Buridan's Principle

Leslie Lamport
Digital Equipment Corporation
Systems Research Center

31 October 1984
revised: 24 February 2012

To appear in *Foundations of Physics*.

1 Buridan's Ass

The problem of Buridan's Ass, named after the fourteenth century French philosopher Jean Buridan, states that an ass placed equidistant between two bales of hay must starve to death because it has no reason to choose one bale over the other. With the benefit of modern mathematics, the argument can be expressed as follows. Assume that, at time 0, the ass is placed at position x along the line joining the bales of hay, where one bale is at position 0 and the other at position 1, so $0 < x < 1$. The position of the ass at time $t > 0$ is a function of two arguments: the time t and the starting position x. Let $A_t(x)$ denote that position. For simplicity, assume that when the ass reaches a bale of hay it stays there forever, so for all $t \geq 0$: $A_t(0) = 0$, $A_t(1) = 1$, and $0 \leq A_t(x) \leq 1$ for any x with $0 < x < 1$. The ass is a physical mechanism subject to the laws of physics. Any such mechanism is continuous, so $A_t(x)$ is a continuous function of x. Since $A_t(0) = 0$ and $A_t(1) = 1$, by continuity there must be a finite range of values of x for which $0 < A_t(x) < 1$. These values represent initial positions of the ass for which it does not reach either bale of hay within t seconds. Such a range of values of x exists for any time t, including times large enough to insure that the ass has starved to death by then. Thus, there exists a finite range of starting positions for which the ass starves to death.

The key assumption in this argument is continuity: the ass's position at a later time is a continuous function of its initial position. Continuity has been a guiding principle in the development of modern physics. Phenomena that appear discontinuous, such as discrete atomic spectral lines, are explained in terms of continuous physical laws, such as Schroedinger's equation. The assumption of continuity is discussed at length in Section 6. For now, let us accept it and investigate its consequences.

The general principle underlying the starvation of Buridan's ass can be stated as follows:

> *Buridan's Principle.* A discrete decision based upon an input having a continuous range of values cannot be made within a bounded length of time.

Buridan's ass starves because it cannot make the discrete decision of which pile of hay to eat, a decision based upon an initial position having a continuous range of values, within the bounded length of time before it starves. A continuous mechanism must either forgo discreteness, permitting a continuous range of decisions, or must allow an unbounded length of time to make the decision.

Let us examine a more modern manifestation of this principle. Consider a car at an unguarded railroad crossing, where the driver (our "ass") stops at the entrance to the crossing and then proceeds when it is safe. The driver must make a discrete decision, to wait for the train or to cross the tracks, in the bounded length of time before the train gets there. By Buridan's Principle, this is impossible. The consequences of this impossibility can be disastrous.

The position of the car at time t will be a function $A_t(x)$ of the time t and some parameter x. For convenience, assume that the train reaches the crossing at time 0, and let the parameter x be the time at which the car reaches the stop sign at the entrance to the crossing. Thus, $A_0(x)$ denotes the position at time 0, the time at which the train reaches the crossing, of a car that reached the stop sign at time x. For a time $x_0 \ll 0$, a car reaching the stop sign at time x_0 will cross before the train arrives, so $A_0(x_0)$ is on the far side of the track. If the car reaches the stop sign at some time $x_1 > 0$, after the train reaches the crossing, then the driver will wait for the train to pass. Hence, $A_0(x_1)$ is on the near side of the track. By continuity, there must be some time x, with $x_0 < x < x_1$, such that $A_0(x)$ is in the middle of the track. Moreover, since the track has a finite width, there is a finite range of values of x for which $A_0(x)$ is on the track—that is, a finite range of arrival times at which the car gets hit by the train.

2 Can Asses Really Starve?

When first told of Buridan's Principle, people usually find it unbelievable and propose mechanisms to circumvent it. Indecision occurs when either decision would be acceptable; a properly cautious driver at a railroad crossing is indecisive only if he is able to cross ahead of the train, but not so far ahead that waiting would be silly. The most common suggestion is that the "ass" take some specific action when he finds himself unable to make a decision; for example, he might stop at the railroad crossing when he cannot decide if it is safe to cross. However, this merely pushes the decision back one level; the driver still must decide whether or not he can decide if it is safe to cross. No such approach can refute the conclusion because Buridan's Principle does not rely upon any assumption about how the decision is made; it rests only on the assumption of continuity.

