t_{fin}-t}$. We also apply $U_{t_{fin}-t}$ on $\langle a_j |$ instead of on $|\Psi_{in}\rangle$ which yields the time-reverse re-formulation of the numerator of eq. 2.6, $\langle U_{t_{fin}-t} | \Psi_{fin} \rangle \langle a_j | U_{t-t_{fin}} | a_j \rangle | \Psi_{in} \rangle$ as depicted in fig. 1.b. Further work is needed to formulate what we mean by the 2-vectors in TSQM. E.g. if we are interested in the probability for possible outcomes of $a_j$ at time $t$, we must consider both $U_{t_{fin}-t} | \Psi_{in} \rangle$ and $\langle U_{t_{fin}-t} | \Psi_{fin} \rangle$, since these expressions propagate the pre-and-post-selection to the present time $t$ (see the conjunction of both figures 1.a and 1.b giving 1.c; these 2-vectors are not just the time-inverse of each other). This represents the basic idea behind the Time-Symmetric re-formulation of Quantum Mechanics (TSQM)\(^3\):

$$Pr(a_j, t | \Psi_{in}, t_{in}; \Psi_{fin}, t_{fin}) = \frac{|\langle U_{t_{fin}-t} | \Psi_{fin} \rangle \langle a_j | U_{t-t_{fin}} | a_j \rangle | \Psi_{in} \rangle|^2}{\sum_n |\langle U_{t_{fin}-t} | \Psi_{fin} \rangle | a_n \rangle \langle a_n | U_{t-t_{fin}} | \Psi_{in} \rangle|^2}$$ (2.6)

While this mathematical manipulation clearly proves that TSQM is consistent with QM, it yields a very different interpretation. For example, the action of $U_{t_{fin}-t}$ on $|\Psi_{fin}\rangle$ (i.e. $\langle U_{t_{fin}-t} | \Psi_{fin} \rangle$) can be interpreted to mean that the time displacement operator $U_{t_{fin}-t}$ sends $|\Psi_{fin}\rangle$ back in time from the time $t_{fin}$ to the present, $t$.

In summary, ABL formulation clarified a number of issues in QM. E.g.: in this formulation, both the probability and the amplitude are symmetric under the exchange of $|\Psi_{in}\rangle$ and $|\Psi_{fin}\rangle$. Therefore, the possibility of wavefunction collapse in QM does not necessarily imply irreversibility of an arrow of time at the QM level [50]. Nevertheless, the real litmus test of any re-formulation is whether conceptual shifts can teach us something fundamentally new or suggest generalizations of QM, etc. The re-formulation to TSQM suggested a number of new experimentally observable effects, one important example of which are weak-measurements (§2.1.1.3), which we now begin to motivate by considering strange pre-and-post-selection effects.

---

\(^3\) We note that because (full) collapses take place at the $t_{in}$ and $t_{fin}$ measurements, there is no meaning to information coming from $t > t_{fin}$ or $t < t_{in}$. Therefore, at least in this context, there is no meaning to a “multi-vector” formalism.

---

Figure 1. Time-reversal symmetry in probability amplitudes.
2.1.1.1. Pre-and-post-selection and Spin-1/2 The simplest, surprising, example of pre-and-post-selection is to pre-select a spin-1/2 system with \( |\Psi_{in}\rangle = |\hat{\sigma}_y = +1\rangle = |\uparrow\rangle_x \) at time \( t_{in} \). After the pre-selection, spin measurements in the direction perpendicular to \( x \) yields complete uncertainty in the result 4, so if we post-select at time \( t_{fin} \) in the \( y \)-direction, we obtain \( |\Psi_{fin}\rangle = |\hat{\sigma}_y = +1\rangle = |\uparrow\rangle_y \) half the time. Since the particle is free, the spin is conserved in time and thus for any \( t \in \{t_{in}, t_{fin}\} \), an ideal-measurement of either \( \hat{\sigma}_x \) or \( \hat{\sigma}_y \), yields +1 for this pre-and-post-selection. This by itself, two non-commuting observables known with certainty, is a most surprising property which no pre-selected-only-ensemble could possess. 5

We now ask a slightly more complicated question about the spin in a direction \( \xi = 45^\circ \) relative to the \( x - y \) axis. This yields:

\[
\hat{\sigma}_\xi = \hat{\sigma}_x \cos 45^\circ + \hat{\sigma}_y \sin 45^\circ = \frac{\hat{\sigma}_x + \hat{\sigma}_y}{\sqrt{2}} \tag{2.7}
\]

From the results \( Pr(\hat{\sigma}_x = +1) = 1 \) and \( Pr(\hat{\sigma}_y = +1) = 1 \), one might wonder why we couldn’t insert both values, \( \hat{\sigma}_x = +1 \) and \( \hat{\sigma}_y = +1 \) into eq. 2.7 and obtain \( \hat{\sigma}_\xi = \frac{1+1}{\sqrt{2}} = \frac{2}{\sqrt{2}} = \sqrt{2} \). Such a result is incorrect for an ideal-measurement because the eigenvalues of any spin operator, including \( \hat{\sigma}_\xi \), must be \( \pm 1 \). The inconsistency can also be seen by noting

\[
\left(\frac{\hat{\sigma}_x + \hat{\sigma}_y}{\sqrt{2}}\right)^2 = \frac{\hat{\sigma}_x^2 + \hat{\sigma}_y^2 + 2\hat{\sigma}_x \hat{\sigma}_y}{2} = \frac{1 + 1 + 0}{2} = 1.
\]

By implementing the above argument, we would expect \( \left(\frac{\hat{\sigma}_x + \hat{\sigma}_y}{\sqrt{2}}\right)^2 = \frac{1 + 1}{2} = 2 \neq 1 \). Performing this step of replacing \( \hat{\sigma}_x = +1 \) and \( \hat{\sigma}_y = +1 \) in eq. 2.7 can only be done if \( \hat{\sigma}_x \) and \( \hat{\sigma}_y \) commute, which would allow both values simultaneously to be definite. Although it appears we have reached the end-of-the-line with this argument, nevertheless, it still seems that there should be some sense in which \( Pr(\hat{\sigma}_x = +1) = 1 \) and \( Pr(\hat{\sigma}_y = +1) = 1 \) manifest themselves simultaneously to produce \( \hat{\sigma}_\xi = \sqrt{2} \).

2.1.1.2. Pre-and-post-selection and 3-Box-paradox Another example of a surprising pre-and-post-selection effect is the 3-box-paradox 40 which uses a single quantum particle that is placed in a superposition of 3 closed, separated boxes. The particle is pre-selected to be in the state \( |\Psi_{in}\rangle = \frac{1}{\sqrt{3}} (|A\rangle + |B\rangle + |C\rangle) \), where \( |A\rangle \), \( |B\rangle \) and \( |C\rangle \) denote the particle localized in boxes \( A \), \( B \), or \( C \), respectively. The particle is post-selected to be in the state \( |\Psi_{fin}\rangle = \frac{1}{\sqrt{3}} (|A\rangle + |B\rangle - |C\rangle) \). If an ideal-measurement is performed on box \( A \) in the intermediate time (e.g. we open the box), then the particle is found in box \( A \) with certainty. This is confirmed by ABL 1 probability for projection in \( A \): \( Pr(\hat{P}_A) = \frac{|\langle \Psi_{fin}|\hat{P}_A|\Psi_{in}\rangle|^2}{(|\langle \Psi_{fin}|\hat{P}_A|\Psi_{in}\rangle|^2 + |\langle \Psi_{fin}|\hat{P}_B + \hat{P}_C|\Psi_{in}\rangle|^2)} = 1 \). This can also be seen intuitively by contradiction: suppose we do not find the particle in box \( A \). In that case, since we do not interact with box \( B \) or \( C \), we would have to conclude that the state that remains after we didn’t find it in \( |A\rangle \) is proportional to \( |B\rangle + |C\rangle \). But this is orthogonal to the post-selection (which we know will definitely be obtained). Because this is a contradiction, we conclude that the particle must be found in box \( A \). Similarly, the probability to find the particle in box \( B \) is 1, i.e. \( Pr(\hat{P}_B = 1) = 1 \). The “paradox” is, what “sense” can these 2 definite statements be simultaneously true. We cannot detect the distinction with ideal-measurements: e.g. \( Pr(\hat{P}_A = 1) = 1 \) if only box \( A \) is opened, while \( Pr(\hat{P}_B = 1) = 1 \) if only box \( B \) is opened. If

4 E.g. in the \( z \)-basis the state is \( \frac{1}{\sqrt{2}} (|\uparrow_z\rangle + |\downarrow_z\rangle) \) which yields equal probability either spin-up or spin-down in the \( z \)-direction.

5 This is also evident from ABL: the probability to obtain \( \hat{\sigma}_\xi = +1 \) at the intermediate time if an ideal-measurement is performed is \( Pr(\hat{\sigma}_\xi = +1) = \frac{\cos(\xi) + \cos(\xi)\sin(\xi)}{1 + \cos(\xi)\sin(\xi)} \). We see that if \( \xi = 0^\circ \) (i.e. \( \hat{\sigma}_x \)) then the intermediate ideal-measurement will yield \( \hat{\sigma}_x = +1 \) with certainty and when \( \xi = 90^\circ \) (i.e. \( \hat{\sigma}_y \)), then the intermediate ideal-measurement will again yield \( \hat{\sigma}_y = +1 \) with certainty. E.g. \( \hat{\sigma}_{\xi=45} = \pm 1 \) is displayed in [39]'s figure 3.a.
ideal-measurements are performed on both box $A$ and box $B$, then obviously the particle will not be found in both boxes, i.e. $\hat{P}_A \hat{P}_B = 0$.\footnote{The mystery is increased by the fact that both $\hat{P}_A$ and $\hat{P}_B$ commute with each other, so how is it possible that measurement of one box can disturb the measurement of another?}

![Diagram](image)

**Figure 2.** a) pre-selected vector $|\Psi_{\text{fin}}\rangle = 1/\sqrt{3} (|A\rangle + |B\rangle + |C\rangle)$ propagates forwards in time from $t_{\text{fin}}$ to $t_1$, and post-selected vector $|\Psi_{\text{fin}}\rangle = 1/\sqrt{3} (|A\rangle + |B\rangle - |C\rangle)$ propagating backwards in time from $t_{\text{fin}}$ to $t_2$. b) ideal-measurement of $\hat{P}_A$ at $t_1$ and of $\hat{P}_B$ at $t_2$.

2.1.1.3. **Counterfactuals** There is a widespread tendency to “resolve” these paradoxes by pointing out that there is an element of counter-factual reasoning: the contradictions arise only because inferences are made that do not refer to actual experiments. Had the experiment actually been performed, then standard measurement theory predicts that the system would have been disrupted so that no paradoxical implications arises. Suppose we applied this to the 3-box-paradox: the resolution then is that there is no meaning to say that the particle is in both boxes without actually measuring both boxes during the intermediate time.

We have proven [7, 41] that one shouldn’t be so quick in throwing away counter-factual reasoning; though indeed counter-factual statements have no observational meaning, such reasoning is actually a very good pointer towards interesting physical situations. Without invoking counter-factual reasoning, we have shown that the apparently paradoxical reality implied counter-factually has new, experimentally accessible consequences. These observable consequences become evident in terms of weak measurements, which allow us to test - to some extent - assertions that have been otherwise regarded as counter-factual.

The main argument against counter-factual statements is that if we actually perform ideal-measurements to test them, we disturb the system significantly, and such disturbed conditions hide the counter-factual situation, so no paradox arises. TSQM also provides some novel insights for this “disturbance-based-argument”. E.g., for the spin-1/2 case (§2.1.1.1), if we verify $\hat{\sigma}_x$ at $t = t_1$ and $\hat{\sigma}_y$ at $t = t_2$, $t_1 < t_1 < t_2 < t_{\text{fin}}$, then $Pr(\hat{\sigma}_x = +1) = 1$ and $Pr(\hat{\sigma}_y = +1) = 1$ are simultaneously true. But if we switch the order and perform $\hat{\sigma}_y$ before $\hat{\sigma}_x$, then $Pr(\hat{\sigma}_x = +1) = 1$ and $Pr(\hat{\sigma}_y = +1) = 1$ are not simultaneously true, since measuring $\hat{\sigma}_y$ at time $t = t_1$ would not allow the information from the earlier ($t_{\text{fin}} < t$) pre-selection of $\hat{\sigma}_x = +1$ to propagate to the later time ($t_2 > t_1 > t_{\text{fin}}$) of the $\hat{\sigma}_x$ measurement. As a consequence, the $\hat{\sigma}_x$ measurement at time $t_2$ would yield both outcomes $\hat{\sigma}_x = \pm 1$.\footnote{The same argument applies in the reverse direction of time. The 4-outcomes are consistent with $|\Psi_{\text{fin}}\rangle = |\hat{\sigma}_x = +1\rangle$ and $|\Psi_{\text{fin}}\rangle = |\hat{\sigma}_y = +1\rangle$. Physically, the ideal-measurement of $\hat{\sigma}_z$ exposes the particle to a magnetic field with a strong gradient in the $\xi = 45^\circ$ direction, which causes the spin to revolve around this axis in an uncertain fashion.} So, in general, the finding that $\hat{\sigma}_x = +1$ with...
certainty or \( \sigma_y = +1 \) with certainty in the pre-and-post-selected ensemble only held when one of these two measurements was performed in the intermediate time, not both. Therefore, we should not expect both \( \sigma_y = +1 \) and \( \sigma_x = +1 \) when measured simultaneously through \( \hat{\sigma}_{\xi=45^\circ} \).

For the spin-1/2 case, the ABL-assignment relied on only the pre-or-post-selection, while in the 3-box-paradox, the ABL assignment relies on both the pre-and-post-selection. However, ABL still only gives an answer for one actual ideal-measurement. What happens if we tried to obtain two answers for the 3-box-paradox? In order to deduce \( \hat{P}_A = 1 \), we used information from both pre-and-post-selected vectors. When we actually measured \( \hat{P}_A \), then this ideal-measurement will limit the “propagation” of the 2-vectors that were relied on to make this determination (see fig. 2.b). If we subsequently were to measure \( \hat{P}_B \), then the necessary information from both the pre-and-post-selected vectors is no longer available (i.e. information from \( t_{in} \) cannot propagate beyond the ideal-measurement of \( \hat{P}_A \) at time \( t_1 \) due to the disturbance caused by the ideal-measurement of \( \hat{P}_A \)). Thus, even though \( \hat{P}_A \) and \( \hat{P}_B \) commute, ideal-measurements of one can disturb ideal-measurement of the other.\(^8\)

Since we have understood the reason why both statements are not simultaneously true as a result of disturbance, we can now see the “sense” in which the definite ABL assignments can be simultaneously relevant. Our main argument is that if one doesn’t perform absolutely precise (ideal) measurements but is willing to accept some finite accuracy, then one can bound the disturbance on the system. For example, according to Heisenberg’s uncertainty relations, a precise measurement of position reduces the uncertainty in position to zero \( \Delta x = 0 \) but produces an infinite uncertainty in momentum \( \Delta p = \infty \). On the other hand, if we measure the position only up to some finite precision \( \Delta x = \Delta \) we can limit the disturbance of momentum to a finite amount \( \Delta p \geq h/\Delta \). By replacing precise measurements with a bounded-measurement paradigm, counter-factual thought experiments become experimentally accessible. What we often find is that the paradox remains - measurements produce surprising and often strange, but nevertheless consistent structures. With limited-disturbance measurements, there is a sense in which both \( Pr(\sigma_x = +1) = 1 \) and \( Pr(\sigma_y = +1) = 1 \) are simultaneously relevant because measurement of one does not disturb the other. Since measurement of \( \sigma_z \) can also be understood as a simultaneous measurement of \( \sigma_x \) and \( \sigma_y \), we will see that with limited-disturbance measurements, we can simultaneously use both \( \sigma_x = +1 \) and \( \sigma_y = +1 \) to obtain

\[
(\sigma_{\xi=45^\circ})_w = \frac{1}{\sqrt{2}} \left( \frac{1}{\sqrt{2}} |\Psi_{fin}\rangle + \frac{1}{\sqrt{2}} |\Psi_{fin}\rangle \right) = \frac{\langle 1 \rangle |\Psi_{in}\rangle + \langle 1 \rangle |\Psi_{in}\rangle}{\sqrt{2} \langle 1 | \Psi_{in} \rangle} = \frac{\sqrt{2}}{\sqrt{2} |\Psi_{in}\rangle} = \sqrt{2}.
\]

2.1.2. Quantum Measurements: ABL considered the situation of measurements between two successive ideal-measurements where one transitions from a pre-selected state \( |\Psi_{in}\rangle \) to a post-selected state \( |\Psi_{fin}\rangle \). An intermediate state between \( |\Psi_{in}\rangle \) and \( |\Psi_{fin}\rangle \) is generally disturbed by an intermediate ideal-measurement. A subsequent theoretical development arising out of the ABL work was the introduction of the weak-value of an observable which can be probed by a new type of measurement called the weak-measurement [4]. The motivation behind these measurements is to explore the relationship between \( |\Psi_{in}\rangle \) and \( |\Psi_{fin}\rangle \) by reducing the disturbance on the system at the intermediate time. This is useful in many ways, e.g. if a weak-measurement of \( \hat{A} \) is performed at the intermediate time \( t \in [t_{in}, t_{fin}] \) then, in contrast to the ABL situation, the basic object in the entire interval \( t_{in} \rightarrow t_{fin} \) for the purpose of calculating other weak-values for other measurements is the pair of states \( |\Psi_{in}\rangle \) and \( |\Psi_{fin}\rangle \).