Another often-suggested escape from Buridan's Principle is noise—the introduction of randomness into the system. In theory, one can balance a ball on a knife edge; in practice, this is impossible because tiny random vibrations will cause the ball to fall, despite our best efforts to balance it. Moreover, balancing the ball on a knife edge requires fixing very precisely both the position and the momentum of the ball, which is forbidden by Heisenberg's Uncertainty Principle. A four-legged or human ass must also have random noise and be subject to the Uncertainty Principle, so it cannot be put into a situation where it will hang forever on a knife edge of indecision.

Randomness can make it impossible deliberately to starve the ass, but it cannot prevent the ass from accidentally starving. Random vibrations make it impossible to balance the ball on the knife edge, but if the ball is positioned randomly, random vibrations are as likely to keep it from falling as to cause it to fall. In classical physics, randomness is a manifestation of a lack of knowledge—if we knew the positions and velocities of all atoms in the universe, then even the tiniest vibration could be predicted. Randomness due to insufficient knowledge does not introduce discontinuity, so it does not invalidate the hypothesis of continuity upon which Buridan's Principle rests.

The status of continuity in quantum physics is less clear than in classical physics. The laws of quantum mechanics (such as Schroedinger's equation) are continuous, and the Uncertainty Principle, like random noise, seems to prohibit only the deliberate starvation of the ass, not its accidental starvation. The relation of quantum mechanics to Buridan's Principle is discussed further in Section 6.

Despite these arguments justifying Buridan's Principle, its consequences may still seem absurd. How could a rational human being knowingly drive in front of a train? Buridan's Principle manifests itself in people as indecision. A driver caught in the "starvation" situation would be unable to decide if he had time to cross safely. He would start to cross and then hesitate, perhaps switching from the brake to the accelerator several times in the process. The reader has probably experienced a similar though less dramatic form of indecision—for example, when trying to decide whether to stop or drive through an intersection as the traffic light is changing from green to red.

Buridan's Principle may seem even less plausible when we realize that our conclusion is unchanged by the presence of a crossing gate. How can a crossing gate fail to prevent such an accident? While a detailed analysis of the physics of a car/crossing gate interaction is impossible, it is clear that if the driver hesitates while the gate is being lowered, then the gate could hit the top of the car and slow it down enough to put it in the path of the train. At best, a crossing gate can only reduce the probability that Buridan's Principle will strike, it cannot eliminate the possibility.

Although everyone has probably experienced Buridan's Principle—a relatively long period of indecision, perhaps when approaching a changing traffic light or choosing a flavor of ice cream—very few people are aware of it. Because real asses are not observed to starve to death when placed between two bales of hay, philosophers have viewed Buridan's Ass as a paradox, discussing it in connection with the problem of free will [4]. Psychologists studying human decision seem to have ignored the phenomenon. A survey article on reaction times [11] mentions two models that describe the time needed to make a binary decision. Both models predict that the decision time increases to infinity as the stimulus approaches the point at which the correct decision changes from zero to one. However, that aspect of the models is completely ignored, and curves plotted from the empirical data tacitly assume that the reaction time is bounded.

An empirical absence of dead asses does not invalidate Buridan's Principle; it simply indicates that the range of initial positions for which the ass starves is so small that the probability of placing it at such a position is negligible, even if we try. However, while it takes an ass days to starve, a few seconds of indecision at a railway crossing can be fatal. Are people hit by trains at railway crossings because of Buridan's Principle? Railroad crossing accidents do happen. When the driver is killed and no explanation can be found, it is natural to assume that he failed to see the train or else misjudged its speed or distance. Few people are aware of Buridan's Principle, and the possibility that it could cause such an accident has not, to my knowledge, been considered. Whether Buridan's Principle is responsible for real accidents seems to be an open question. An investigation is unlikely to implicate a phenomenon that the investigators have never considered. All we can safely say is that the probability of such an accident is very small, but no one knows if "very small" means that it will happen once in an eon or once a year.

3 Other Asses

The railroad crossing is an example of how a discrete decision is impossible in a bounded time. Let us now consider an example in which a discrete decision is possible when one does not bound the decision time. Consider a driver approaching a tree in the middle of the road. He must make the discrete decision of whether to drive to the left or the right. If he maintains his speed, then he must decide in the bounded time before he hits the tree, and he is in the same situation as at the railroad crossing: there must be some initial conditions that result in his driving into the tree. However,

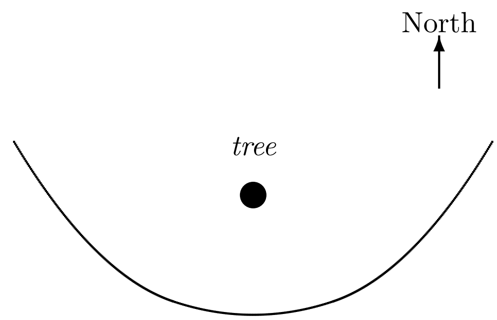

Figure 1: Position of car at time t for different initial east-west positions.

suppose that he has the option of stopping before hitting the tree. The driver can then adjust his speed as a (continuous) function of how close he will come to the tree, ensuring that he stops before he might hit it. Once stopped, he is in the same position as Buridan's original ass and could starve to death before deciding which way to go. However, he will never hit the tree, and the probability of starvation is quite low—much lower than the probability of his hitting the tree if he tried to make a decision without stopping.

It is instructive to compare this situation with the railroad crossing. Assume that the car is traveling due north, crossing a fixed east-west line at time zero, and let $A_t(x)$ represent the car's position at time t if its initial east-west position at time zero is x. For a fixed time t, the locus of all points $A_t(x)$ might look like the curve drawn in Figure 1. The locus does not pass through the tree, so there is no chance of a crash. This is possible because the position of the car is a point in a two-dimensional space, unlike the situation at the railroad crossing where there is only one degree of freedom: the distance along the road.

A more common manifestation of Buridan's Principle occurs when two people, say Alice and Bob, are walking toward each other and must decide whether to pass to the left or the right. Choosing a coordinate system fixed with respect to Alice converts this into mathematically the same situation as the car and the tree. In the correspondence, Alice becomes the tree and Bob becomes the car. Buridan's Principle is independent of whether it is the car, the tree, or both that are responsible for making the decision. Alice and Bob will stop before they collide, but it could take them arbitrarily long to pass one another.

One way to circumvent Buridan's Principle is to eliminate the decision. The ass's action could be decided in advance—always to wait for the train to pass or always to go to the right of the tree. However, such an approach is impossible in practice. A driver cannot be expected to wait when he arrives at the crossing an hour before the train, nor will he go to the right of a tree that is all the way at the right-hand side of the road.

One cannot eliminate decisions, though one can sometimes circumvent the problem by eliminating discreteness. If you cannot decide between chocolate and vanilla ice cream, try half chocolate and half vanilla. However, the process of making discrete decisions based upon continuous inputs seems to be a fundamental part of life. It is likely that this process lies at the heart of human (and animal) perception, the act of perceiving being the identification of an object as one of a discrete set of possibilities (chair, table, dog, etc.) based upon an input (a pattern of light intensities) from a continuous range of possible values.

4 Flying Asses

An interesting variation on the problem of people walking toward each other is that of two airplanes on a collision course. Buridan's Principle cannot be applied here because it is no longer a discrete decision. The pilots are not constrained only to turn right or left; they can also simultaneously climb or dive, giving them a continuous range of directions in which to pass each other. However, as I now show, the consequences are the same: if the airplanes cannot stop, but must proceed to their destinations, then there must exist a finite range of initial conditions that lead to a collision.