Weak-measurements [4] originally grew out of the quantum measurement theory developed

---

\(^8\) This is related to a violation of the product rule. In general, if \( |\Psi_1\rangle \) is an eigenvector of \( \hat{A} \) with eigenvalue \( a \) and \( |\Psi_2\rangle \) is an eigenvector of \( \hat{B} \) with eigenvalue \( b \) and \( [\hat{A}, \hat{B}] = 0 \), then if \( \hat{A} \) and \( \hat{B} \) are known only by either pre-selection or post-selection, then the product rule is valid, i.e. \( \hat{A}\hat{B} = ab \). However if \( \hat{A} \) and \( \hat{B} \) are known by both pre-selection and post-selection, then the product rule is not valid, i.e. \( \hat{A}\hat{B} \neq ab \), i.e. they can still disturb each other, even though they commute. [42]
by von Neumann [24]⁹. First we consider ideal-measurements of observable $\hat{A}$ by using an interaction Hamiltonian $H_{int}$ of the form $H_{int} = -\lambda(t)\hat{Q}_{md}\hat{A}$ where $\hat{Q}_{md}$ is an observable of the measuring-device (e.g. the position of the pointer) and $\lambda(t)$ is a coupling constant which determines the duration and strength of the measurement. For an impulsive measurement we need the coupling to be strong and short and thus take $\lambda(t) \neq 0$ only for $t \in (t_0 - \varepsilon, t_0 + \varepsilon)$ and set $\lambda = \int_{t_0 - \varepsilon}^{t_0 + \varepsilon} \lambda(t)dt$. We may then neglect the time evolution given by $H_s$ and $H_{md}$ in the complete Hamiltonian $H = H_s + md + H_{int}$. Using the Heisenberg equations-of-motion for the momentum $P_{md}$ of the measuring-device (conjugate to the position $\hat{Q}_{md}$), we see that $P_{md}$ evolves according to $\frac{dP_{md}}{dt} = \lambda(t)\hat{A}$. Integrating this, we see that $P_{md}(T) - P_{md}(0) = \lambda\hat{A}$, where $P_{md}(0)$ characterizes the initial state of the measuring-device and $P_{md}(T)$ characterizes the final state. To make a more precise determination of $\hat{A}$ requires that the shift in $P_{md}$, i.e. $\delta P_{md} = P_{md}(T) - P_{md}(0)$, be distinguishable from it’s uncertainty, $\Delta P_{md}$. This occurs, e.g., if $P_{md}(0)$ and $P_{md}(T)$ are more precisely defined and/or if $\lambda$ is sufficiently large (see figure 3.a). However, under these conditions (e.g. if the measuring-device approaches a delta function in $P_{md}$), then the disturbance or back-reaction on the system is increased due to a larger $H_{int}$, the result of the larger $\Delta Q_{md}$ ($\Delta Q_{md} \geq \frac{1}{\Delta P_{md}}$) When $\hat{A}$ is measured in this way, then any operator $\hat{O}$ ($\langle \hat{A}, \hat{O} \rangle \neq 0$) is disturbed because it evolved according to $\frac{d}{dt} \hat{O} = \lambda(t)\langle \hat{A}, \hat{O} \rangle \hat{Q}_{md}$, and since $\lambda \Delta Q_{md}$ is not zero, $\hat{O}$ changes in an uncertain way proportional to $\lambda \Delta Q_{md}$.

In the Schrödinger picture, the time evolution operator for the complete system from $t = t_0 - \varepsilon$ to $t = t_0 + \varepsilon$ is $\exp\{-i \int_{t_0 - \varepsilon}^{t_0 + \varepsilon} H(t)dt\} = \exp\{-i\lambda \int_{t_0 - \varepsilon}^{t_0 + \varepsilon} \hat{Q}_{md}\hat{A}\}$. This shifts $P_{md}$ (see figure 3.a). If before the measurement the system was in a superposition of eigenstates of $\hat{A}$, then the measuring-device will also be superposed proportional to the system. This leads to the “quantum measurement problems”. A conventional solution to this problem is to argue that because the measuring-device is macroscopic, it cannot be in a superposition, and so it will “collapse” into one of these states and the system will collapse with it.

2.1.2.1. Weakening the interaction between system and measuring device: Following our intuition we now perform measurements which do not disturb either the pre-or-post-selections. The interaction $H_{int} = -\lambda(t)\hat{Q}_{md}\hat{A}$ is weakened by minimizing $\lambda \Delta Q_{md}$. For simplicity, we consider $\lambda \ll 1$ (assuming without lack of generality that the state of the measuring-device is a Gaussian with spreads $\Delta P_{md} = \Delta Q_{md} = 1$). We may then set $e^{-i\lambda \hat{Q}_{md}\hat{A}} \approx 1 - i\lambda \hat{Q}_{md}\hat{A}$ and use a theorem [37]¹⁰:

$$\hat{A}\langle \Psi \rangle = \langle \hat{A} \rangle \langle \Psi \rangle + \Delta A \langle \Psi \perp \rangle,$$

(2.8)

to show that before the post-selection, the system state is:

$$e^{-i\lambda \hat{Q}_{md}\hat{A}}\langle \Psi_{in} \rangle = (1 - i\lambda \hat{Q}_{md}\langle \hat{A} \rangle)\langle \Psi_{in} \rangle = (1 - i\lambda \hat{Q}_{md}\langle \hat{A} \rangle)\langle \Psi_{in} \rangle - i\lambda \hat{Q}_{md}\Delta \hat{A}\langle \Psi_{in} \perp \rangle$$

(2.9)

Using the norm of this state $\| (1 - i\lambda \hat{Q}_{md}\hat{A})\langle \Psi_{in} \rangle \|^2 = 1 + \lambda^2 \hat{Q}_{md}^2 \langle \Delta A^2 \rangle$, the probability to leave

---

⁹ Weak-measurements and their outcome, weak-values, can be derived in all approaches to quantum measurement theory. E.g. the usual projective measurement typically utilized in quantum experiments is a special case of these weak-measurements [26].

¹⁰ E.g. in the spin-1/2 example, the conditions for an ideal-measurement $\delta P_{md}^c = \lambda \delta_{\sigma_z} \gg \Delta P_{md}^c$ will also necessitate $\Delta Q_{md}^c \gg \frac{\lambda}{\sigma_z}$, which will thereby create a back-reaction causing a precession in the spin such that $\Delta \hat{\hat{\theta}} \gg 1$ (i.e. more than one revolution), thereby destroying (i.e. making completely uncertain) the information that in the past we had $\sigma_z = +1$, and in the future we will have $\sigma_z = +1$.

¹¹ where $\langle \hat{A} \rangle = \langle \Psi_{in} | \hat{A} | \Psi_{in} \rangle$, $\langle \Psi | \hat{A} | \Psi \rangle$ is any vector in Hilbert space, $\Delta A^2 = \langle \Psi | (\hat{A} - \langle \hat{A} \rangle)^2 | \Psi \rangle$, and $\langle \Psi \perp \rangle$ is a state such that $\langle \Psi | \Psi \perp \rangle = 0$. 

Figure 3. a) with an ideal or “strong” measurement at \( t \) (characterized e.g. by \( \delta P_{\text{md}} = \lambda a_1 > \Delta P_{\text{md}} \)), then ABL gives the probability to obtain a collapse onto eigenstate \( a_1 \) by propagating \( |\Psi_{\text{fin}}\rangle \) backwards in time from \( t_{\text{fin}} \) to \( t \) and \( |\Psi_{\text{fin}}\rangle \) forwards in time from \( t_{\text{in}} \) to \( t \); in addition, the collapse caused by ideal-measurement at \( t \) creates a new boundary condition \( |a_1\rangle \langle a_1| \) at time \( t \in [t_{\text{in}}, t_{\text{fin}}] \); b) if a weak-measurement is performed at \( t \) (characterized e.g. by \( \delta P_{\text{md}} = \lambda a_w < \Delta P_{\text{md}} \)), then the outcome of the weak-measurement, the weak-value, can be calculated by propagating the state \( |\Psi_{\text{fin}}\rangle \) backwards in time from \( t_{\text{in}} \) to \( t \) and the state \( |\Psi_{\text{in}}\rangle \) forwards in time from \( t_{\text{in}} \) to \( t \); the weak-measurement does not cause a collapse and thus no new boundary condition is created at time \( t \).

\(|\Psi_{\text{in}}\rangle\) un-changed after the measurement is:

\[
\frac{1 + \lambda^2 \hat{Q}_{\text{md}}^2 \langle \hat{A} \rangle^2}{1 + \lambda^2 \hat{Q}_{\text{md}}^2 \langle \hat{A}^2 \rangle} \rightarrow 1 \quad (\lambda \to 0) \quad (2.10)
\]

while the probability to disturb the state (i.e. to obtain \( |\Psi_{\text{in}}\rangle \)) is:

\[
\frac{\lambda^2 \hat{Q}_{\text{md}}^2 \Delta \hat{A}^2}{1 + \lambda^2 \hat{Q}_{\text{md}}^2 \langle \hat{A}^2 \rangle} \rightarrow 0 \quad (\lambda \to 0) \quad (2.11)
\]

The final state of the measuring-device is now a superposition of many substantially overlapping Gaussians with probability distribution given by \( Pr(P_{\text{md}}) = \sum_i |\langle a_i|\Psi_{\text{in}}\rangle|^2 \exp \left\{ -\frac{(P_{\text{md}} - \lambda a_i)^2}{2\Delta P_{\text{md}}^2} \right\} \).

This sum is a Gaussian mixture, so it can be approximated by a single Gaussian \( \tilde{\Phi}_{\text{md}}(P_{\text{md}}) \approx \langle P_{\text{md}}| e^{-i\lambda \hat{Q}_{\text{md}}(\hat{A})} \Phi_{\text{md}}^{\text{in}} \rangle \approx \exp \left\{ -\frac{(P_{\text{md}} - \lambda \langle \hat{A} \rangle)^2}{2\Delta P_{\text{md}}^2} \right\} \text{ centered on } \lambda \langle \hat{A} \rangle \).

2.1.2.2. Information gain without disturbance: safety in numbers: It follows from eq. 2.11 that the probability for a collapse decreases as \( O(\lambda^2) \), but the measuring-device’s shift grows linearly \( O(\lambda) \), so \( \delta P_{\text{md}} = \lambda a_i \). For a sufficiently weak interaction (e.g. \( \lambda \ll 1 \)), the probability for a collapse can be made arbitrarily small, while the measurement still yields information but becomes less precise because the shift in the measuring-device is much smaller than its uncertainty \( \delta P_{\text{md}} \ll \Delta P_{\text{md}} \) (figure 3.b). If we perform this measurement on a single particle, then two non-orthogonal states will be indistinguishable. If this were possible, it would violate unitarity because these states could time evolve into orthogonal states \( |\Psi_1\rangle|\Phi_{\text{md}}^{\text{in}}\rangle \rightarrow |\Psi_1\rangle|\Phi_{\text{md}}^{\text{in}}(1)\rangle \) and \( |\Psi_2\rangle|\Phi_{\text{md}}^{\text{in}}\rangle \rightarrow |\Psi_2\rangle|\Phi_{\text{md}}^{\text{in}}(2)\rangle \), with \( |\Psi_1\rangle|\Phi_{\text{md}}^{\text{in}}\rangle \) orthogonal to \( |\Psi_2\rangle|\Phi_{\text{md}}^{\text{in}}\rangle \). With weakened measurement interactions, this does not happen because the measurement of these two non-orthogonal states causes a smaller shift in the measuring-device than it’s uncertainty. We conclude that the shift \( \delta P_{\text{md}} \) of the measuring-device is a measurement error because \( \tilde{\Phi}_{\text{md}}(P_{\text{md}}) = \langle P_{\text{md}} - \lambda \langle \hat{A} \rangle|\Phi_{\text{md}}^{\text{in}}\rangle \approx \langle P_{\text{md}}|\Phi_{\text{md}}^{\text{in}}\rangle \text{ for } \lambda \ll 1 \).

Nevertheless, if a large \( (N \geq \frac{\lambda}{\Delta P_{\text{md}}}) \) ensemble of particles is used, then the shift of all the measuring-devices \( \langle \delta P_{\text{md}}^{\text{tot}} \rangle \approx \lambda \langle \hat{A} \rangle \frac{N}{N} = N\langle \hat{A} \rangle \) becomes distinguishable because of repeated
integrations, while the collapse probability still goes to zero. That is, for a large ensemble of particles which are all either $|\Psi_2\rangle$ or $|\Psi_1\rangle$, this measurement can distinguish between them even if $|\Psi_2\rangle$ and $|\Psi_1\rangle$ are not orthogonal.\footnote{Because the scalar product $\langle \Psi_1^{(N)} | \Psi_2^{(N)} \rangle = \cos^n \theta \longrightarrow 0$.}

Using these observations, we now emphasize that the average of any operator $\hat{A}$, i.e. $\langle \hat{A} \rangle \equiv \langle \Psi | \hat{A} | \Psi \rangle$, can be obtained in three distinct cases\cite{43,51}:

(i) **Statistical method with disturbance:** the traditional approach is to perform ideal-measurements of $\hat{A}$ on each particle, obtaining a variety of different eigenvalues, and then manually calculate the usual statistical average to obtain $\langle \hat{A} \rangle$.

(ii) **Statistical method without disturbance:** as demonstrated by using $\hat{A}|\Psi\rangle = \langle \hat{A} | \Psi \rangle + \Delta A |\Psi_\perp\rangle$. We can also verify that there was no disturbance: consider the spin-1/2 example (§2.1.1.1), pre-selecting an ensemble, $|\uparrow_x\rangle$, then performing a weakened-measurement of $\hat{\sigma}_x$ and finally a post-selection again in the $x$-direction (figure 4). For every post-selection, we will again find $|\uparrow_x\rangle$ with greater and greater certainty (in the weakness limit), verifying our claim of no disturbance. Each measuring device is centered on $\langle \uparrow_x | \hat{\sigma}_x | \uparrow_x \rangle = \frac{1}{2} \sqrt{\delta P_{\text{ind}}}^2$ and the whole ensemble can be used to reduce the spread (see [39]'s figure 2.c and figure 5). The weakened interaction for $\hat{\sigma}_x$ means that the inhomogeneity in the magnetic field induces a shift in momentum which is less than the uncertainty $\delta P_{\text{ind}} < \Delta P_{\text{ind}}$ and thus a wave packet corresponding to $\frac{\delta_x + \delta_\perp}{\sqrt{2}} = 1$ will be broadly overlapping with the wave packet corresponding to $\frac{\delta_x + \delta_\perp}{\sqrt{2}} = -1$. A particular example is depicted in [39]'s figure 2.c and a general example is depicted in figure 5.

(iii) **Non-statistical method without disturbance:** is the case where $\langle \Psi | \hat{A} | \Psi \rangle$ is the "eigenvalue" of a single "collective operator," $\hat{A}^{(N)} = \frac{1}{N} \sum_{i=1}^{N} \hat{A}_i$ (with $\hat{A}_i$ the same operator $\hat{A}$ acting on the $i$-th particle). Using this, we are able to obtain information about $\langle \Psi | \hat{A} | \Psi \rangle$ without causing disturbance (or a collapse) and without using a statistical approach because any product state $|\Psi^{(N)}\rangle$ becomes an eigenstate of the operator $\hat{A}^{(N)}$. To see this, we apply the theorem $\hat{A}|\Psi\rangle = \langle \hat{A} | \Psi \rangle + \Delta A |\Psi_\perp\rangle$ [37] to $\hat{A}^{(N)}|\Psi^{(N)}\rangle$, i.e.:

$$\hat{A}^{(N)}|\Psi^{(N)}\rangle = \frac{1}{N} \left[ N\langle \hat{A} | \Psi^{(N)} \rangle + \Delta A \sum_i |\Psi_\perp^{(N)}(i)\rangle \right]$$

where $\langle \hat{A} \rangle$ is the average for any one particle and the states $|\Psi_\perp^{(N)}(i)\rangle$ are mutually orthogonal and are given by $|\Psi_\perp^{(N)}(i)\rangle = |\Psi_1\rangle|\Psi_2\rangle \ldots |\Psi_N\rangle$. That is, the $i$th state has particle $i$ changed to an orthogonal state and all the other particles remain in the same state. If we further define a normalized state $|\Psi_\perp^{(N)}\rangle = \sum_i \frac{1}{\sqrt{N}} |\Psi_\perp^{(N)}(i)\rangle$ then the last term of eq. 2.12 is $\frac{\Delta A}{\sqrt{N}} |\Psi_\perp^{(N)}\rangle$ and it's size is $\left| \frac{\Delta A}{\sqrt{N}} |\Psi_\perp^{(N)}\rangle \right|^2 \propto \frac{1}{N} \longrightarrow 0$. Therefore, $|\Psi^{(N)}\rangle$ becomes an eigenstate of $\hat{A}^{(N)}$, with the value $\langle \hat{A} \rangle$ and not even a single particle has been disturbed (as $N \rightarrow \infty$).

In the last case, the average for a single particle becomes a robust property over the entire ensemble, so a single experiment is sufficient to determine the average with great precision. There is no longer any need to average over results obtained in multiple experiments.

Tradition has dictated that when measurement interactions are limited so there is no disturbance on the system, when no information can be gained. However, we have now shown that when considered as a limiting process, the disturbance goes to zero more quickly than the shift in the measuring-device, which means for a large enough ensemble, information (e.g. the
expectation value) can be obtained even though not even a single particle is disturbed. This viewpoint thereby shifts the standard perspective on two fundamental postulates of QM.

2.1.2.3. Adding a post-selection to the weakened interaction: Weak-Values and Weak-Measurements: Having established a new measurement paradigm -information gain without disturbance- it is fruitful to inquire whether this type of measurement reveals new values or properties. With weak-measurements (which involve adding a post-selection to this ordinary -but weakened- von Neumann measurement), the measuring-device registers a new value, the weak-value. As an indication of this, we insert a complete set of states \( \{ |\Psi_{\text{fin}} j\rangle \} \) into the outcome of the weak interaction of §2.1.2.1 (i.e. the expectation value \( \langle \hat{A} \rangle \)):

\[
\langle \hat{A} \rangle = \langle \Psi_{\text{in}} | \left( \sum_j |\Psi_{\text{fin}} j\rangle \langle \Psi_{\text{fin}} j| \right) \hat{A} |\Psi_{\text{fin}} \rangle = \sum_j |\langle \Psi_{\text{fin}} j| \Psi_{\text{in}} \rangle|^2 \frac{\langle \Psi_{\text{fin}} j| \hat{A} |\Psi_{\text{fin}} \rangle}{\langle \Psi_{\text{fin}} j| \Psi_{\text{fin}} \rangle} \tag{2.13}
\]

If we interpret the states \( |\Psi_{\text{fin}} j\rangle \) as the outcomes of a final ideal-measurement on the system (i.e. a post-selection) then performing a weak-measurement (e.g. with \( \lambda \Delta Q_{\text{md}} \rightarrow 0 \)) during the intermediate time \( t \in [t_{\text{fin}}, t_{\text{in}}] \), provides the coefficients for \( |\langle \Psi_{\text{fin}} j| \Psi_{\text{in}} \rangle|^2 \) which gives the probabilities \( \text{Pr}(j) \) for obtaining a pre-selection of \( |\Psi_{\text{in}} \rangle \) and a post-selection of \( |\Psi_{\text{fin}} j\rangle \). The intermediate weak-measurement does not disturb these states and the quantity \( A_w(j) \equiv \frac{\langle \Psi_{\text{fin}} j| \hat{A} |\Psi_{\text{fin}} \rangle}{\langle \Psi_{\text{fin}} j| \Psi_{\text{fin}} \rangle} \) is the weak-value of \( \hat{A} \) given a particular final post-selection \( |\Psi_{\text{fin}} j\rangle \). Thus, from the definition \( \langle \hat{A} \rangle = \sum_j \text{Pr}(j) A_w(j) \), one can think of \( \langle \hat{A} \rangle \) for the whole ensemble as being constructed out of sub-ensembles of pre-and-post-selected-states in which the weak-value is multiplied by a probability for a post-selected-state.