As in the case of the walkers, a change of coordinates converts the problem into one of avoiding a fixed obstacle—say an airplane that must pass a stationary balloon floating in the air. Suppose the airplane is flying north, and let Q be a vertical, east-west plane through the balloon. The existence of initial positions that lead to a collision rests upon the following assumptions:

1. For every initial position, the airplane must cross plane Q.

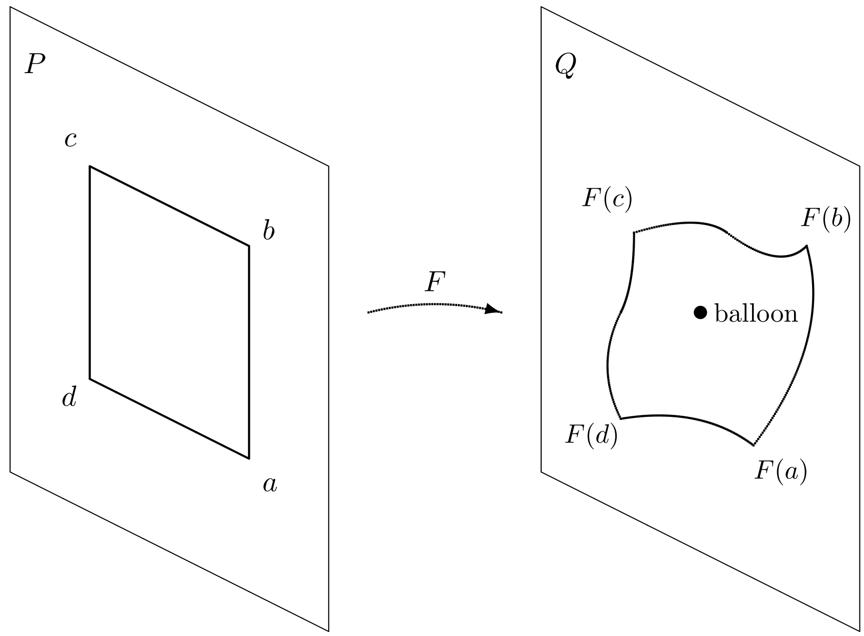

Figure 2: The airplane and the balloon.

2. If the airplane starts far enough east (west) of the balloon, then it must pass east (west) of the balloon; and if the airplane starts far enough below (above) the balloon, then it must pass below (above) the balloon.

The following argument is illustrated in Figure 2.

Let P be a vertical plane representing possible initial positions of the airplane at some time before it reaches the balloon, so P lies south of Q, and let F be the mapping from P to Q such that $F(x)$ is the point in Q reached by the airplane if it is initially at point x of P. Draw the rectangle $abcd$ in P such that: line ab is far enough east so $F(x)$ is east of the balloon for all points x on that line (that is, if the airplane is started anywhere on line ab, then it will pass east of the balloon); line bc is high enough so $F(x)$ is above the balloon for all points x on that line; line cd is far enough west so $F(x)$ is west of the balloon for all points x on that line; and line da is low enough so $F(x)$ is below the balloon for all points x on that line. The existence of rectangle $abcd$ is guaranteed by our second assumption. The function F maps rectangle $abcd$ into a closed curve in Q whose interior contains the balloon. Our first assumption implies that F maps every point in the interior of $abcd$ into a point of Q, and the continuity of F implies that there must be some point x in the interior of $abcd$ such that $F(x)$

equals the position of the balloon. (As one continuously shrinks rectangle *abcd* to a point, its image under F also shrinks continuously to a point, so the image of one of the shrunken rectangles must meet the balloon; the mathematical proof uses the Brouwer Fixed-Point Theorem [9].) Moreover, since the balloon has a finite extent, there is a finite region of initial values (points on P) for which the airplane collides with the balloon.

5 Computer Asses

Buridan's Principle has appeared as a fundamental problem in computer design. In computing, a device that makes a discrete (usually binary) decision based upon a continuous input value is called an *arbiter*, and Buridan's Principle is usually known as the Arbiter Problem [1].

The traditional manifestation of the Arbiter Problem is in the design of a computer's interrupt-handling mechanism. A computer interacts with a peripheral device through interrupts; the device signals that it needs attention by setting an interrupt flag, which the computer tests during normal execution. When the computer sees that the flag is set, it interrupts normal processing and executes a special program to service the device and reset the interrupt flag. On every instruction, the computer must make the binary decision of whether or not the interrupt flag has been set.