The weak-value arises naturally from a weakened measurement with post-selection: taking \( \lambda << 1 \), the final state of measuring-device in the momentum representation becomes (see also Appendix Appendix A):

\[
\langle P_{\text{md}} | e^{-i \lambda Q_{\text{md}} \hat{A}} |\Psi_{\text{fin}} \rangle |\Phi_{\text{MD}} \rangle \approx \langle P_{\text{md}} | \Psi_{\text{fin}} \rangle \langle 1 + i \lambda Q_{\text{md}} \hat{A} |\Psi_{\text{fin}} \rangle |\Phi_{\text{MD}} \rangle \approx \langle P_{\text{md}} | \Psi_{\text{fin}} \rangle \{ 1 + i \lambda Q_{\text{md}} \langle \Psi_{\text{fin}} | \hat{A} |\Psi_{\text{fin}} \rangle \} |\Phi_{\text{MD}} \rangle \approx \langle \Psi_{\text{fin}} | \Psi_{\text{in}} \rangle \langle P_{\text{md}} | e^{-i \lambda Q_{\text{md}} A_w} |\Phi_{\text{MD}} \rangle \rightarrow \langle \Psi_{\text{fin}} | \Psi_{\text{in}} \rangle \exp \left\{ -\left( P_{\text{md}} - \lambda A_w \right)^2 \right\} \tag{2.14}
\]

13This is also helpful to understand the quantum to classical transition because typical classical interactions involve these collective observables which do not disturb each other.
Figure 5. a)-c): Standard Ideal Measurement, a) short impulsive coupling, b) initial and final wavefunctions for measuring device, c) conjugate momentum variable; d)-f): Measurement with a weak interaction without a post-selection, d) short impulsive coupling, e) initial and final wavefunctions for measuring device, f) conjugate momentum variable.
where \( A_w = \frac{\langle \Psi_{\text{fin}}|\hat{A}|\Psi_{\text{in}}\rangle}{\langle \Psi_{\text{fin}}|\Psi_{\text{in}}\rangle} \)

The final state of the measuring-device is almost un-entangled with the system; it is shifted by a very unusual quantity, the weak-value, \( A_w \), which is not in general an eigenvalue of \( \hat{A} \). We have used such limited disturbance measurements to explore many paradoxes (see, e.g. [7, 41, 44, 45]). A number of experiments have been performed to test the predictions made by weak-measurements and results have proven to be in very good agreement with theoretical predictions [25, 27–30]. Since eigenvalues or expectation values can be derived from weak-values [46], we believe that the weak-value is indeed of fundamental importance in QM. In addition, the weak-value is the relevant quantity for all generalized weak interactions with an environment, not just measurement interactions. The only requirement being that the 2-vectors, i.e. the pre-and-post-selection, are not significantly disturbed by the environment.

2.1.3. Fundamentally new features of weak-values

2.1.3.1. Weak-values and 3-box-paradox: Returning to the 3-box-paradox (§2.1.1.2), we can calculate the weak-values of the number of particles in each box, e.g.:

\[
(|A\rangle\langle A|)_w = \frac{\langle \Psi_{\text{fin}}|A\rangle\langle A|\Psi_{\text{in}}\rangle}{\langle \Psi_{\text{fin}}|\Psi_{\text{in}}\rangle} = \frac{1}{\sqrt{3}} \{\langle A| + \langle B| - \langle C|\} = \frac{1}{\sqrt{3}} \{\langle A| + \langle B| + \langle C|\} \Rightarrow \frac{1}{\sqrt{3}} = \frac{1}{2}
\]

However, we can more easily ascertain the weak-values without calculation due to the following theorems:

**Theorem 1:** The sum of the weak-values is equal to the weak-value of the sum (see Appendix G):

\[
if \ (\hat{P}_A)_w = (\hat{P}_B + \hat{P}_C)_w then \ (\hat{P}_A)_w = (\hat{P}_B)_w + (\hat{P}_C)_w \tag{2.15}
\]

**Theorem 2** (see Appendix H): If a single ideal-measurement of an observable \( \hat{P}_A \) is performed between the pre-and-post-selection, then if the outcome is definite (e.g. \( \text{Prob}(\hat{P}_A = 1) = 1 \)) then the weak-value is equal to this eigenvalue (e.g. \( (\hat{P}_A)_w = 1 \)) [5].

This also provides a direct link to the counterfactual statements (§2.1.1.3) because all counterfactual statements which claim that something occurs with certainty, and which can actually be experimentally verified by separate ideal-measurements, continue to remain true when tested by weak-measurements. However, given that weak-measurements do not disturb each other, all these statements can be measured simultaneously.

Applying Theorem 2 to the 3-box-paradox, we know the following weak-values with certainty:

\[
(\hat{P}_A)_w = 1, \ (\hat{P}_B)_w = 1, \ (\hat{P}_{\text{total}}) = (\hat{P}_A + \hat{P}_B + \hat{P}_C)_w = 1. \tag{2.16}
\]

Using theorem 1, we obtain:

\[
(\hat{P}_C)_w = \frac{\langle \Psi_{\text{fin}}|\hat{P}_{\text{total}} - \hat{P}_A - \hat{P}_B|\Psi_{\text{in}}\rangle}{\langle \Psi_{\text{fin}}|\Psi_{\text{in}}\rangle} = (\hat{P}_A + \hat{P}_B + \hat{P}_C)_w - (\hat{P}_A)_w - (\hat{P}_B)_w = -1. \tag{2.17}
\]

\(14\) Thereby challenging another fundamental postulate of QM.
This surprising theoretical prediction of TSQM has been verified experimentally using photons [47]. What interpretation should be given to $(\hat{P}_C)_w = -1$? Any weak-measurement which is sensitive to the projection operator $\hat{P}_C$ will register the opposite effect from those cases in which the projection operator is positive, e.g. a weak-measurement of the amount of charge in box $C$ in the intermediate time will yield a negative charge (assuming it is a positively charged particle). For numerous reasons, we believe the most natural interpretation is: there are $-1$ particles in box $C$.

2.1.3.2. How the weak-value of a spin-$1/2$ can be 100: The weak-value for the spin-$1/2$ considered in §2.1.1.1 (which was confirmed experimentally for an analogous observable, the polarization [30]) is:

$$
(\hat{\sigma}_{\xi=45^\circ})_w = \frac{\langle \uparrow_y | \hat{\sigma}_y + \hat{\sigma}_x | \uparrow_x \rangle}{\langle \uparrow_y | \uparrow_x \rangle} = \frac{\{\{ \uparrow_y | \hat{\sigma}_y \} + \{ \hat{\sigma}_x | \uparrow_x \} \}}{\sqrt{2} \langle \uparrow_y | \uparrow_x \rangle} = \frac{\langle \uparrow_y | 1 \rangle}{\sqrt{2} \langle \uparrow_y | \uparrow_x \rangle} = \sqrt{2} \quad (2.18)
$$

Normally, the component of spin $\hat{\sigma}_z$ is an eigenvalue, $\pm 1$, but the weak-value $(\hat{\sigma}_z)_w = \sqrt{2}$ is $\sqrt{2}$ times bigger, (i.e. lies outside the range of eigenvalues of $\hat{\sigma} \cdot \mathbf{n}$) 15. How do we obtain this? Instead of post-selecting $\hat{\sigma}_x = 1$ (figure 4), we post-select $\hat{\sigma}_y = 1$ which will be satisfied in one-half the trials (figure 6).16

![Figure 6. Statistical weak-measurement ensemble.](image)

To show this in an actual calculation, we use eq. A.6 and the post-selected state of the quantum system in the $\sigma_\xi$ basis ($| \uparrow_y \rangle \equiv \cos(\pi/8)| \uparrow_\xi \rangle - \sin(\pi/8)| \downarrow_\xi \rangle$), the MD probability distribution is:

$$
Pr(P_{md}) = N^2[\cos^2(\pi/8)e^{-(P_{md}-1)^2/\Delta^2} - \sin^2(\pi/8)e^{-(P_{md}+1)^2/\Delta^2}]^2 \quad (2.19)
$$

With a strong measurement, $\Delta \ll 1$, the distribution is localized again around the eigenvalues $\pm 1$, as illustrated in [39]'s figure 3.a and 3.b, similar to what occurred in [39]'s figure 2.a and figure 5. What is different, however, is that when the measurement is weakened, i.e. $\Delta$ is made

15 Weak-values even further outside the eigenvalue spectrum can be obtained by post-selecting states which are more anti-parallel to the pre-selection: e.g. if we post-select the $+1$ eigenstate of $(\cos \alpha)\sigma_x + (\sin \alpha)\sigma_z$, then $(\hat{\sigma}_z)_w = \lambda \tan \frac{\alpha}{2}$, yielding arbitrarily large values such as spin-100.

16 If a post-selection does not satisfy $\hat{\sigma}_y = +1$, then that member of the sub-ensemble must be discarded. This highlights a fundamental difference between pre-and-post-selection due to the macroscopic arrow-of-time: in contrast to post-selection, if the pre-selection does not satisfy the criteria, then a subsequent unitary transformation unitary transformation can transform to the proper criteria.
larger, then the distribution changes to one single distribution centered around $\sqrt{2}$, the weak-value, as illustrated in [39]'s figure 3.c-f, the width of which can again be reduced when we consider an ensemble [39]'s figure 3.f. Using eq. 2.13, we can see that the weak-value is just the pre-and-post-selected sub-ensemble arising from within the pre-selected-only ensembles. That is, [39]'s figure 3.f is sub-ensemble within the expectation value, see [39]'s figure 2.c.

The non-statistical aspect mentioned in case-3 (§2.1.2.2) can also be explored by changing the problem slightly. Instead of considering an ensemble of spin-1/2 particles, we now consider “particles” which are composed of many ($N$) spin-1/2 particles, as occurs with a ferromagnet. The spin in any direction can then be obtained by measuring the magnetic-field with a compass, which is not a very precise measurement. If the compass were to flip a few spins, then this would be insignificant compared to other uncertainties (such as the position of the compass), and since these measurements barely disturb the ferromagnet, all directions commute. For example, we could measure the ferromagnet with an error of $\sqrt{N}$ while disturbing at most $\sqrt{N}$ spins which is insignificant compared to large $N$. This is therefore a relatively precise measurement which produces little disturbance and we expect $\hat{\sigma}_{45\circ} \approx \frac{N+\sqrt{N}}{\sqrt{2}} \pm O(\sqrt{N}) = \sqrt{2}N \pm O(\sqrt{N})$.

More precisely, we perform a weak-measurement of the collective observable $\hat{\sigma}_{45\circ}^{(N)} \equiv \frac{1}{N} \sum_{i=1}^{N} \hat{\sigma}_{45\circ}^{(i)}$ in the $45\circ$-angle to the $x-y$ plane. Using $H_{\text{int}} = -\frac{\Delta(t)}{N} \hat{Q}_{\text{md}} \sum_{i=1}^{N} \hat{\sigma}_{45\circ}^{(i)}$, a particular pre-selection of $|x\rangle$ (i.e. $|\Psi_{\text{in}}^{(N)}\rangle = \prod_{j=1}^{N} |x\rangle_{j}$) and post-selection $|y\rangle$ (i.e. $|\Psi_{\text{in}}^{(N)}\rangle = \prod_{k=1}^{N} |y\rangle_{k} = \prod_{n=1}^{N} (|x\rangle_{n} + i|y\rangle_{n})$), the final state of the measuring-device is:

$$|\Phi_{\text{fin}}^{\text{MD}}\rangle = \prod_{j=1}^{N} (|y\rangle_{j} \exp \left\{ \frac{\lambda}{N} \hat{Q}_{\text{md}} \sum_{k=1}^{N} \hat{\sigma}_{k}^{(j)} \right\} \prod_{i=1}^{N} |x\rangle_{i} |\Phi_{\text{in}}^{\text{MD}}\rangle)$$

(2.20)

Since the spins do not interact with each other, we can calculate one of the products and take the result to the $N$th power:

$$|\Phi_{\text{fin}}^{\text{MD}}\rangle = \prod_{j=1}^{N} (|y\rangle_{j} \exp \left\{ \frac{\lambda}{N} \hat{Q}_{\text{md}} \hat{\sigma}_{y} \right\} |\Phi_{\text{in}}^{\text{MD}}\rangle) = \left\{ (|y\rangle \exp \left\{ \frac{\lambda}{N} \hat{Q}_{\text{md}} \hat{\sigma}_{y} \right\} |x\rangle \right\}^{N} |\Phi_{\text{in}}^{\text{MD}}\rangle$$

(2.21)

Using the following identity $\exp \{ i\alpha \hat{\sigma}_{45\circ} \} = \cos \alpha + i \hat{\sigma}_{45\circ} \sin \alpha$ [38], this becomes:

$$|\Phi_{\text{fin}}^{\text{MD}}\rangle = \left\{ (|y\rangle \cos \left\{ \frac{\lambda}{N} \hat{Q}_{\text{md}} \right\} \sin \left\{ \frac{\lambda}{N} \hat{Q}_{\text{md}} \hat{\sigma}_{y} \right\} |x\rangle \right\}^{N} |\Phi_{\text{in}}^{\text{MD}}\rangle$$

(2.22)

where we have substituted $\alpha_{w} \equiv (\hat{\sigma}_{y})_{w} = \langle \langle y | \hat{\sigma}_{y} | x \rangle \rangle$. We consider only the second part (the first bracket, a number, can be neglected since it does not depend on $\hat{Q}$ and thus can only affect the normalization):

$$|\Phi_{\text{in}}^{\text{MD}}\rangle \approx \left\{ 1 - \frac{\lambda^{2} \langle \hat{Q}_{\text{md}} \rangle^{2}}{N^{2}} - \frac{i \lambda \alpha_{w} \langle \hat{Q}_{\text{md}} \rangle}{N} \right\}^{N} |\Phi_{\text{in}}^{\text{MD}}\rangle \approx e^{i \lambda \alpha_{w} \langle \hat{Q}_{\text{md}} \rangle} |\Phi_{\text{in}}^{\text{MD}}\rangle$$

(2.23)

When$^{17}$ projected onto $P_{\text{md}}$, i.e. the pointer, we see that the pointer is robustly shifted by the the same weak-value obtained with the previous statistical method, i.e. $\sqrt{2}$:

$$(\hat{\sigma}_{y})_{w} = \frac{\prod_{k=1}^{N} |y\rangle_{k} \sum_{i=1}^{N} \{ \hat{\sigma}_{x}^{(i)} + \hat{\sigma}_{y}^{(i)} \} \prod_{j=1}^{N} |x\rangle_{j}}{\sqrt{2} \ N(|\langle y | x \rangle|)^{N}} = \sqrt{2} \pm O\left( \frac{1}{\sqrt{N}} \right).$$

(2.24)

$^{17}$ The last approximation was obtained as $N \rightarrow \infty$, using $(1 + \frac{a}{N})^{N} = (1 + \frac{a}{N})^{\frac{N}{a}} \approx e^{a}$. 

16
A single experiment is now sufficient to determine the weak-value with great precision and there is no longer any need to average over results obtained in multiple experiments as we did in the previous section. Therefore, if we repeat the experiment with different measuring-devices, then each measuring-device will show the very same weak-values, up to an insignificant spread of \( \frac{1}{N} \) and the information from both boundary conditions, i.e. \( |\Psi_{in}\rangle = \prod_{i=1}^{N}|x_i\rangle \) and \( |\Psi_{fin}\rangle = \prod_{i=1}^{N}|y_i\rangle \), describes the entire interval of time between pre-and-post-selection. Following [39], we consider an example with \( N = 20 \). The probability distribution of the measuring-device after the post-selection is:

\[
\text{prob}(Q^{(N)}_{md}) = N^2 \left( \sum_{i=1}^{N} (-1)^i (\cos^2(\pi/8))^{N-i} (\sin^2(\pi/8))^{i} e^{-i Q^{(N)}_{md} (2-2i)/(2\Delta^2)} \right)^2.
\]

and is drawn for different values of \( \Delta \) in [39]’s figure 4. While this result is rare, we have recently shown [43,51] how any ensemble can yield robust weak-values like this in a way that is not rare and for a much stronger regime of interaction. We have thereby shown that weak-values are a general property of every pre-and-post-selected ensemble. \(^{18}\)

2.2. Superoscillations as weak values

Superoscillations were originally discovered through the study of weak-values. By way of example, consider again eq. 2.22:

\[
|\Phi_{fin}^{MD}\rangle = \left\{ \cos \frac{\lambda Q_{md}}{N} - i \alpha w \sin \frac{\lambda Q_{md}}{N} \right\}^N |\Phi_{in}^{MD}\rangle
\]

\[
= \left\{ \frac{e^{i \lambda Q_{md}}}{2} - \frac{e^{-i \lambda Q_{md}}}{2} + \frac{i \alpha w}{N} \left( e^{i \lambda Q_{md}} - e^{-i \lambda Q_{md}} \right) \right\}^N |\Phi_{in}^{MD}\rangle
\]

\[
\equiv \psi(x) \text{ i.e. eq. 1.2}
\]

We already saw how this could be approximated as \( e^{i \lambda \alpha w Q_{md}} |\Phi_{in}^{MD}\rangle \) which produced a robust-shift in the measuring-device by the weak-value \( \sqrt{2} \). However, the bracket in eq. 2.26 is identical to eq. 1.2, so we can also view \( \psi(x) = \left\{ e^{i \lambda Q_{md} \frac{1 + \alpha w}{2}} + e^{-i \lambda Q_{md} \frac{1 - \alpha w}{2}} \right\}^N \) in a very different way, by performing a binomial expansion:

\[
\psi(x) = \sum_{n=0}^{N} \frac{(1+\alpha w)^n(1-\alpha w)^{N-n}}{2^N} \frac{N!}{n!(N-n)!} \exp \left\{ \frac{in\lambda Q_{md}}{N} \right\} \exp \left\{ \frac{-i\lambda Q_{md} \lambda}{2N} \right\}
\]

\[
= \sum_{n=0}^{N} c_n \exp \left\{ \frac{i\lambda Q_{md}(2n-N)}{N} \right\} = \sum_{n=0}^{N} c_n \exp \left\{ \frac{i\lambda Q_{md} n}{N} \right\}
\]

We see that this wavefunction is a superposition of waves with small wavenumbers \( |k| \leq 1 \) (because \(-1 < \frac{2n-N}{N} < 1 \)). For a small region (which can include several wavelengths \( 2\pi/\alpha w \), depending on how large one chooses \( N \)), \( \psi(x) \) appears to have a very large momentum, since \( \alpha w \) can be arbitrarily large, i.e. a super-oscillation. However, the super oscillation is created at the

\(^{18}\) We have also proposed this as another innovative new laser-technology, e.g. in the amplification of small non-random signals by minimizing uncertainties in determining the weak value and by minimizing sample size. [43,51]
expense of having the function grow exponentially in other regions as can be seen by considering
the size of the function:

\[ \Phi_{\text{fin}}^{\text{md}} = \{1 + (\alpha^2 - 1) \sin^2 \frac{\lambda Q_{\text{md}}}{N}\}^{\frac{\Delta}{A}} \exp \left\{ -\frac{(Q_{\text{md}})^2}{2(\Delta Q_{\text{md}})^2}\right\} \]

(2.28)

For large \( Q_{\text{md}} \), the \( A \) term grows exponentially as \( \alpha^N \). Because these regions of superoscillations are created at the expense of having the function grow exponentially in other regions, it would be natural to conclude that the superoscillations would be quickly “over-taken” by tails coming from the exponential regions and would thus be short-lived. However, it has been shown that superoscillations are remarkably robust [9] and can last for a surprisingly long time. This has therefore led to proposed/practical applications of superoscillations to situations which were previously probed by evanescent waves (e.g. in the superresolution of very fine features with lasers).