While the interrupt flag is being set, its state is a continuous function of the time at which the device began setting it. If, as is usually the case, the peripheral device's setting of the flag is not synchronized with the computer's execution, then the computer's binary decision is based upon an input having a continuous range of values. Buridan's Principle asserts that the decision cannot be made in a bounded length of time. However, the computer must make that decision before beginning its next instruction, which generally happens in a fixed length of time.

The computer is thus trying to do something that is impossible. Just as the driver at the railroad crossing has a finite probability of being hit by the train, the computer has a finite probability of not making its decision in time. The physical manifestation of the computer's indecision is that bad voltage levels are propagated. For example, if a 0 is represented by a zero voltage and a 1 is represented by +5 volts, then some wire might have a level of 2.5 volts. This leads to errors, because a 2.5 volt level could be interpreted as a 0 by some circuits and a 1 by others. The computer stops acting like a digital device and starts acting like a continuous (analog) one, with unpredictable results.

The Arbiter Problem is a classic example of Buridan's Principle. The problem is not one of making the "right" decision, since it makes little difference if the interrupt is handled after the current instruction or after the following one; the problem is simply making a decision. The Arbiter Problem went unrecognized for a number of years because engineers did not believe that their binary circuit elements could ever produce "1/2's". The problem is solved in modern computers by allowing enough time for deciding so the probability of not reaching a decision soon enough is much smaller than the probability of other types of failure. For example, rather than deciding whether to interrupt execution after the current instruction, the computer can decide whether to interrupt it after the third succeeding instruction. With proper circuit design, the probability of not having reached a decision by time t is an exponentially decreasing function of t, so allowing a little extra time for the decision can make the probability of failure negligible.

Another possible solution to the Arbiter Problem is to allow an unbounded length of time for the decision. The computer simply waits until it makes the decision before starting its next instruction. With properly designed circuits, the expected time to reach the decision is small, so this would not appreciably slow down the computer. However, most computers are designed so the start of the next instruction is controlled by a pulse from a continuously-running clock. There are sound engineering reasons for such a design. For example, if the computer were to wait for the decision, then loss through component malfunction of the single pulse signaling the decision would hang up the computer.

Buridan's Principle might lead one to suspect that a digital computer is an impossibility, since every step in its execution requires making discrete decisions within a fixed length of time. However, those decisions are normally based upon a discontinuous set of inputs. Whenever the value of a memory register is tested, each bit will be represented by a voltage whose value lies within two separate ranges—the range of values representing a zero or the range representing a one. Intermediate voltages are not possible because the register is never examined while it is in an intermediate state—for example, while a bit is changing from zero to one. The Arbiter Problem arises when the computer must interact *asynchronously* with an external device, since synchronization is required to prevent the computer from seeing an intermediate voltage level by reading a bit while the device is changing it. A similar problem occurs in interactions between the computer and its environment that require analog to digital conversion, such as video input.

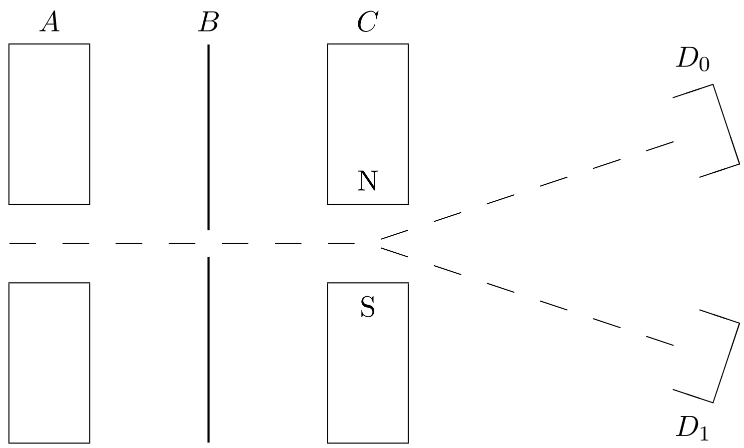

Figure 3: A quantum-mechanical arbiter.