As we mentioned in the introduction, TSQM is a re-formulation of QM, and therefore it must be possible to view the novel effects from the traditional single-vector perspective. This is precisely what super-oscillations teach us. In summary, there are 2 ways to understand weak-values:

- the measuring-device is registering the weak-value as a property of the system as characterized by TSQM
- the weak-value is a result of a complex interference effect in the measuring-device; the system continues to be described with a single-vector pursuant to QM

Oftentimes, calculations are either much simplified or can only be performed by utilizing the first approach (e.g. when the measuring-device is classical) [41].

2.2.1. Quantum Random Walk: Another fundamental discovery arising out of TSQM is the Quantum-Random-Walk [19] which has also stimulated discoveries in other areas of physics (for a review, see [48]). In the second bullet above, the measuring-device is shifted by the operator \( \sigma_\xi^{(N)} \) with it’s \( N + 1 \) eigenvalues equally spaced between \(-1\) and \(+1\) [41]. How can a superposition of small shifts between \(-1\) and \(+1\) give a shift that is arbitrarily far outside \( \pm 1 \)? The answer is that states of the measuring-device interfere constructively for \( \hat{P}_{\text{md}}^{(N)} = \alpha_w \) and destructively for all other values of \( \hat{P}_{\text{md}}^{(N)} \) such that \( \Phi_{\text{fin}}^{\text{MD}}(P) \rightarrow \Phi_{\text{inf}}^{\text{MD}}(P - A_w) \), the essence of quantum-random-walk [19]. If the coefficients for a step to the left or right were probabilities, as would be the case in a classical random walk, then \( N \) steps of step size 1 could generate an average displacement of \( \sqrt{N} \), but never a distance larger than \( N \). However when the steps are superposed with probability amplitudes, as with the quantum-random-walk, and when one considers probability amplitudes that are determined by pre-and-post-selection, then the random walk can produce any displacement. In other words, instead of saying that a “quantum step” is made up of probabilities, we say that a quantum step is a superposition of the amplitude for a step “to the left” and the amplitude for a “step to the right,” then one can superpose small Fourier components and obtain a large shift. This phenomenon is very general: if \( f(t-a_n) \) is a function shifted by small numbers \( a_n \), then a superposition can produce the same function but shifted by a value \( \alpha \) well outside the range of \( a_n \): \( \sum_{n=0}^{N} c_n f(t-a_n) \approx f(t-\alpha) \). The same values of \( a_n \) and \( c_n \) are appropriate for a wide class of functions and this relation can be made arbitrarily precise by increasing the number of terms in the sum, see [39]’s figure 5. The key to this phenomenon is the extremely rapid oscillations in the coefficients \( c_n \equiv \frac{(1+\alpha)(1-\alpha)^N - n}{2^n n!(N-n)!} \) in \( \sum_{n=0}^{N} c_n \exp \left\{ i\lambda Q_{\text{md}}k_n \right\} \).
2.3. **Is the high-momenta really there?**
Suppose we were to create a photon described by the wavefunction eq. 1.2 (e.g. a superposition of infra-red frequencies [31]) and put it in a box [35]. Suppose the box is prepared with a window and a shutter placed in the region where a photon is in the superoscillatory state. Suppose also that the shutter is opened for a short period of time. In the unlikely event that the photon is emitted, then it would be a gamma ray (with real high-momenta, with corrections due to truncation of the wavefunction which can be minimized by increasing the time the shutter is left open.) We could then arrange a situation in which relativistic causality would prevent information from the other parts of the wavefunction from reaching the window.

However, when the particle is still in the box and thus described by the entire quantum wave eq. 1.2, then we must continue to consider the interference between the superoscillatory portion of the wave and the other portions, in which case, from the perspective of the binomial distribution of eq. 2.27, there is no “high-momenta,” only superpositions of small momenta.

Where then did the high energy of the gamma ray come from since it was not contained in the original wavefunction? The energy had to come from the shutter which opens the window. The problem is that the shutter would have no effect if there was a true gamma photon there. The shutter shouldn’t be able to tell that the photon is actually a superoscillatory function (i.e. by the far-away tails). The situation can be made even more extreme by slightly modifying eq. 1.2, \( \alpha \rightarrow \alpha + \delta \alpha \):

\[
\Psi = \int \psi(x) \exp \{i(\alpha - \delta \alpha)x\} \exp \left\{ \frac{\delta \alpha^2}{2\Delta^2} \right\} d\delta \alpha
\]

\[
= e^{i\alpha x} \int \psi(x) \exp \{i\delta \alpha x\} \exp \left\{ \frac{\delta \alpha^2}{\Delta^2} \right\} d\delta \alpha = e^{i\alpha x} e^{-x^2\Delta^2}
\]

(2.29)

In this case, the wavefunction goes to zero just outside the region of superoscillation. Therefore, there is practically no connection between the region of superoscillation and the exponential region of the wavefunction. Yet still, the window will emit a gamma photon!

3. **Relationship between modular variables and superoscillations**
In this section, we present a new answer to the question: “Where does the photon take the extra energy from?” [35] All superoscillatory phenomenon have the property that you see something in one region of space that does not exist everywhere. That is, if we look at the entire wavefunction, we will not see the hi-momentum. But if we project the particle onto the region of superoscillation, then the hi-momentum is found. We will argue that an uncertain phase, which characterizes the process of projecting the particle into a certain region, can create the hi-momentum. If an uncertain phase can localize the particle, then a definite phase can also localize it. This corresponds to the exchange of the modular variable. Again, the original wavefunction was a superposition of allowed eigenvalues with different amplitudes. If the photon is found in the superoscillatory region, then when one looks in the Fourier transform the impossible (e.g. gamma) energy is seen. Therefore, the answer to “Where does the photon take the extra energy from?” is that it obtains the “impossible” energy by exchanging \( E_{mod} \) with the opener. I.e. we show how an exchange of a definite modular variable will create the hi-momentum which lies outside the eigenvalue spectrum. In other words, before we do a phase change, if we look at the Fourier transform there are only momenta \( < p_0 \) while if we introduce a relative phase then what before was just a weak value for the high-momentum, now becomes a strong value.
3.1. Imposing a relative phase can create the high-momentum

Let’s consider the state of the photon, eq. 2.29, as broken up into three components [35], one $|\Psi_{I\text{in so}}\rangle$ inside the super oscillatory region, and the other two $|\Psi_{I\text{out so}}\rangle$ and $|\Psi_{II\text{out so}}\rangle$ outside the super oscillatory region (see fig. 7):

$$|\Psi\rangle = |\Psi_{I\text{out so}}\rangle + |\Psi_{II\text{in so}}\rangle + |\Psi_{III\text{out so}}\rangle$$

No high momenta

Now, we know that $|\Psi\rangle$ does not have the high-momenta. However, if we consider $|\Psi_{II\text{in so}}\rangle$ alone, then the high-momenta is actually there. Therefore, if we subtract 2$|\Psi_{II\text{in so}}\rangle$ from eq. 3.30, then the resultant will have high-momenta:

$$|\Psi\rangle = |\Psi_{I\text{out so}}\rangle - |\Psi_{II\text{in so}}\rangle + |\Psi_{III\text{out so}}\rangle$$

Has high momenta

Consider now a von Neumann measurement of the projection operator in the space region $(0, L)$, $\Pi = \int_{0}^{\infty} |x\rangle\langle x| \, dx$ (if $x \in (0, L)$, then $\Pi = 1$, otherwise $\Pi = 0$):

$$H_{\text{int}} = g(t)p_{q}\Pi$$

$q$ is the location of the pointer of the MD and $p_{q}$ is the conjugate momentum. Let’s consider an initial state of the pointer, $|q = 0\rangle$ and suppose the interaction takes place between $t - \epsilon$ and $t + \epsilon$ with $\int_{t-\epsilon}^{t+\epsilon} g(t) \, dt = 1$. After the interaction, $U = \exp\{ip_{q}\Pi\}$, the state of the photon and MD is (for simplicity, considering just 2 regions):

$$U[|\Psi_{II\text{in so}}\rangle + |\Psi_{I\text{out so}}\rangle]|q = 0\rangle = [\exp\{ip_{q}\Pi\}|\Psi_{II\text{in so}}\rangle + \exp\{ip_{q}\Pi\}|\Psi_{I\text{out so}}\rangle]|q = 0\rangle$$

$$= [\exp\{ip_{q} \cdot 1\}|\Psi_{II\text{in so}}\rangle + \exp\{ip_{q} \cdot 0\}|\Psi_{I\text{out so}}\rangle]|q = 0\rangle$$

$$= \exp\{ip_{q}\}|\Psi_{II\text{in so}}\rangle|q = 0\rangle + |\Psi_{I\text{out so}}\rangle|q = 0\rangle$$

There are now two ways to look at this result:

- We apply $\exp\{ip_{q}\}$ to the position state of MD ($\exp\{ip_{q}L\}$ simply translates $q$ by a the parameter distance $L$):

$$U|\Psi\rangle|q = 0\rangle = |\Psi_{II\text{in so}}\rangle|q = 1\rangle + |\Psi_{I\text{out so}}\rangle|q = 0\rangle$$

In this view, the photon is entangled with the pointer.
• Or, if we only look at the photon:

\[ U|\Psi\rangle_q = 0 = [\exp\{ip_q\}|\Psi_{in\,so}^I\rangle + |\Psi_{out\,so}^f\rangle]q = 0 \]  

(3.35)

In which case, we see that there is a phase \( \exp\{i\alpha\} \) between the two components \( |\Psi_{in\,so}^I\rangle \) and \( |\Psi_{out\,so}^f\rangle \). This phase is uncertain because \( q \) is well defined and therefore \( p_q \) is uncertain. Now, measurement of \( \Pi \) changes the distribution of the momentum of the photon so that now there is a non-zero probability to find the super-oscillatory high momentum. Then it must be the case that even a well-defined relative phase will also cause the high momentum to appear, since an uncertain phase is built out of many fixed, certain phases. Otherwise if no definite relative phase would make the high momentum appear then there would be no way to see the high momentum in the mixture of states with different relative phases. So, one of the phases should bring out the high-momentum by itself.

The change of relative phase does not correspond to any force exerted on the particle (because the wavefunction goes to 0 where the potential is applied). It’s hard to imagine a more delicate influence on the particle aside from this change of relative phase, yet the consequences are very curious. In summary, although the emitted gamma photon has a genuine high-momenta, there is no change to any of the moments of the momenta. I.e. if \( Pr(p) \) represents the probability distribution before the relative phase change, and \( Pr(p, p_q) \) afterwards, then there is no change to any of the moments, i.e. \( \{ Pr(p, p_q) - Pr(p) \} p^a dp = 0 \). However, we will suggest that there is a kind of loophole in the conservation laws, which allows us to exchange a new quantity (called modular variables), without changing the distribution of any of the moments of the energy of the opener, and therefore in a sense, that high-energy was in the photon all along. That is, in order to say that the photon had this high-energy, we need to say that something could make it realized without paying a penalty. So, when a system is pre- and post-selected to manifest superoscillatory phenomenon, it gets into a state such that it is ready to exchange a modular variable which is in fact a non-local interaction. The change of relative phase to create the high-momenta can be understood as a non-local exchange of modular momentum or modular energy \( E_{mod} \).

3.2. Modular Variables: a new kind of nonlocality

In this section, we review the formalism of modular variables [10–12, 33, 35] and their relationship for quantum interference phenomenon. Consider two disjoint wavefunctions which are approaching each other at a velocity \( 2\vec{v} \) and let \( \psi_1(x) = \psi_1(x - L) \) at \( t = 0 \). These wavefunctions define an entire class of wavefunctions given by:

\[ \psi_\alpha = \psi_1 + e^{i\alpha} \psi_2 \]  

(3.36)

While different \( \alpha \)'s yield different wavefunctions, the wavefunction \( e^{-i\alpha}\psi_\alpha \) is equivalent to \( \psi_\alpha \) (since absolute phase cannot be measured). This property, combined with the assumption that \( \psi_1 \) and \( \psi_2 \) do not overlap, means that no local experiment can measure \( \alpha \): i.e., if an apparatus only interacts with \( \psi_2 \), it cannot determine \( \alpha \), unless it is later brought into contact with an apparatus which has interacted with \( \psi_1 \). We could also wait a time \( t > \frac{L}{v} \) for the packets to overlap (where \( L \) is the separation between the packets), thereby producing interference and the possibility of determining \( \alpha \). As a working example, we consider a one dimensional problem, with \( \psi \) of the form \( \exp\{-\frac{x^2}{\Delta x^2}\} \), i.e., \( \psi_1(x, 0) = Ne^{-\frac{(x-x_0)^2}{\Delta x^2}}e^{ik_o x} \) and \( \psi_2(x, 0) = Ne^{-\frac{(x-x_0)^2}{\Delta x^2}}e^{-ik_o x} \) where \( N \) is a normalization factor, \( x_o = \frac{L}{2} \), and \( \Delta x \ll L \). By neglecting the spread of the wavefunction over time, we obtain:

\[ \psi_1(x, T) \approx \psi_1(x - vT, 0) = Ne^{-\frac{(x-x_0-vT)^2}{\Delta x^2}}e^{ik_o(x-vT)} \]  

(3.37)

\[ \psi_2(x, T) \approx \psi_2(x + vT, 0) = Ne^{-\frac{(x+x_0+vT)^2}{\Delta x^2}}e^{ik_o(x+vT)} \]  

(3.38)
where \( v \) is the average speed \( v = \frac{\langle \psi \rangle}{m} \gg \Delta v \). So, at the time of overlap, \( T \), \( (x_\alpha = vT) \)

\[
\psi_\alpha(x, T) = 2Ne^{i(\frac{\omega}{2} - k_\alpha x_\alpha)}e^{-\frac{x^2}{\Delta x^2}}\cos(k_\alpha x - \alpha) \tag{3.39}
\]

and the probability density \( P \) yields interference depending on \( \alpha \):

\[
P_\alpha(x, T) = \psi_\alpha^*(x, T)\psi_\alpha(x, T) = 4N^2 Ne^{-2\frac{x^2}{\Delta x^2}}\cos^2(k_\alpha x - \frac{\alpha}{2}) = 2N^2 Ne^{-2\frac{x^2}{\Delta x^2}}[1 + \cos(2k_\alpha x - \alpha)] \tag{3.40}
\]

This interference pattern in the probability for the two cases of \( \alpha = 0, \pi \) is illustrated in Fig. 8, and we can see the interference is determined by \( \alpha \).

![Figure 8. \( P_\alpha(x, T) \) for (a) \( \alpha = 0 \) (b) \( \alpha = \pi \)](image)

We can now state a basic theorem for interference (refer to Appendix E for its proof):

**Theorem 3:** Let \( \psi_\alpha = \psi_1(x, t) + e^{i\alpha}\psi_2(x, t) \) such that there is no overlap of \( \psi_1(x, 0) \) and \( \psi_2(x, 0) \). If \( n \) is an integer, then for all values of \( t \), and choices of \( \alpha, \beta \):

\[
\int [\psi_\alpha^*(x, t)\psi_\alpha(x, t) - \psi_\beta^*(x, t)\psi_\beta(x, t)]x^n dx = 0 \tag{3.41}
\]

This is a very important result, which characterizes all interference phenomena. 19 In summary, although the simple wave-packets in eq. (3.36) do overlap, we use them because they are easy to handle mathematically and are a good approximation to having a region of zero probability between the packets. Nevertheless, the tails give some initial overlap in which case the left hand side of eq. (3.41) approaches 0 as the weight of the overlap approaches 0. Usually, the physical arrangement of interference is such that there is no overlap tail, i.e., usually in interference, we know the particle went through one slit or the other, with zero probability between. In general, mathematicians would prefer to use analytic functions which have no regions of zero weight (by analytical functions, we mean functions which have no singularities on the complex plane). However, experimentalists have been using non-analytic functions in the laboratory for years.
we have two basic properties associated with the interference of non-overlapping wave packets:

(i) interference depends on the relative phase $\alpha$, and

(ii) all moments of both position and momentum are independent of the relative phase parameter $\alpha$ (see eq. (3.41) and Appendix E).

Thus, information in an interference pattern is stored in the wavefunction in a subtle fashion.