6 Continuity

The validity of Buridan's Principle rests upon the assumption of continuity: the state $A_t(x)$ at time t of a physical system started in state x is a continuous function of x. Is this a correct assumption about the physical world? If the system is described by a set of differential equations satisfying certain reasonable conditions, then the assumption follows from known mathematical results [6]. If the system is described by partial differential equations, we know of no similar general result. However, the assumption that the behavior of a system is a continuous function of its initial conditions has generally been taken by physicists to be a requirement for a meaningful theory.

It might be argued that the assumption of continuity is no longer valid in quantum mechanics, since quantum mechanics predicts the existence of values, like the angular momentum of the electron, whose measurement must yield a discrete result. The problem of measurement in quantum mechanics is a puzzling one, since the act of measurement, which takes a system from a linear combination of eigenstates to a pure eigenstate (the "reduction of the wave packet"), cannot be described quantum mechanically. (See [7] and [12] for a discussion of the measurement problem in quantum mechanics.) I believe that discrete eignenvalues do not permit one to circumvent the principle. While I cannot give any general argument to support this belief, an analysis of a particular "ass" should indicate why it is plausible.

Consider the arbiter, shown in Figure 3, that is based upon a modified Stern-Gerlach apparatus. A and C are magnets, where the direction of magnetization of A is controlled by the interrupt flag and that of C is fixed, B is an accelerating device, and D_0 and D_1 are detectors. The dashed lines show the two possible paths of an electron through the device.

The value of the interrupt flag is read by detecting the direction of magnetization of magnet A. An electron is inserted into the apparatus from the left; its spin is set by magnet A and measured by the rest of the apparatus—device B accelerating it to the right and magnet C deflecting it up or down, depending upon whether its spin is up or down, into detector D_0 or D_1. Magnet A is presumed strong enough so that if the interrupt flag has a stable value, not in the process of changing, then the electron will be given the correct spin by A and will strike the appropriate detector. If the flag is in the process of changing while the electron is passing through it, so we don't care what value is read, we cannot predict the spin direction of the electron after it has passed through magnet A. However, because a measurement of the electron's spin must give either *up* or *down*, the electron will strike one of the two detectors. Thus, this device seems to contradict Buridan's Principle; we can measure the state of the flag in a bounded length of time, getting either a zero or a one, by simply inserting an electron.

The reason this device does not work is rather subtle. To read the flag in a bounded length of time, we must constrain the horizontal position of the electron to a bounded region of space when we insert it into the device. By the Uncertainty Principle, this introduces an uncertainty in its horizontal momentum. More precisely, since the momentum wave function is the Fourier transform of the position wave function, bounding its horizontal position implies that the statistical distribution of its horizontal momentum must be unbounded. Thus, there is a finite probability of the electron moving very fast to the right—so fast that it is not deflected enough by magnet C to hit either of the detectors, instead going between them.

A real Stern-Gerlach apparatus does not produce the discrete statistical distribution of electron trajectories usually ascribed to it in simplified descriptions. Instead, it produces a continuous distribution having two maxima, but with a nonzero probability of finding an electron in any finite region between them. Trying to decide if the electron is deflected up or down then becomes just another instance of the problem of making a discrete decision based upon a continuous input value, so nothing has been gained by measuring the discrete spin value.

Validity of Buridan's Principle implies the following:

> *Buridan's Law of Measurement.* If $x < y < z$, then any measurement performed in a bounded length of time that has a nonzero

probability of yielding a value in a neighborhood of x and a nonzero probability of yielding a value in a neighborhood of z must also have a nonzero probability of yielding a value in a neighborhood of y.

If this law is not valid, then one can find a counterexample to Buridan's Principle, with the discrete decision being: "Is the value greater or less than y?"

7 The Meaning of Buridan's Principle

Buridan's Principle rests upon mathematical concepts of continuity and boundedness that are not physically observable. No real experiment, having finite precision, can demonstrate the presence or absence of continuity, which is defined in terms of limits. No experiment can demonstrate that an arbiter requires an unbounded length of time to reach a decision. An experiment in which the arbiter failed to decide within a week does not prove that it would not always decide within a year.