### 3.2.1. The Hidden Information of Future Interference

Examining eq. (3.40), we see that $\langle \cos x \rangle$ will depend upon $\alpha$. In particular,

$$
\langle \cos \{2k_o x(T)\} \rangle_{\alpha=0} = 2N^*N \int e^{-\frac{2\pi^2}{\hbar^2}} \left[1 + \cos(2k_o x)\right] \cos(2k_o x) dx
$$

$$
\langle \cos \{2k_o x(T)\} \rangle_{\alpha=\pi} = 2N^*N \int e^{-\frac{2\pi^2}{\hbar^2}} \left[1 - \cos(2k_o x)\right] \cos(2k_o x) dx
$$

The first term of each of these integrals are equal, while the second are of opposite sign and obviously not zero. Thus, $\langle \cos \{2k_o x(T)\} \rangle$ is clearly dependent upon $\alpha$. We can find the dependence upon $\alpha$ for our example by noting that $\cos \{2k_o x(T)\} = \cos \left\{2k_o x(0) + \frac{\hbar}{2m}(2k_o L)\right\}$ and remembering that $\frac{2k_o L}{m} = v(0)$, and $v(0)T = x_o$. Then dropping the (0) as being understood, and remembering that $k_o$ and $x_o$ are constants, we obtain: $\cos \{2k_o x(T)\} = \cos \{2k_o x + \frac{2\pi p}{\hbar}\}$. Now we need the following theorem:

**Theorem 4:** Let $p$ be an $x$ component of linear momentum. If $A$ is an operator such that $A = A(x)$ and $A = \frac{\partial A}{\partial x}$, then:

$$
e^{i\frac{\hbar}{\pi}L} = e^{-\frac{i\hbar}{\pi}A} e^{\frac{i\hbar}{\pi}pL} e^{\frac{i\hbar}{\pi}A}.
$$

This result can be used with the fact that $e^{i\frac{\hbar}{\pi}pL} \psi(x) = \psi(x + L)$ to show that:

$$
e^{i\frac{\hbar}{\pi}(p+A)L} = \exp \left\{ \frac{i}{\hbar} \int x' A(x') dx' \right\} e^{i\frac{\hbar}{\pi}pL} \tag{3.42}
$$

Applying this identity to the representation of $\cos(2k_o x + \frac{2\pi p}{\hbar})$ as exponentials yields

$$
\langle \cos \{2k_o x(T)\} \rangle = \frac{1}{2} \cos(\alpha) \tag{3.43}
$$

From this example, we have learned that:

(i) although the exact value of eq. (3.43) depends upon the wavefunction assumed, the $\alpha$ dependence holds even when there is no overlap;

(ii) operators that are sensitive to $\alpha$ are operators which involve exponentials of the momentum.

The reason is that these exponentials translate wavefunctions so they overlap, thus manifesting the interference;

(iii) since

$$
\cos \{2k_o x(T)\} = 1 - \frac{(2k_o x)^2}{2!} + \frac{(2k_o x)^4}{4!} - \ldots \tag{3.44}
$$

it might be argued that $\langle \cos \{2k_o x(T)\} \rangle = 1 - \frac{(2k_o)^2(x^2(T))}{2!} + \frac{(2k_o)^4(x^4(T))}{4!} - \ldots$ But, by Theorem 3, each of the moments are independent of $\alpha$, so term by term the result must also

---

20 This can be easily generalized to 3 dimensions, with $\vec{p}$ and $\vec{A}(\vec{r}) = \vec{V}(\vec{r})$.

21 This is proven by expanding the exponentials in a power series, and establishing the identity term by term.
be independent of $\alpha$. We can’t use the argument that we have an overlap effect because any such effect should be proportional to the initial weight of the overlap, which is not the case for eq. (3.43). In fact, it is the lack of analyticity which prevents us from making the expansion given by eq. (3.44). In general $e^{ixL}\psi(x) = \sum_n (ipL)^n \psi(x)$ only for $\psi(x)$ which are analytic in the region $x < L$. However, $e^{ixL}\psi(x) = \psi(x + L)$ holds as long as $\psi(x)$ has a Fourier transform which thus accommodates a much broader class of functions:

$$e^{ipL}\psi(x) = \frac{1}{\sqrt{2\pi\hbar}} \int \phi(p)e^{ip(x+L)}dp = \frac{1}{\sqrt{2\pi\hbar}} \int \phi(p)e^{ip(x)}dp = \psi(x + L)$$

### 3.2.2. Modular Momentum

The above discussion has shown that dynamical variables like $\sin pL$ and $\cos pL$ are relevant for the description of interference phenomena in the quantum domain. In this section we present further indications of the significance of these variables for the understanding of the general properties of quantum interference.

It is well known that quantum system behaves similarly to a classical system if the conditions are such that the quantum system can be represented by a single wave packet moving in a smoothly varying potential. The variables that are adequate for describing the motion of a single wave packet are the average position and the average momentum, and the uncertainties of position and momentum. The equations of motion for the averages are similar to the classical equations of motion in the given potential, and the uncertainties $\Delta x$ and $\Delta p$, describing the spread in the relevant variables, have properties similar to those of the spread of variables in a classical situation with unsharply defined initial conditions.

However, if we consider a quantum problem in which the object is represented by more than one wave packet, then the above variables cease to be sufficient for describing the problem. The following argument provides a simple indication of the inadequacy of the usual variables in situations involving several non-overlapping wave packets. The usual view of quantum uncertainties, derived from the study of single wave packet problems, makes us expect that if we increase the uncertainty in position without introducing a correlation between position and momentum, then the uncertainty in momentum will decrease. This expectation is wrong in the case of non-overlapping wave packets. In fact, when we suppose any number, even infinitely many, of such wave packets, so that $\Delta x$ becomes arbitrarily large, the uncertainty in momentum, $\Delta p$, remains the same as in the case of a single wave packet. To see this, consider a simple example. Let $\psi_o(x)$ be a wave packet centered at $x = 0$. We assume that $\psi_o(x)$ is exactly equal to zero for $x \in [-\frac{L_o}{2}, \frac{L_o}{2}]$. For this wave function we have

$$\Delta p \approx \frac{\hbar}{\Delta x}.$$  

We now define

$$\psi_n(x) = \psi_o(x + nL)$$  

(3.45)

Where $n = 1, 2, \ldots$ and $L > L_o$. In Appendix C, we show that for the class of states

$$\Psi(x) = \sum_n a_n \psi_n(x)$$  

(3.46)

with $\sum_n |a_n|^2 = 1$, the uncertainty in momentum is independent of $a_n$.

From this surprising result we learn, for example, that if in a typical quantum interference experiment we want to measure through which slit the particle passes, then we do not have to increase the uncertainty in momentum. In fact, what our intended experiment does is to change the situation from one in which $a_1 = a_2 = \frac{1}{\sqrt{2}}$ to one in which, e.g., $a_1 = 1, a_2 = 0$. In both situations $\Delta p$ is the same. Therefore, it is misleading to say that in measuring which slit the particle passes through we destroy the interference pattern by increasing the uncertainty in momentum.

On the other hand, the variables that are sensitive to the choice of $a_n$ are, e.g.:

$$\langle \cos \frac{pL}{\hbar} \rangle = \frac{1}{2} \left\{ \langle e^{i\frac{pL}{\hbar}} \rangle + \langle e^{-i\frac{pL}{\hbar}} \rangle \right\}$$  

(3.47)
and

\[ \langle e^{i\frac{pL}{\hbar}} \rangle = \int_{-\infty}^{+\infty} dx \sum_{m} a_m^* \psi_m^* e^{i\frac{mx}{\hbar}} \sum_{n} a_n^* \psi_n^* = \int_{-\infty}^{+\infty} dx \sum_{m} a_m^* \psi_m^* \sum_{n} a_n^* \psi_{n+1}^* = \sum_{n} a_{n+1}^* a_n \]

(3.48)

(Similarly for \( \langle e^{-i\frac{pL}{\hbar}} \rangle \)) which obviously depends on the choice of \( a_n \). In particular, if all \( a_n \) are equal, \( \langle \cos \frac{pL}{\hbar} \rangle = 1 \), while if only one \( a_n \) is different from zero, \( \langle \cos \frac{pL}{\hbar} \rangle = 0 \). Thus, a measurement to find which slit the particle passes through in an interferometer will change the value of \( \langle \cos \frac{pL}{\hbar} \rangle \). This result is a further suggestion that the ordinary dynamical variables, \( \langle p \rangle \) and \( \Delta p \), are inappropriate for the description of quantum interference phenomena, and that such phenomena should instead be described in terms of variables like \( \langle \cos \frac{pL}{\hbar} \rangle \).

By way of example, we apply this to the double-slit experiment [11] where the interference pattern is built up one particle at a time. Consider first two non-overlapping wavefunctions, representing what happens to a particle immediately after it passes through the two slits and still being in a superposition of two possibilities, i.e. one slit or the other:

\[ \Psi(x, y, t) = \Psi_1(x, y, t) + e^{i\alpha} \Psi_2(x, y, t) \]

(3.49)

with \( \Psi_2(x, y, t) = \Psi_1(x - L, y, t) \) where \( L \) is the distance between the slits (i.e. \( \Psi_1 \) represents a wave-packet emerging from the left slit and \( \Psi_2 \) represents a wave-packet emerging from the right slit). The uncertainty relation implies that if \( \Delta x \) is decreased, then \( \Delta p_x \) can be increased such that \( \Delta x \Delta p_x \geq \hbar / 2 \). If we perform an experiment which gives us the location of the particle, then we destroy the interference pattern. If we try to deduce this from the uncertainty principle, then we might argue that since the measurement decreases the uncertainty in position, it increases the momentum uncertainty. However, if we calculate \( \Delta p^2 = \langle p^2 \rangle - \langle p \rangle^2 \) (using, e.g. \( \Psi = \frac{1}{\sqrt{2}} (\Psi_1 + \Psi_2) \)) and \( \Psi = \Psi_1 \) as the two wavefunctions), we find that \( \Delta p \) is unchanged (see Appendix D). Here we started out with the superposition of two acceptable wave packets, so \( \Delta x \Delta p \geq \hbar / 2 \) before the measurement, and hence \( \Delta p \) can remain constant. In fact, since there is, by assumption, no overlap, then by Theorem 3, every power of the momentum remains unchanged. If we were to ascertain through which slit an individual particle traveled, then the uncertainty in momentum would not be changed: therefore, destruction of the interference pattern cannot be explained by a change in the uncertainty of the momentum. This shows that the usual uncertainty principle has no connection with interference phenomenon.

Why then, does the interference pattern disappear when the particle is localized? Although the moments of the momentum do not contain the interference information, there are periodic functions which do contain this information. In particular, the action of detecting the particle at a particular slit makes the modular momentum (see definition below) completely uncertain.

First we define some terminology. Consider a particle moving on a circle, and describe its position by \( \theta \), and its probability distribution by \( Prob(\theta) = \psi^*(\theta)\psi(\theta) \). Then \( \theta \) is completely uncertain when \( Prob(\theta) = \frac{1}{2\pi} \). If we consider operators such as \( e^{i\frac{pL}{\hbar}} \) (with \( p_o = \frac{\hbar}{2} \) so that \( e^{i\frac{pL}{\hbar}} = e^{i\frac{\theta}{\tau}} \)), then, if we only know \( p \) such that \( e^{i\frac{2\pi p}{p_o}} = e^{i\alpha} \) with \( \alpha \) definite, we do not know \( p \), but only \( p \mod p_o \). Here the symbol \( \mod \) means modulo, i.e. if \( n \) is the largest integer such that \( np_o < p \), then \( p \mod p_o = p - np_o \). It is clear that \( p \mod p_o \) has the topology of a circle, as would any periodic function, every point on the circle is another possible value for the modular variable. This leads us to the complete uncertainty principle:

**Complete Uncertainty Principle**: If \( \theta \) is a periodic function with period \( \tau \), then \( \theta \) is completely uncertain \( (Prob(\theta) = constant = \frac{1}{\tau}) \) if and only if \( \langle e^{in\theta} \rangle = 0 \) for any integer \( n \neq 0 \) (Appendix F).
An angle like $\theta$ can be defined for any particle whose wavefunction is nonzero over a finite range. Complete uncertainty means that every value of the modular variable is equally probable which means that we would be unable to see any change in the modular variable. As an important example, consider a particle in a box of length $L_o$ such that $\psi(x) = 0$ for $|x| > \frac{L_o}{2}$. Then \( \int_{-\infty}^{+\infty} \psi^*(x)\psi(x + nL)dx = 0 \) for $n \neq 0$ and $L > L_o$. But, $\psi(x + nL) = e^{i\frac{2\pi n}{L_o}}\psi(x)$, so \( \langle e^{i\frac{2\pi n}{L_o}} \rangle = 0 \), $n \neq 0$, $L > L_o$. If $m$ is any integer, we have \( \exp\{i2\pi n\frac{p}{p_o}\} = \exp\{i2\pi n\frac{p_{mod}}{p_o}\} \) By defining the Modular Momentum $\Theta(p) = 2\pi \frac{p_{mod}}{p_o}$ (noting that $\Theta$ is periodic) we find by the complete uncertainty principle, that $\Theta$ is completely uncertain.

The final point in our review is to emphasize that modular variables have different kinds of conservation laws. For example, suppose we have a collision between two systems and then momentum is conserved: $p_1 + p_2 = p_1' + p_2'$, where the prime represents the momentum after the collision. What is the analog of this in terms of modular variables? Using $\pi_1(p_1) = \cos(2\pi p_1/p_o)$ and $\pi_2(p_2) = \cos(2\pi p_2/p_o)$ (which is another expression for $p_{mod}$) we see that using $p_1 + p_2 = p_1' + p_2'$, conservation of modular momentum assumes the form:

\[
\cos[2\pi \{p_1 + p_2\}/p_o] = \cos[2\pi \{p_1' + p_2'\}/p_o]
\]

in other words: $\pi_1\pi_2 - \sqrt{1 - \pi_1^2}\sqrt{1 - \pi_2^2} = \pi_1'\pi_2' - \sqrt{1 - (\pi_1')^2}\sqrt{1 - (\pi_2')^2}$ which gives:

\[
(\pi_1')^2 + (\pi_2')^2 - 2\cos[2\pi \{p_1 + p_2\}/p_o]\pi_1'\pi_2' = 1 - \cos[2\pi \{p_1 + p_2\}/p_o]^2
\]

Instead of a line $p_1 + p_2 = constant$, the conservation law for modular variables is an ellipse.

### 3.2.3. The dynamical difference between Quantum and Classical Mechanics

Let us compare the classical and quantum equations of motion of variables like $\frac{p}{p_o}$ (we set $\frac{h}{L} = p_o$). Classically, we have

\[
\frac{d}{dt} \sin \frac{p}{p_o} = \frac{dp}{dt} \cos \frac{p}{p_o} = -\frac{\partial V}{\partial x} \cos \frac{p}{p_o}
\]

Thus, if the force vanishes at the position of the particle, $\sin \frac{p}{p_o}$ is a constant of the motion for every value of $p_o$. To obtain the QM equations of motion, we write

\[
\sin \frac{p}{p_o} = \frac{1}{2i} \left\{ e^{i\frac{p}{p_o}} - e^{-i\frac{p}{p_o}} \right\}
\]

and with $H = \frac{p^2}{2m} + V(x)$, we have:

\[
\frac{d}{dt} e^{i\frac{p}{p_o}} = \frac{i}{\hbar} [H, e^{i\frac{p}{p_o}}] = \frac{i}{\hbar} \frac{p^2}{2m} + V(x), e^{i\frac{p}{p_o}} = \frac{i}{\hbar} [V(x), e^{i\frac{p}{p_o}}]
\]

\[
= \frac{i}{\hbar} \left\{ V(x)e^{i\frac{p}{p_o}} - e^{-i\frac{p}{p_o}}V(x) \right\} = \frac{i}{\hbar} \left\{ V(x) - V(x + L) \right\} e^{i\frac{p}{p_o}}
\]

Similarly:

\[
\frac{d}{dt} e^{-i\frac{p}{p_o}} = \frac{i}{\hbar} \left\{ V(x) - V(x - L) \right\} e^{i\frac{p}{p_o}}
\]

\footnote{Modular variables such as this, can be defined for all bounded wavefunctions. It is important to notice, that uncertainty in the modular momentum $\Theta(p)$ implies an uncertainty in the regular momentum $\Delta p$ (see Appendix C). This can be seen if, e.g., the modular momentum is completely uncertain, then we know that the particle's wavefunction is bounded, $\Delta x < L_o$, and hence $\Delta p > \frac{\hbar}{L_o}$. On the other hand, if we have a great uncertainty in the regular momentum $\Delta p \gg 1$, it does not imply that the modular momentum is uncertain at all.}
Therefore:

\[
\frac{d}{dt} \sin \frac{p}{p_0} = \frac{1}{2i} \left\{ \frac{d}{dt} e^{i \frac{p}{p_0}} - e^{-i \frac{p}{p_0}} \right\} \\
= \frac{1}{2i} i \hbar \left\{ V(x) e^{i \frac{p}{p_0}} - V(x + L) e^{i \frac{p}{p_0}} - V(x) e^{-i \frac{p}{p_0}} + V(x - L) e^{-i \frac{p}{p_0}} \right\} \\
= \frac{1}{2\hbar} \left\{ V(x - L) - V(x + L) \right\} \left\{ \frac{e^{i \frac{p}{p_0}} + e^{-i \frac{p}{p_0}}}{2} \right\} \\
+ i \left\{ 2V(x) - V(x - L) - V(x + L) \right\} \left\{ \frac{e^{i \frac{p}{p_0}} - e^{-i \frac{p}{p_0}}}{2} \right\} \\
= \frac{1}{2\hbar} \left\{ V(x - L) - V(x + L) \right\} \cos \frac{p}{p_0} \\
+ i \left\{ 2V(x) - V(x - L) - V(x + L) \right\} \sin \frac{p}{p_0} \tag{3.56}
\]

We see that for variables like \(\sin \frac{p}{p_0}\) there is a remarkable difference between the classical and the quantum mechanical equations of motion. While the quantum equation of motion for the average of the momentum itself is identical to the classical equation of motion for \(p\), there is no such correspondence in the case of \(\sin \frac{p}{p_0}\). In contrast to the classical equation of motion for \(\sin \frac{p}{p_0}\), the quantum equation of motion is non-local. As a consequence, the classical variable \(\sin \frac{p}{p_0}\), may be a constant of the motion under conditions when the average of the quantum variable \(\sin \frac{p}{p_0}\) is not conserved.

In particular, quantum interference involves a non-local exchange of \(\sin \frac{p}{p_0}\) between the particle and the screen with the slits. The quantum equation of motion indicates where the loophole lies in the usual approach to the problem of quantum interference. It seemed natural to believe that if we have a point particle in a local potential, then it is impossible to understand how both slits of the interferometer can affect the particle simultaneously. We now see that this is not the case since variables like \(\sin \frac{p}{p_0}\) are affected by a local potential which is different from zero at a distance \(\sin \frac{p}{p_0}\) from the position of the particle. Thus, even when the particle passes through one slit, the question of whether the other slit a distance away is open or closed is relevant for the behavior of the variable \(\sin \frac{p}{p_0}\).

More abstractly for any \(\hat{A}\), the QM equation of motion is

\[
i\hbar \frac{d\hat{A}}{dt} = [\hat{A}, \hat{H}] \tag{3.57}
\]

while the CM equation for such an operator would be

\[
\frac{dA}{dt} = \{A, H\}_{P.B.} = \sum_i \left( \frac{\partial A}{\partial x_i} \frac{\partial H}{\partial p_i} - \frac{\partial A}{\partial p_i} \frac{\partial H}{\partial x_i} \right) \tag{3.58}
\]

(where \(x_i\) and \(p_i\) are the coordinates and associated momenta of the system and the sum is taken over all coordinates.) Now, there are two types of time dependence that the operator \(\hat{A}\) may have: implicit time dependence, such as \(\hat{X}(t)\), and explicit time dependence, such as \(\hat{X} \cos(t) + \hat{P}_x \sin(t)\). Assuming for now that there is no explicit time dependence, then using the Schrodinger equation it is easy to show that

\[
\frac{d}{dt} \langle \hat{A} \rangle = \frac{i}{\hbar} \langle [\hat{H}, \hat{A}] \rangle \tag{3.59}
\]
which is Ehrenfest’s Theorem for operators without explicit time dependence. Thus if
\[
\frac{-i}{\hbar} \langle [\hat{A}, \hat{H}] \rangle = [A, H]_{P.B.}.
\] (3.60)
then the time average dependence of QM and CM are the same (e.g. \([X, P_x]_{P.B.} = 1\) and \([\hat{X}, \hat{P}_x] = i\hbar\)). This has led many physicists to assert that when 3.60 holds, then the classical and the quantum dynamics are the same and the only difference between QM and CM lies in the kinematics (uncertainty relations) so that an understanding of the classical dynamics gives a complete understanding of the quantum dynamics. One of the objectives of this paper is to show that this viewpoint is incorrect: there is a fundamental difference between classical and quantum dynamics which is necessary to obtain an insight into the meaning of the uncertainty relations. Although historically the uncertainty relations (kinematics) came first, logically it is the dynamics that we must understand before the uncertainty relations.

Again, if we consider \(H = \frac{p^2}{2m} + V(x)\), such that \(\frac{dV}{dx} = 0\). Then since \(\frac{\partial f}{\partial x} = \frac{\partial H}{\partial x} = 0\), we have:
\[
\frac{df(p)}{dt} = \{f(p), H\}_{P.B.} = -\frac{\partial f}{\partial p} \frac{\partial H}{\partial x} + \frac{\partial f}{\partial x} \frac{\partial H}{\partial p} = 0
\] (3.61)
Therefore, classically, the time derivative of any function of momentum is different from zero only if there is a change in the potential in the same physical location where the particles are located. It is here where the argument that there is no essential difference between the quantum dynamics of eq. (3.57) and the local classical dynamics of eq. (3.58) breaks down.

Dirac first argued that the Poisson bracket could be simply replaced by the commutator by assuming that corrections were small (in the higher order terms in \(A\) or \(H\)). However, we shall learn by focusing on DSO that the quantum eqns. 3.57 are local only if \(A\) and \(H\) are bilinear in \(x\) and \(p\) or if \(A\) is linear in \(x\) and \(p\) and \(H\) is arbitrary, i.e. eq. (3.58) is local if
\[
A = a_0 + a_1 x + a_2 p + a_3 x^2 + a_4 p^2 + a_5 x p,
\]
However, for infinite series such as \(e^{\frac{iiV}{\hbar}}\), then the EOM are intrinsically non-local [10–12]:
\[
\frac{d}{dt} e^{\frac{iV}{\hbar}} = \frac{i}{\hbar} \{V(x) e^{\frac{iV}{\hbar}} - e^{\frac{iV}{\hbar}} V(x)\} = \frac{i}{\hbar} [V(x) - V(x + L)] e^{\frac{iV}{\hbar}}
\] (3.62)
Consider a potential \(V(x)\), such that \(V(x + L) = V(x) + V_o\). Suppose we have a single particle characterized by a wavefunction that is in a superposition of being isolated in region \(0, L\) and \(2L, 3L\) (etc). Initially we take \(\exp \frac{i}{\hbar} \{pL\} = 1\) and thus, initially, \(p(t)mod \frac{L}{\hbar} = 0\). Subsequently, \(\frac{d}{dt} e^{\frac{iV}{\hbar}} = V_o e^{\frac{iV}{\hbar}}\) leading to \(e^{\frac{iV}{\hbar}}(t) = \exp \frac{i}{\hbar} \{pL - V_o t\}\) and thus \(p(t)mod \frac{L}{\hbar} = V_o t \frac{L}{\hbar}\). Therefore, irrespective of the system state and even though the potential vanishes and no forces are acting in the region \(0, L\), the modular momentum is changing at a constant rate based on the particles’ nonlocal interaction with the potential in other regions. We showed that these functions will depend on exponentials of the form \(e^{\frac{iV}{\hbar}}\) and therefore obey non-local EOM. Classically, however, the equation of motion for \(e^{\frac{iV}{p_0}}\) gives \((p_o = h/L, p_0\) constant, \(h \rightarrow 0)\):
\[
\frac{d}{dt} e^{\frac{iV}{p_0}} = -\frac{2\pi}{p_0} \frac{dV}{dx} e^{\frac{iV}{p_0}}
\] (3.63)
Eq. (3.63) is local, but eq. (3.62) is nonlocal because the operator \(e^{\frac{iV}{\hbar}}\) can change its value even at a location where \(\frac{dV}{dx} = 0\).\(^{23}\)
We now consider an example of the exchange of modular variables in order to distinguish it from the dramatic case illustrated in §3.1.

\(^{23}\) There was an attempt to give a counter-example to the non-local phenomenon (e.g. the Aharonov-Bohm effect)
3.3. Example: Particle in a Piston

Consider a particle in a long cylinder of length $L$ with a frictionless and movable piston at the end [12, 33, 35]. When an external quantum particle hits the piston, the piston barely moves because it is much heavier than the particle and therefore we can neglect its motion. The particle inside the cylinder has a Gaussian wavefunction of width $\Delta x \ll L$ that falls off sufficiently quickly, so that when the center of the internal particle is far from the piston, we can neglect any probability that the internal particle can be found at the piston. If $\langle v \rangle$ is large, such that, $m \langle v \rangle \gg \hbar / \Delta x$, then the internal particle will make many reflections before it spreads an appreciable amount. The period of the internal particle $T = (2L / \langle v \rangle)$, is well defined and therefore the energy of the particle will be ill-defined, because $\Delta ET \geq \hbar$. The modular energy $E \mod \hbar$ is however definite because $e^{iET/\hbar}$ is the time displacement operator and will evolve the internal particle back to its original shape:

$$e^{iET/\hbar} \Psi(x, t = 0) = \Psi(x, t = T) \approx \Psi(x, t = 0) \quad (3.64)$$

therefore, $E = n\hbar$ and thus $E \mod \hbar = 0$. The expectation value, computed at $t = 0$ is:

$$\langle \Psi(0)|e^{iET/\hbar}|\Psi(0) \rangle = 1 \quad (3.65)$$

Now consider an external particle which makes a completely elastic collision (when the internal particle is far away from the piston) between 2 plates that are attached to the piston which moves the piston inwards a distance $\delta L \ll \Delta x$. If the internal particle is fast, such that it makes many reflections from the piston in the time that the external particle bounces off the two plates, then the internal particle’s energy will increase between the 2 collisions (at the expense of the external particle), because the energy of a particle in a box is $E = \frac{n^2 \pi^2 \hbar^2}{2mL^2}$ and $L$ is decreasing. Consider now the situation in which the internal particle is very slow and does not reflect from the piston before the external particle bounces off the 2nd plate. In this case, the length of the piston is $L - \delta L$ and $|\Psi(T)\rangle$ will be displaced a distance $2\delta L$ from $|\Psi(0)\rangle$, i.e.

$$|\Psi(T)\rangle = e^{i2\delta L/\hbar}|\Psi(0)\rangle \quad (3.66)$$

For purpose of illustration, we assume that $\psi(x) \approx e^{-(x-a)^2}$ and hence that $\psi$ has a Fourier transform $\hat{\psi}(p) \approx e^{-(p-a)^2}$. Then

$$e^{-2i\delta L p} \hat{\psi} = e^{-i(\frac{2\delta L p}{\hbar} + 2 \frac{\Delta x L}{\hbar})} \hat{\psi}$$

described by modular variables by considering a formulation of QM in terms of the density $\rho(x, p) = \psi^\ast(x, p)\psi(x, p)$ and the current $\vec{j} = \psi^\ast \nabla \psi - \psi \nabla \psi^\ast$. The EOM for $\rho$ and $\vec{j}$ are non-linear, only involve the fields (rather than potentials and are thus gauge invariant) and are local. Consider again the superposed states eq. (3.49) with relative phase $\alpha$ between the two wavepackets. Because $\vec{j} = \rho \nabla \alpha$, the relative phase can be found by integrating $\int \vec{j}$ over all space. However, for wavefunctions that go to zero in between the wavepackets, this approach does not work. Nevertheless, this approach does reveal an interesting connection between non-locality and chaos. For example, there is an instability characterized by infinitesimal changes in $\rho$ and/or $\vec{j}$ in the region between the wavepackets which will result in a large change later when the wavepackets overlap.

Figure 9. Particle in a piston
The last part of the exponential can be neglected if

\[ 1 \gg \frac{\Delta L}{\Delta x} = \frac{\hbar \Delta L}{\hbar \Delta x} \geq \frac{\Delta p \delta L}{\hbar}. \]

Thus the modular energy remains definite if \( \Delta x \gg \delta L \). Alternatively, we can write

\[ \psi(x + 2\delta L) = \frac{1}{\sqrt{2\pi\hbar}} \int e^{-\left(\frac{p-p_0}{\Delta p}\right)^2} e^{i\frac{p_0 x}{\hbar}} e^{i\frac{2p\delta L}{\hbar}} dp \]

and complete the square. We again find that if \( \Delta x \gg \delta L \) we get

\[ \psi(x + 2\delta L) = e^{2i\frac{p_0 \delta L}{\hbar}} \psi(x) \]

so

\[ e^{-i\frac{HT}{\hbar}} \psi_{\text{new}} = e^{2i\frac{p_0 \delta L}{\hbar}} \psi_{\text{new}} \]

and so we get

\[ E_{\text{mod}} \frac{\hbar}{T} = \frac{2p_0 \delta L}{T} \quad (3.67) \]

In other words, because \( \Delta x \gg \delta L \), the scalar product still produces considerable overlap, \( |\langle \Psi(T) | \Psi(0) \rangle| \sim 1 \). However, there has been a change in the expectation value of the modular energy \( E_{\text{mod}} \frac{\hbar}{T} = 2p_0 \delta L/T \). By conservation of modular energy, the modular energy of the external particle has also changed by \( 2p_0 \delta L/T \) as long as \( \delta L \neq \frac{n\hbar}{2p} \) (which would produce no change in \( E_{\text{mod}} \)). If we shift the internal particle by one wavelength, then there will not be an exchange of \( E_{\text{mod}} \). If we shift by less than one wavelength, then we can actually increase the energy of the internal particle, exactly opposite to what one would expect classically.

Before the piston is moved, \( E = E_{\text{mod}} + N_E \hbar/T \) with \( E_{\text{mod}} = 0 \) and \( N_E \) is uncertain since \( E_{\text{mod}} \) is definite and the particle is well localized. Therefore, the probability distribution for the energy is given by fig. 10.a. After the piston is moved, \( E_{\text{mod}} = \frac{\hbar}{2T} \) and the probability distribution is represented by 10.b. Now \( \Delta E \approx \frac{N_E \hbar}{2T} \) in both cases since \( \Delta N_E \gg 1 \), because the particle is localized to \( L \gg \Delta x \).

The important point is that the external particle has changed the internal particle’s energy even though the 2 particles could have been outside each other’s light cones. To analyze this, we first need to see how the conservation of modular energy works. We again consider two systems denoted by 1 and 2, each isolated before and after they interact so that \( E_1^i + E_2^i = \text{const} = E_1^f + E_2^f \), where \( E^i \) is the energy before the interaction and \( E^f \) is the energy after the interaction. If we were to just consider the energy, a plot of all possible cases for a given constant energy is a straight line. For modular variables such as \( \cos(\frac{pL}{\hbar}) \), since \( p_1 + p_2 = \text{const} \), we have \( \cos(\frac{p_1 + p_2 L}{\hbar}) = C \), which is an ellipse determined by the value of \( C \). Modular energy is very similar in structure to modular momentum.

On the flip side, by measuring the change in \( E_{\text{mod}} \) of the external particle, it seems that we could determine whether or not there is a particle inside the piston because the \( E_{\text{mod}} \) of the internal and external particle changes. However, in order for a violation of causality to have occurred, we need to know that the external particle hits the piston before the internal particle hits it and therefore we must know the time of the collision to within \( T \). But this means that \( E_{\text{mod}} \frac{\hbar}{T} \) is completely uncertain for the particle which moves the piston. So any change in the modulo energy is unobservable, since the initial internal \( E_{\text{mod}} \) is completely uncertain. Thus, there is no problem with causality since no observable change occur.
4. Conservation Laws

In the particle in a piston example of §3.3, the exchange of $E_{\text{mod}}$ only slightly shifted the fringes in the probability distribution (compare fig. 10.a with 10.b). However, in the super-oscillation example analyzed in §3.1, a new peak appeared in momentum, far outside the “allowed” values, corresponding to the superscillatory high-momentum (compare fig. 11.b with fig. 11.a). This extra energy was obtained from the particle which opened the window by a non-local exchange of modular energy (or, equivalently, modular momentum). However, if we could determine the change in the modular momentum of the particle which opened the window, then it would be an acausal interaction [35]. In this section, we prove that the change in the modular momentum is completely uncertain which therefore protects causality.

4.1. Localization Creates Uncertainty in Modular Variables

We therefore explore the non-local aspect of modular variables in greater depth and prove in general that if a system is localized to within a region $L$, then all modular variables for $L' > L$ become completely uncertain [35]. The implication of this is that if there is an interaction in which momentum is conserved and modular momentum is exchanged, then this will show up as a change in the distribution of one of the modular momentum, not both. In summary we will show that if you want to change the distribution of one variable without changing the distribution of another, then this can be done only if there is a whole range of modular variables that are exchanged.

Consider a variable $\Theta$ that behaves like an angle. This $\Theta$ is completely uncertain if the probability distribution for $\Theta$ is a constant. We can expand this probability distribution as a
Figure 11. (a) Energy distribution before and (b) Energy distribution after imposing relative phase change between superoscillatory region and exponential region of eq. 2.29

Fourier series because it is periodic in $\Theta$:

$$P_r(\Theta) = \sum a_n e^{i n \Theta}$$

where the coefficients are given by:

$$a_n = \int_0^{2\pi} P_r(\Theta) e^{i n \Theta} d\Theta$$

However, stating that the probability is a constant is equivalent to saying that $a_n = 0$ for $n \neq 0$, because this integral for $a_n$ is also the average of $e^{i n \Theta}$:

$$\int_0^{2\pi} P_r(\Theta) e^{i n \Theta} d\Theta = e^{i n \Theta}$$

So, instead of saying that $\Theta$ is uncertain, we can say that the average of $\langle e^{i n \Theta} \rangle = 0$. This simply says that the average of an oscillating function is zero and this is a necessary and sufficient condition for $P_r(\Theta)$ to be a constant.

Suppose we have a particle that we know is localized within a region $L$. As long as $L' > L$ we can deduce that $e^{-i \frac{p}{\hbar} L'} = 0$ because $e^{-i \frac{p}{\hbar} L'}$ translates this wavefunction out of the region $L$, so the scalar product of this translated wavefunction and the original wavefunction is zero. Now, $p$ can be expressed as $p = p \mod \frac{\hbar}{L} + N \frac{\hbar}{L}$. When this is put into the exponent, the second term just gives zero, therefore $e^{-i \frac{p}{\hbar} L'}$ depends on the modular momentum, not on any integral part of $p$.

$$e^{-i \frac{p \mod \frac{h}{L}}{\hbar} \frac{n L}{L}} = 0$$

for $n \neq 0$. The range of $p \mod \frac{\hbar}{L}$ is $(0, \frac{\hbar}{L})$. If we normalize by dividing $p \mod \frac{\hbar}{L}$ by $\frac{\hbar}{L}$ then this also behaves like an operator with a range of $(0, 2\pi)$, i.e. an angle. Similarly, if $p \mod$ is
completely uncertain, then the values of \( P_x \) which are from 0 to 2\( \pi \) are all equally probable and therefore the expectation value \( \langle e^{i \frac{p_{\text{mod}} L}{\hbar}} \rangle = 0 \). On the flip side, when this expectation value vanishes, then \( p_{\text{mod}} \) is completely uncertain. So if \( \psi(x) \) is confined to a region of length \( L \), then

\[
\langle e^{i \frac{p_{\text{mod}} L}{\hbar}} \rangle = \langle \psi(x) | e^{i p_{\text{mod}} L/\hbar} \psi(x) \rangle = 0
\]

and this means that \( p_{\text{mod}} \) is completely uncertain. So, if the particle is localized to within \( L \), then all modular variables for \( L' > L \) are completely uncertain. This works in the same way as the angular momentum, angle commutation relations: e.g., if I know the total angular momentum, then the angle is completely uncertain. As \( L \) is decreased, i.e. as an object becomes more localized, then more modular momenta become uncertain: i.e. the old modular momenta that were uncertain before \( L \) was decreased remain uncertain and new modular momenta becomes uncertain when \( L \) is decreased. When these new modular momenta become completely uncertain, this means that they can operate nonlocally. We therefore have situations in which non-local equations of motion are at play but in which causality is preserved since we have complete uncertainty in the non-local quantities.

### 4.2. Localization in Time creates uncertainty in Modular Energy

We can now apply the general results of the previous section to superoscillations [12, 35]. Let \( P^1_r(E_1) \) be the energy distribution of the particle described by eq. 1.2, and \( P^2_r(E_2) \) the energy distribution of a second particle which opens the window the localize the particle to \( (0, L) \). The particle needs to open and close the window sufficiently quickly so that the exponential part of the wavefunction for the first particle does not have time to reach the window. Because the timing needs to be relatively definite, the energy of the opener has to be comparatively uncertain. We will use the fact that there are tails in the momentum of MD that was used to project the particle onto the superoscillatory region because the MD needed to know the position of the superoscillation. These tails in the wavefunction of the particle which opens the window and are used to “manifest” the high-momentum. However, as we will now show, although the tails are used, they are used in a way that the probability distribution of the opener does not change, thereby preserving causality.