To understand the meaning of Buridan's Principle as a scientific law, consider the analogous problem with classical mechanics. Kepler's first law states that the orbit of a planet is an ellipse. This is not experimentally verifiable because any finite-precision measurement of the orbit is consistent with an infinite number of mathematical curves. In practice, what we can deduce from Kepler's law is that measurement of the orbit will, to a good approximation, be consistent with the predicted ellipse.

Similarly, we cannot conclude from Buridan's Principle that the decision will take an unbounded time. What we can conclude is that it may take a very long time to make the decision. A cleverly designed testing circuit that locates "indecision points" can cause an arbiter that normally decides within a microsecond to hang up for milliseconds before deciding.

The significance of Buridan's Principle lies in its warning that decisions may, in rare circumstances, take much longer than expected. Before the problem was recognized by computer designers, some computer systems probably failed regularly (perhaps once or twice a week) because arbiters took longer than expected to reach a decision. Real accidents may occur because people cannot decide in time which of two alternative actions to take, even though either would prevent the accident. Although Buridan's Principle implies that the possibility of such an accident cannot be eliminated, awareness of the problem could lead to methods for reducing its probability.

Buridan's Principle appears to be a universal law of nature—perhaps a fundamental law not derivable from others, similar in that respect to the second law of thermodynamics. It manifests itself not only in the world of computers, but in such mundane actions as driving.

8 More Recent Developments

Modern interest in Buridan's Principle stems from the arbiter problem, its manifestation in computers. There is a long history of computer scientists being unaware of the problem—and even denying its existence. The original version of this paper, almost identical to the preceding seven sections, was written in 1984. Its history, as well as some amusing anecdotes about the problem, appear in this paper's entry in the author's publications Web page [8]. The arbiter problem is now well known, in part because of interest in self-timed (clockless) computers [10], and it has been subject to rigorous mathematical analysis [2]. However, it still leads to errors [5].

In 1984, there did not seem to be a quantum-mechanical theory of measurement from which Buridan's Principle could be derived. More recent developments in quantum mechanics may make such a derivation possible [3]. A careful analysis of the Stern-Gerlach apparatus in [3] confirms the conclusion of Section 6 that it does not contradict Buridan's Principle.

References

[1] J. C. Barros and B. W. Johnson. Equivalence of the arbiter, the synchronizer, the latch, and the inertial delay. *IEEE Transactions on Computers*, C-32(7):603–614, July 1983.

[2] Michael S. Branicky. Universal computation and other capabilities of hybrid and continuous dynamical systems. *Theoretical Computer Science*, 138(1):67–100, February 1995.

[3] Paul Busch, Marian Grabowski, and Pekka J. Lahti. *Operational Quantum Physics*, volume m31 of *Lecture Notes in Physics*. Springer, Berlin, 1995.

[4] Paul Edwards, editor. *The Encyclopedia of Philosophy*, volume 1. The Macmillan Company & The Free Press, New York, 1973.

[5] Ran Ginosar. Fourteen ways to fool your synchronizer. In *9th International Symposium on Advanced Research in Asynchronous Circuits and*

Systems (ASYNC 2003), pages 89–97, Vancouver, BC, Canada, May 2003.

[6] J. Hale. *Functional Differential Equations*. Springer-Verlag, New York, 1971.

[7] Max Jammer. *The Philosophy of Quantum Mechanics: the Interpretations of Quantum Mechanics in Historical Perspective*. John Wiley and Sons, Inc., New York, 1974.

[8] Leslie Lamport. My writings. `http://research.microsoft.com/users/lamport/pubs/pubs.html`, can be found by searching the Web for the 23-letter string obtained by removing the - characters from all-lamports-pubs-onthe-web.

[9] James R. Munkres. *Topology: A First Course*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1975.

[10] Ivan E. Sutherland. Computers without clocks. *Scientific American*, 287(2):62–69, August 2002.

[11] Douglas Vickers. Discrimination. In A. T. Welford, editor, *Reaction Times*, pages 25–72, London, 1980. Academic Press.

[12] John A. Wheeler and W. H. Zurek, editors. *Quantum Theory and Measurement*. Princeton University Press, Princeton, New Jersey, 1982.