As we have argued in §4.1, the particles undergo an interaction in such a way that \( P^1_r(E_1) \) changes but \( P^2_r(E_2) \) does not change (otherwise causality would be violated). Furthermore, since energy is conserved, the distribution of the combined energy given by the convolution of the two does not change:

\[
P^{12}_r(E_1 + E_2 = E_{\text{tot}}) = \int P^1_r(E - E') P^2_r(E') dE'
\]  

(4.72)

How is it consistent that \( P^{12}_r \) and \( P^2_r \) do not change yet \( P^1_r \) does change? If we look at the Fourier transform of each of these functions we may obtain an answer:

\[
\int P_r(E) e^{i E t / \hbar} dE = \langle \Psi | e^{i H t / \hbar} | \Psi \rangle = F(t)
\]  

(4.73)

What is the meaning of \( F(t) \)? Consider a wavefunction at time \( t = 0 \), \( |\Psi(0)\rangle = \int |\Psi(E)\rangle |E\rangle dE \) and at time \( t = T \), \( |\Psi(T)\rangle = \int |\Psi(E)\rangle |E\rangle e^{i E t / \hbar} dE \). Looking at the overlap between \( |\Psi(0)\rangle \) and \( |\Psi(T)\rangle \) will tell us to what extent \( |\Psi(T)\rangle \) is different from \( |\Psi(0)\rangle \):

\[
\langle \Psi(0) | \Psi(T) \rangle = \int \Psi^* E' |\Psi(E)\rangle |E\rangle e^{i E t / \hbar} dE = F(t)
\]  

(4.74)

So \( F(t) \) tells us how much the wavefunction was displaced. Let’s take the Fourier transform of \( P^1_r(E_1) \), \( P^2_r(E_2) \), and \( P^{12}_r(E_1 + E_2 = E_{\text{tot}}) \):

\[
P^1_r(E - E') = \int F_1(t_1) e^{i (E - E') t_1} dt_1
\]
when a nonlocal exchange of modular variables could violate causality, the quantum uncertainty is confined to a region.

These results are similar to the general statement of uncertainty proven in the previous section: if a wavefunction (i.e. fig. 11.b) and the non-superoscillatory distribution (i.e. fig. 11.a) for $T > T'$ then we could calculate the average of the modular variables for $t > T$, but according to this result, by the time we can measure $E_{mod}$, it is completely uncertain. Thus we have proven that precisely when a nonlocal exchange of modular variables could violate causality, the quantum uncertainty masks the nonlocal change so that it is un-measurable. In other words, $F_1(t) \neq F_1(0)$ only for $t > T$. Furthermore, if we calculate all the moments of the energy, we will see that none of them change. However, if we calculate the average of the modular variables for $t > T$, then we will see a difference in $F_1$ between the original distribution and the distribution of $F_1$ after the nonlocal exchange.

In summary, the fact that $F(t) = \int P(E)e^{iEt/h}dE$ is zero in a certain range means that only modular energy could be exchanged in this range. However, if we know that a particle passes a point by the time $T$, then $E \mod h/T$ is completely uncertain. Now $E \mod h/T'$ is exchanged for $T' > T$ and that precisely characterizes the difference in the superoscillatory distribution (i.e. fig. 11.b) and the non-superoscillatory distribution (i.e. fig. 11.a)\textsuperscript{24}

\textsuperscript{24} These results are similar to the general statement of uncertainty proven in the previous section: if a wavefunction is confined to a region $L$, then $p \mod h/L$ is completely uncertain.
4.3. New way to measure the nonlocality in the equation of motion for modular variables by using weak measurements

It is a bit anti-climatic if in the only place a system acts non-locally, the non-locality cannot be seen because it is uncertain. However, as will be shown in a forthcoming paper [34, 35] using the weak-measurements discussed in §2.1.1.3, the non-local equations of motion for each individual particle can be given a more physical meaning. In summary, we consider again eq. 2.12

$$\hat{A}^{(N)}|\Psi^{(N)}\rangle = \langle \hat{A} | |\Psi^{(N)}\rangle + \frac{\Delta A}{N} \sum_{i=1}^{N} |\Psi_\perp^{(N)}(i)\rangle,$$

where $\hat{A}_i \equiv 1/N \sum_{i=1}^{N} \hat{A}_i$, but in this case, we will take $\hat{A}_i = e^{i\pi L/H}$. We consider a double-slit setup in which the right slit is always open and the left slit can be either open or closed. Suppose we consider $N$ single particles, each in the state eq. 3.36, $\Psi = \Psi_1 + e^{i\alpha} \Psi_2$, where $\Psi_1$ is localized around the left slit and $\Psi_2$ is localized around the right slit. If a weak-measurement of $\hat{A}^{(N)} = 1/N \sum_{i=1}^{N} e^{i\pi L/H}$ is performed on the $N$ particles before entering the double-slit setup, then MD will register $\bar{A} = e^{i\alpha}$, and, as shown in §2.1.2.2, not even a single particle will be disturbed. But, if instead of eq. 3.36, we sent in $N$ single wavepackets, each in the state $\Psi_2$ localized around the right slit, then MD will yield $\bar{A} = 0$, and again not even a single particle is disturbed.

Suppose we now perform an ideal measurement of $e^{i\pi L/H}$ on each particle after they travel through the double-slit setup. Since $e^{i\pi L/H}$ just translates a wavepacket by a distance $L$, the measurement will simply kick the particles localized around the right slit to be localized around the left slit which is a distance $L$ away.

We now consider two possibilities: a) the left slit is open or, b) the left slit is closed.

- If the left slit is open, and under the situation in which the subsequent ideal measurement yields the outcome $\bar{A} = e^{i\alpha} L$, then by conservation of modular momentum (see §3.2), then the earlier weak-measurement must also have yielded $\bar{A} = e^{i\alpha} L$.
- If the left slit is closed, then all the particles initially in the state $\Psi_1 + e^{i\alpha} \Psi_2$ and upon which a weak-measurement yielded $\bar{A} = e^{i\alpha} L$ will simply travel through the right slit, so the subsequent ideal measurement will yield $\bar{A} = 0$. Under the rare situation in which the subsequent measurement yields $\bar{A} = e^{i\alpha} L$, then this cannot be explained by conservation of modular momentum.

Therefore, each of these particles “knew” that the left slit was open or closed (because otherwise, there is no way to explain why the same answer $\bar{A} = e^{i\alpha} L$ is obtained before and after in the first case), even though the probability that a single particle is kicked from right to left goes to zero. Now, from the arguments of §3.2, although we have one equation of motion in the case the left slit is open and another equation of motion in the case the left slit is closed, we argued that these could never be distinguished because the non-local exchange was uncertain. However, according to the argument presented in this section, two situations can now be distinguished and we can thus verify that non-local equations are at work.

5. Conclusion

In this article, we have shown that all superoscillatory phenomenon have the following general characteristic:

- the superoscillations only exist in a finite region and do not exist in the global region,
- the superoscillations are created by interacting with a system which was suitably pre- and post-selected,
- the seemingly contradictory creation of high-momentum can be understood as a non-local exchange of modular energy with the window opener, even though the probability distribution of the opener did not change,
only the average of modular energy for $t > T'$ will change as compared to the original
distribution and this precisely characterizes the high-momentum corresponding to the
super-oscillation.

Acknowledgments: The author thanks Yakir Aharonov and Sandu Popescu for many
fascinating conversations and NIST for support.

Appendix A. Additional material on derivation of Weak-values and
Weak-measurements
In the Schroedinger picture of measurement, the system and MD state are:

$$|\Phi_{\text{tot}}\rangle = |\Psi_{\text{in}}\rangle|\Phi_{\text{md}}^{\text{in}}\rangle \rightarrow e^{-i\int H_{\text{int}} dt} |\Psi_{\text{in}}\rangle|\Phi_{\text{md}}^{\text{in}}\rangle = e^{i\lambda Q_{\text{md}} A}|\Psi_{\text{in}}\rangle|\Phi_{\text{md}}^{\text{in}}\rangle \quad (A.1)$$

where the state of the system is $|\Psi_{\text{in}}\rangle$ and the MD state, $|\Phi_{\text{md}}^{\text{in}}\rangle$, is given by $|\Phi_{\text{md}}^{\text{in}}\rangle = \int dQ_{\text{md}}|\Phi_{\text{md}}^{\text{in}}(Q_{\text{md}})\rangle|Q_{\text{md}}\rangle = \int dP_{\text{md}}|\Phi_{\text{md}}^{\text{MD}}(P_{\text{md}})\rangle|P_{\text{md}}\rangle$. A good approximation for realistic experiments is to consider MD’s initial state as a Gaussian (without loss of generality), i.e.

$$\Phi_{\text{md}}^{\text{in}}(Q_{\text{md}}) \equiv \langle Q_{\text{md}} | \Phi_{\text{md}}^{\text{in}} \rangle = \exp(-\frac{Q_{\text{md}}^2}{4\Delta^2})$$ and $\Phi_{\text{md}}^{\text{MD}}(P_{\text{md}}) \equiv \langle P | \Phi_{\text{md}}^{\text{in}} \rangle = \exp(-\Delta^2P^2)$ (substituting

$\Delta \equiv \Delta Q_{\text{md}}, \Delta P_{\text{md}} \equiv \frac{1}{\Delta}$, leaving off the normalizations). Expanding $|\Phi_{\text{in}}\rangle$ in eigenstates of $A$,

$$|\Phi_{\text{in}}\rangle = \sum_i |a_i\rangle \langle a_i | \Psi_{\text{in}}\rangle = \sum_i a_i |a_i\rangle \quad \text{then eq. A.1, becomes:}$$

$$\sum a_i \int dQ_{\text{md}} e^{i\lambda Q_{\text{md}} a_i} e^{-\frac{Q_{\text{md}}^2}{4\Delta^2}} |a_i\rangle|Q_{\text{md}}\rangle = \sum a_i \int dPe^{-i(P - a_i)^2} |a_i\rangle|P\rangle \quad (A.2)$$

When the uncertainty $\Delta P_{\text{md}} = \frac{1}{\Delta Q_{\text{md}}} |a_i\rangle$ in MD is much smaller than the shift of MD $\delta P_{\text{md}} = \lambda a_i$

corresponding to the strength of interaction and different eigenvalues, then the final state of MD is a density matrix representing a series of peaks, each corresponding to a different eigenvalue $a_i$, i.e. after tracing over the state of the measured system, the absolute square of eq. A.2 yields $\text{Pr}(P_{\text{md}}) = \sum |a_i|^2 e^{-2\Delta^2(P_{\text{md}} - \lambda a_i)^2}$. In other words, MD goes into a state of superposition proportional to the system. If $\Phi_{\text{fin}}^{\text{in}}(P_{\text{md}} - \lambda a_i)$ is orthogonal to $\Phi_{\text{fin}}^{\text{in}}(P_{\text{md}} - \lambda a_j)$ when $i \neq j$ (which occurs if $\delta P_{\text{md}} \gg \Delta P_{\text{md}}$), then this macroscopic superposition collapses into a single peak with probability given by the Born rule $\text{Pr}(a_i,t|\Psi_{\text{fin}},t_{\text{in}}) = |\langle a_i | U_{t_{\text{fin}} - t} | \Psi_{\text{fin}} \rangle|^2 \equiv |a_i|^2$, i.e. depending only on the initial state of the measured system.

To see how the WV arises from a weakened measurement with post-selection more precisely, we expand the post-selected state $|\Psi_{\text{fin}}\rangle$ in eigenstates of $A$, i.e. $|\Psi_{\text{fin}}\rangle = \sum_i |a_i\rangle \langle a_i | \Psi_{\text{fin}}\rangle = \sum_i \beta_i |a_i\rangle$ and then consider the final state of MD in the position representation:

$$|\Phi_{\text{fin}}^{\text{MD}}\rangle = \langle \Psi_{\text{fin}} | e^{iH_{\text{int}} dt} | \Psi_{\text{in}}\rangle |\Phi_{\text{md}}^{\text{MD}}\rangle = \sum_i \alpha_i \beta_i^* \int dP_{\text{md}} e^{iP_{\text{md}} a_i} e^{-\Delta^2 P_{\text{md}}^2} |P_{\text{md}}\rangle \quad (A.3)$$

However, Aharonov, Albert, and Vaidman performed this calculation in a different way [3]:

$$\Phi_{\text{fin}}^{\text{MD}}(Q_{\text{md}}) = \langle Q_{\text{md}} | \Psi_{\text{fin}} \rangle |\Phi_{\text{tot}}\rangle = \langle \Psi_{\text{fin}} | e^{i\lambda Q_{\text{md}} A} | \Psi_{\text{in}}\rangle e^{-\frac{Q_{\text{md}}^2}{4\Delta^2}}$$

$$= \sum_{n=0}^{\infty} \frac{(-i\lambda Q_{\text{md}})^n}{n!} \langle \Psi_{\text{fin}} | \hat{A}^n | \Psi_{\text{in}}\rangle e^{-\frac{Q_{\text{md}}^2}{2\Delta^2}} = \langle \Psi_{\text{fin}} | \Psi_{\text{in}}\rangle \sum_{n=0}^{\infty} \frac{(-i\lambda Q_{\text{md}})^n}{n!} (A^n)_{w} e^{-\frac{Q_{\text{md}}^2}{2\Delta^2}}$$

$$= \langle \Psi_{\text{fin}} | \Psi_{\text{in}}\rangle \{ 1 + i\lambda Q_{\text{md}} A_w + \sum_{n=2}^{\infty} \frac{(i\lambda Q_{\text{md}})^n}{n!} A^n_w e^{-\frac{Q_{\text{md}}^2}{2\Delta^2}} \}$$

36
momentum, we apply the Baker-Campbell-Hausdorff identity
\[ \exp(i\lambda Q_{md}) A_w \exp(-i\lambda Q_{md}) = A_w + \frac{(i\lambda Q_{md})^n}{n!} (A_w)^n \]

The second term in the last part of eq. A.4 can be neglected by:

(i) minimizing \( \lambda \Delta Q_{md} \) by either using a small \( \lambda \) (setting \( \Delta P_{md} = \Delta Q_{md} = 1 \)), or by minimizing the spreads in MD, e.g. so that \( P_{md} \) is measured to a finite precision \( \Delta P_{md} \), (which limits the disturbance by a finite amount \( \Delta Q_{md} \geq 1/\Delta P_{md} \)), or

(ii) minimizing \( [A^n_w - (A_w)^n]/n! \) even if \( \lambda \Delta Q_{md} \) is not small.

For the first case, by way of example, the first moment in the Taylor’s expansion (from the second term in the last part of eq. A.4) can be neglected if \( (\lambda \Delta Q_{md})^2 \Delta A_w \ll 1 \) where \( \Delta A_w \equiv ||(A^2)_w - (A_w)^2||^{1/2} \) [4]. When eq. A.4 is transformed back to the \( P_{md} \) representation, then the final state of the measuring-device after weak-measurement and post-selection is (up to normalization):

\[ \tilde{\Phi}_{\text{fin}}^{MD}(P_{md}) \approx \langle \Psi_{\text{fin}} | \Psi_{\text{in}} \rangle \langle P_{md} | e^{-i\lambda Q_{md} A_w} | \Psi_{\text{in}} \rangle \approx \exp \left\{ -\Delta^2 (P_{md} - \lambda A_w)^2 \right\} \]

(A.5)

where \( A_w = \langle \Psi_{\text{fin}} | \hat{A} | \Psi_{\text{in}} \rangle \)

(A.6)

The final state of MD is almost unentangled with the system and has the same initial shape but shifted by a very surprising amount, the WV, \( A_w \) (the factor \( \langle \Psi_{\text{fin}} | \Psi_{\text{in}} \rangle \) arises as a result of the exclusion of other post-selections). Since the value of \( \hat{A} \) is given by \( P_{md}(T) - P_{md}(0) \), we may conclude \( \hat{A} \approx A_w \). We have used such limited disturbance measurements to explore many paradoxes (see, e.g. [7]). There have also been a number of experiments to test the predictions made by the weak-measurement and their results are in very good agreement with theoretical predictions [25, 27-30].

Appendix B. Modular position
To establish the relationship between modular position, \( x_{\text{mod}} = x \mod L \), and modular momentum, we apply the Baker-Campbell-Hausdorff identity

\[ e^A e^B = e^{A+B+\frac{1}{2}[A,B]+\frac{1}{12}[A,[A,B]]-[B,[A,B]]} + \ldots \]

(B.1)

to calculate the commutation relation of \( e^{i2\pi x_{\text{mod}}/L'} \) and \( e^{\frac{i}{\hbar} P_{\text{mod}} L} \) and we obtain:

\[ e^{\frac{i}{\hbar} P_{\text{mod}} L} e^{i2\pi x/L'} = e^{\frac{i}{\hbar} P_{\text{mod}} L + i2\pi x/L' + iL\pi/L'} \]

(B.2)

and

\[ e^{i2\pi x/L'} e^{\frac{i}{\hbar} P_{\text{mod}} L} = e^{\frac{i}{\hbar} P_{\text{mod}} L + i2\pi x/L' - iL\pi/L'} \]

(B.3)

therefore, if \( L/L' \) is an integer, then \( P_{\text{mod}} \) and \( x_{\text{mod}} \) commute:

\[ [e^{i2\pi x_{\text{mod}}/L'}, e^{\frac{i}{\hbar} P_{\text{mod}} L}] = 0 \]

(B.4)
Figure B1. (a) phase space of particles with definite modular position, \( x \mod L = L/4 \) and definite modular momentum \( p \mod \hbar/L = \frac{1}{2}\hbar/L \) and cell size \( \Delta x \Delta p = \hbar \); (b) intermediate situation: more knowledge of which cell we are in means less knowledge of modular variable.

We may thus state the following relations:

\[
\left[ e^{i\hbar p_{\text{mod}}L}, N_x \right] = e^{i\hbar p_{\text{mod}}L} \tag{B.5}
\]

and

\[
\left[ e^{i2\pi x_{\text{mod}}/L}, N_p \right] = e^{i2\pi x_{\text{mod}}/L} \tag{B.6}
\]

where \( N_x \) and \( N_p \) are given by \( x = N_x L + x_{\text{mod}} \) and \( p = N_p (\hbar/L) + p_{\text{mod}} \). These two commutation relations are similar to the angular momentum-angle relation \( [L_z, e^{i\phi}] = e^{i\phi} \): if the angular momentum is known precisely then the angle is completely uncertain. Thus, \( p_{\text{mod}} \) is conjugate to \( N_x \) and \( x_{\text{mod}} \) is conjugate to \( N_p \). Similarly, if the system is localized in space, i.e. \( N_x \) is known, then the modular momentum is completely uncertain:

\[
\Delta N_x \Delta p_{\text{mod}} \geq \frac{\hbar}{L} \tag{B.7}
\]

Or, if the momentum is well known, then the modular position is uncertain:

\[
\Delta N_p \Delta x_{\text{mod}} \geq L \tag{B.8}
\]

The standard Heisenberg uncertainty principle does not allow us to localize a particle in phase space to anything less than an area of \( \hbar \). The uncertainty principle for modular variables, however, allows us to precisely locate a spot within a cell of area \( \hbar \), i.e. \( x_{\text{mod}} \) and \( p_{\text{mod}} \) are certain, but it gives no information about which cell it is in, i.e. \( N_x \) and \( N_p \) are completely uncertain. This suggests that we can have precise but partial information about the momentum, namely the modular momentum, and simultaneously precise but partial information about the location, namely the modular position (see figure B1.a). We can also have intermediate situations, i.e. less precise information of the exact point within a phase cell but more information of which cell we’re in (see figure B1.b).

Appendix C. Proof that uncertainty in momentum is independent of \( a_a \)

We have \( \Delta p = \left( \langle p^2 \rangle - \langle p \rangle^2 \right)^{1/2} \). For \( \langle p \rangle \) we have:

\[
\langle p \rangle = \int_{-\infty}^{+\infty} dx \Psi^*(x)(-i\hbar) \frac{\partial}{\partial x} \Psi(x) = \int_{-\infty}^{+\infty} dx \sum_{n,m} a_n^* a_m \Psi_n^*(x)(-i\hbar) \frac{\partial}{\partial x} \psi_m(x) \tag{C.1}
\]
But $\psi_n$ and $\psi_m$ do not overlap, so that $\psi_m$ and its derivatives are different from zero only where $\psi_n = 0$. $\psi^*(x) \frac{\partial}{\partial x} \psi(x + L) = 0$ for all $x$ if $\psi(x)$ vanishes outside the interval, then so does its derivatives. Thus

$$\langle p \rangle = \sum_{n,m} \alpha_n^* \alpha_m \frac{\hbar}{i} \int_{-\infty}^{+\infty} dx \psi^*_n(x) \frac{\partial}{\partial x} \psi_n(x) \tag{C.2}$$

Further, since $\psi_n(x) = \psi_n(x+nL)$, we have

$$\int_{-\infty}^{+\infty} dx \psi^*_n(x) \frac{\partial}{\partial x} \psi_n(x) = \int_{-\infty}^{+\infty} dx \psi^*_n(x+nL) \frac{\partial}{\partial x} \psi_n(x+nL) = \int_{-\infty}^{+\infty} dx \psi^*_n(x) \frac{\partial}{\partial x} \psi_n(x) \tag{C.3}$$

Therefore,

$$\langle p \rangle = \sum_{n} |\alpha_n|^2 \frac{\hbar}{i} \int_{-\infty}^{+\infty} dx \psi^*_n(x) \frac{\partial}{\partial x} \psi_n(x) = \frac{\hbar}{i} \int_{-\infty}^{+\infty} dx \psi^*_n(x) \frac{\partial}{\partial x} \psi_n(x) \tag{C.4}$$

I.e. $\langle p \rangle$ is independent of the choice of $\alpha_n$. In a similar way, one can show that

$$\langle p^2 \rangle = \sum_{n} |\alpha_n|^2 \frac{\hbar^2}{i} \int_{-\infty}^{+\infty} dx \psi^*_n(x) \left( \frac{\partial}{\partial x} \right)^2 \psi_n(x) = \frac{\hbar^2}{i} \int_{-\infty}^{+\infty} dx \psi^*_n(x) \left( \frac{\partial}{\partial x} \right)^2 \psi_n(x) \tag{C.5}$$

Thus $\langle \Delta p \rangle^2 = \langle (p^2) - \langle p \rangle^2 \rangle$ is independent of the choice of $\alpha_n$.

**Appendix D. Proof.**

$$\Delta p = \Delta p \text{ for the state } \Psi \equiv \psi_n(x) = \frac{1}{\sqrt{2}} (\psi(x) + e^{i\alpha} \psi(x + L))$$

Following Appendix C very closely, we note again that $\psi^*(x) \frac{\partial}{\partial x} \psi(x + L) = 0$ for all $x$. If $\psi(x)$ vanishes outside the interval, then so does its derivatives. For any non-negative integer $n$, $\langle p^n \rangle_{\Psi} = \langle \psi \rangle \left( -i\hbar \frac{\partial}{\partial x} \right)^n |\psi\rangle \equiv \langle p^n \rangle$. Then: $\Delta p_n = \langle p^n \rangle - \langle p^n \rangle^2 / 2 = \langle p^n \rangle / 2 = \langle p^n \rangle / 2 = \Delta p$ Which is proven by:

$$\langle p^n \rangle_{\Psi} = \int_{-\infty}^{+\infty} dx \psi^*_n(x) \left( -i\hbar \right)^n \frac{\partial^n}{\partial x^n} \psi_n(x) = \frac{1}{\sqrt{2}} \left( \int_{-\infty}^{+\infty} dx \left( \psi^*(x) + e^{i\alpha} \psi^*(x + L) \right) \left( -i\hbar \right)^n \frac{\partial^n}{\partial x^n} \left( \psi(x) + e^{i\alpha} \psi(x + L) \right) = \frac{1}{2} \left( \int_{-\infty}^{+\infty} dx \psi^*(x) \left( -i\hbar \right)^n \frac{\partial^n}{\partial x^n} \psi(x) + \int_{-\infty}^{+\infty} dx \left( \psi^*(x + L) \left( -i\hbar \right)^n \frac{\partial^n}{\partial x^n} \psi(x + L) \right) \right) \right) \cdot$$

If we change the variables $x + L \rightarrow x$ in the second integral then we obtain:

$$\langle p^n \rangle_{\Psi} = \int_{-\infty}^{+\infty} dx \psi^*(x) \left( -i\hbar \right)^n \frac{\partial^n}{\partial x^n} \psi(x) = \langle p^n \rangle$$

**Appendix E. Proof of 3.36.**

If $[\psi_n^* \psi_n - \psi_n^* \psi_m] x^n dx = 0$ if no overlap of $\psi_1$ and $\psi_2$, $n \in \text{integer}$

Refer to eq 3.36, we see that $\psi^*_n(x,0) \psi_n(x,0)$ is independent of $\alpha$, and hence $\psi^*_n(x,0) \psi_n(x,0) - \psi^*_m(x,0) \psi_m(x,0) = 0$. Therefore, at $t = 0$, $\langle x^n \rangle$ is independent of $\alpha$, as is $\langle p^n \rangle$. The latter follows from the non-overlapping nature of $\psi_1(x,0)$ and $\psi_2(x,0)$. It is also easy to show that $\langle x^n \rangle$ is independent of $\alpha$ by using the Heisenberg representation $\langle x^n(t) \rangle = \int \psi^*_n(x,0)x^n(t) \psi_n(x,0) dx$ and noting that $x(t) = x(0) + p(t) \frac{\partial}{\partial x}$ and $p(t) = p(0)$ in this representation, we must have $\langle x^n(t) \rangle = \int \psi^*_n(x,0)x^n(0) + p(0) \frac{\partial}{\partial x} \int \psi_n(x,0) dx$. This is clearly independent of $\alpha$, since term by term it is independent of $\alpha$. Eq. (3.41) then follows, and holds for $p^a$, as long as we retain the proper $\psi^*_n p^n \psi_n$ order.

**Appendix F. Proof of complete uncertainty principle in §3.2.2**

We expand the probability density $Prob(\theta)$ to a Fourier series $Prob(\theta) = \sum_{n=-\infty}^{+\infty} a_n e^{i\theta} \text{ (integer $n$ is a requirement for the function to be periodic in $\theta$)}$, where $a_n = \int Prob(\theta) e^{i\theta} d\theta = |e^{i\theta}|$ (since the average of any function is given the integral of the function with the probability). We see that $Prob(\theta) = \text{const}$ if and only if $a_n = 0$ for all $n \neq 0$, and therefore $|e^{i\theta}| = 0$ for $n \neq 0$.

**Appendix G. Proof of Theorem 1**

Proof: from linearity $\frac{\langle p_{\text{fin}} | p_{\text{fin}} \rangle - \langle p_{\text{coh}} | p_{\text{fin}} \rangle + \langle p_{\text{coh}} | p_{\text{fin}} \rangle - \langle p_{\text{coh}} | p_{\text{fin}} \rangle}{\langle p_{\text{coh}} | p_{\text{coh}} \rangle - \langle p_{\text{coh}} | p_{\text{coh}} \rangle + \langle p_{\text{coh}} | p_{\text{coh}} \rangle - \langle p_{\text{coh}} | p_{\text{coh}} \rangle}$
Appendix H. Proof of Theorem 2
Proof: Given that \( \hat{P}_A = \sum_n \alpha_n |\alpha_n\rangle\langle\alpha_n| \), if an eigenvalue, e.g. \( \hat{P}_A = \alpha_m \), is obtained with certainty, then for \( n \neq m \), \( \hat{P}_A \equiv |\alpha_m\rangle\langle\alpha_m| = 0 \) because the probability to obtain another eigenvalue by ABL is \( \propto \langle \Psi_{\text{fin}} | \alpha_m \rangle \langle \alpha_m | \Psi_{\text{in}} \rangle = 0 \). In this case, the weak-value \( \langle \hat{P}_A \rangle_w = (|\alpha_m\rangle\langle\alpha_m|)_w = \frac{\langle \Psi_{\text{fin}} | \alpha_m \rangle \langle \alpha_m | \Psi_{\text{in}} \rangle}{\langle \Psi_{\text{fin}} | \Psi_{\text{in}} \rangle} = 0 \). In addition, \( \sum_m \frac{\langle \Psi_{\text{fin}} | \alpha_m \rangle \langle \alpha_m | \Psi_{\text{in}} \rangle}{\langle \Psi_{\text{fin}} | \Psi_{\text{in}} \rangle} = 1 \) because \( \sum_m |\alpha_m\rangle\langle\alpha_m| = 1 \). But since \( \langle \Psi_{\text{fin}} | \alpha_m \rangle \langle \alpha_m | \Psi_{\text{in}} \rangle = 0 \) for \( n \neq m \), the only term left is \( n \). Therefore, the weak-value is 1, the same as the ideal value.

References
[1] Y. Aharonov, P. G. Bergmann, and J. L. Lebowitz, Phys. Rev. 134, B1410 (1964), reprinted in Quantum Theory and Measurement, eds. J. A. Wheeler and W. H. Zurek (Princeton University Press), 1983, pp. 680-686.
[2] Y. Aharonov and D.Z. Albert, Phys. Rev. D12 (1980) 3316, D24 (1981) 359.
[3] Y. Aharonov, D. Albert, and L. Vaidman, Phys. Rev. Lett. 60 (1988), 1351.
[4] Y. Aharonov, L. Vaidman, Phys. Rev. A, 41 (1990), 11.
[5] Y. Aharonov and L. Vaidman, J. Phys. A 24, 2315 (1991).
[6] Y. Aharonov, S. Massar, S. Popescu, J. Tollaksen, and L. Vaidman, “Adiabatic Measurements on Metastable Systems,” Phys. Rev. Lett., 77, p. 983, (1996).
[7] Aharonov Y, Botero A, Popescu S, Reznik B, Tollaksen J, “Revisiting Hardy’s Paradox: Counterfactual Statements, Real Measurements, Entanglement and Weak Values,” PHYS LETT A 301 (3-4): 130-138 AUG 26 2002.
[8] “Evanescent and real waves in quantum billiards and Gaussian beams,” M.V. Berry, 2000, J. Phys. A. 27 L391, and Faster than Fourier, 1994, Fundamental Problems in Quantum Theory ed. J.A. Anandan and J. Saiko.
[9] Berry, M V and Popescu, S, 2006, 'Evolution of quantum superoscillations, and optical superresolution without evanescent waves', J.Phys.A 39 6965-6977.
[10] Y. Aharonov, H. Pendleton and A. Petersen, Int. J. Theor. Phys., 2, (1969), 213.
[11] Y. Aharonov, H. Pendleton and A. Petersen, Int. J. Theor. Phys., 3, (1970), 443.
[12] Aharonov, Yakir, Rohrlich, Daniel, Quantum paradoxes: quantum theory for the perplexed, Weinheim : Cambridge : Wiley-VCH, 2005.
[13] Aharonov, Y., unpublished notes.
[14] B.-G Englert, C. Kurtsiefer, H. Weinfurter Phys Rev A, Vol. 63, art. 032303.
[15] A. Beige, B.-G. Englert, C. Kurtsiefer, and H. Weinfurter, quant-ph/0101066.
[16] Chiao Ry, Steinberg AM, Prog in Opt 37: 345-405 1997; Solli DR, McCormick CF, Chiao Ry, et al. Phys Rev Lett 92 (4): Art. No. 043601 JAN 30 2004.
[17] M. Gell-Mann and J.B. Hartle, Phys. Rev. D 47, 3345 (1993).
[18] A. M. Steinberg, Phys. Rev. Lett. 74, 2405(1995).
[19] Y. Aharonov, L. Davidovich, N. Zagury, Phys. Rev. A48 (1993) 1687.
[20] A. Botero and B. Reznik, Phys. Rev. A, Vol. 61, 050301-4, (2000).
[21] Bub, J., Phys. Rev. A 63 (2001) 032309.
[22] Aharonov Y., Gruss, E., quant-ph/0507269.
[23] M.Duck, P.M. Stevenson, and E.C.G. Sudarshan, Phys. Rev.D 40, 2112 (1989).
[24] J. von Neumann, Mathematical Foundations of Quantum Theory, Princeton, University Press, New Jersey (1983).
[25] Ahnert SE, Payne MC, Phys. Rev. A, 70 (4): Art. No. 042102 OCT 2004.
[26] The usual projective measurement typically utilized in quantum experiments is a special case of these weak measurements, see Ognyan Oreshkov, Todd A. Brun, “Weak measurements are universal,” Phys. Rev. Lett. 95, 110405 (2005).
[27] Parks AD, Cullin DW, Stoudt DC, Proc. of the Royal Soc. of London Series A, 454 (1979): 2997-3008 NOV 8 1998.
[28] Wiseman HM Phys. Rev. A 65 (3): Art. No. 032111 Part A MAR 2002.
[29] Pryde GJ, O’Brien JL, White AG, Ralph TC, Wiseman HM, Phys. Rev. Lett., 94 (22): Art. No. 220405 JUN 10 2005.
[30] N.W. M. Ritchie, J. G. Story and R. G. Hulet, Phys. Rev. Lett. 66, 1107 (1991).
[31] Sandu Popescu and Daniel Rohrlich, “How a Soft Photon Can Emit a Hard Photon”, TAUP 1847-90, May 1991.
[32] Y. Aharonov, A. Casher, D. Albert, and L. Vaidman. Phys. Lett. A124, 199 (1987).
[33] Y. Aharonov, “Non-Local Phenomenon and the Aharonov-Bohm Effect,” in Proc. Int. Symp. Foundations of Quantum Mechanics, Tokyo, 1983, pp. 10-19.
D. Albert, Y. Aharonov, and S. D’Amato, Y. Aharonov, L Vaidman, in “Time in Quantum Mechanics,” In: VISIONS OF DISCOVERY: New Light on Physics, Cosmology, and Consciousness, ed. R. Y. Chiao, M. L. Cohen, A. J. Leggett, W. D. Phillips, and C. L. Harper, Jr. Cambridge: Cambridge University Press, 2007.

Theorem: for every observable $A$ and a normalized state $|\psi\rangle$, we have: $A|\psi\rangle = (A|\psi\rangle + \Delta A|\psi_{\perp}\rangle)$ for some state $|\psi_{\perp}\rangle$ which is orthogonal to $|\psi\rangle$. To prove this, we begin with: $A|\psi\rangle = (A|\psi\rangle + A|\psi\rangle - A|\psi\rangle)$ now we set: $|\psi_{\perp}\rangle = A|\psi\rangle - A|\psi\rangle$, so: $\langle\psi_{\perp}|\psi\rangle = ((\langle\psi_{\perp}|A)\langle\psi_{\perp}|\psi\rangle = \langle\psi|A|\psi\rangle - \langle A|\psi\rangle = 0$ now we set: $|\psi_{\perp}\rangle = b|\psi_{\perp}\rangle$, where $|\psi_{\perp}\rangle$ is normalized and $b$ real (note that $\langle\psi_{\perp}|\psi\rangle = 0$). so: $A|\psi\rangle = (A|\psi\rangle + b|\psi_{\perp}\rangle$. Now we multiply from the left by $|\psi_{\perp}\rangle$, and we get: $\langle\psi_{\perp}|A|\psi\rangle = b$. Now we can see that: $\langle\psi|A^{2}|\psi\rangle = \langle\psi|A(A|\psi\rangle + b|\psi_{\perp}\rangle) = \langle\psi|(A|A|\psi\rangle + b(A|\psi_{\perp}\rangle) = \langle A|^{2} + b\langle\psi|A|\psi_{\perp}\rangle$ so: $\langle A^{2} - \langle A\rangle^{2} = 0$ which means that: $b = \sqrt{\langle A^{2} - \langle A\rangle^{2}} = \Delta A$ and the result: $A|\psi\rangle = (A|\psi\rangle + \Delta A|\psi_{\perp}\rangle)$ is proved.

The identity $\exp\{i\alpha\sigma_{y}\} = \cos\alpha + i\sigma_{y}\sin\alpha$ is easily proven using the fact that for any integer $k$: $\sigma_{y}^{2k} = I$ and $\sigma_{y}^{2k+1} = \sigma_{y}$, and now it follows that: $e^{i\alpha\sigma_{y}} = \sum_{k=0}^{\infty} \frac{(i\alpha)^{2k}}{2k!} = \sum_{k=0}^{\infty} \frac{(i\alpha)^{2k}}{2k!} + \sigma_{y} \sum_{k=0}^{\infty} \frac{(i\alpha)^{2k+1}}{(2k+1)!} = e^{i\alpha\sigma_{y}} = \cos\alpha + i\sigma_{y}\sin\alpha$ and the identity is proven.

Y Aharonov, L Vaidman, in “Time in Quantum Mechanics”, ed J Muga, R Sala Mayato and I Egusquiza; quant-ph/0105101.

Resch KJ, Lundeen JS, Steinberg AM Phys Lett A 324 (2-3): 125-131 APR 12 2004.

Tollaksen, J, “Quantum properties that are extended in time,” Quantum Information and Computation V, Ed by E Donkor, A Pirich, H Brandt, Proc of SPIE Vol. 6573 (SPIE, Bellingham, WA, 2007), CID 6573-36.

Tollaksen, J, “Pre-and-post-selection, weak-values, and contextuality”, to appear in Journal of Physics A: Mathematical and General (2007); quant-ph/0602226.