Behavior: The Control of Perception

What is the *definition* of behavior? What is the difference between human and computer behavior, or between natural and artificial intelligence? What limits a computer's ability to reason, plan, learn, and communicate? Could we design a computer which duplicates a human's behavior? If so, its necessary for us to conceptually understand the architecture of the human brain as well as the fundamental difference between the brain and the computer.

First, let's understand the relation between a computer's mistakes and our own. Assume the following: a computer is only supposed to execute instructions. A computer's "mistake" would therefore be directly related to a human programmer's error, because the programmer is the one supplying instructions to the computer. Therefore, the only reason a computer would logically err independent of its programmer is if its physical components were faulty. For instance, if the Arithmetic Logic Unit within the CPU were incorrectly wired, the computer will not perform arithmetic correctly and thus 1 + 1 will not equal 2, clearly a logical fallacy. For completeness, we'll mention that even with impeccable wiring, there's still a (vanishingly small) probability that a computer might make such an arithmetic error, but that is of no real concern.

Turning the attention to human errors, we now ask: what are the different types of errors *we* make? Well, for one, humans make tons of math errors. We could discuss the various reasons why humans make mistakes in their arithmetic, but let's not - let's just assume that humans shouldn't be negligent when it comes to math. Another type of error we make is a syntactical error. Syntax, or grammar, is the study of the principles and rules by which sentences are constructed in particular languages or mathematical systems. It's important to be extremely familiar with the rules of syntax, because using improper grammar leads to ambiguation of meaning and thus improper articulation of a communicated message. A third type of error involves perceptions. Perceptions are obviously fundamental to the way our brain works, and by forming perceptions, we learn, we understand, and we communicate through language.

Rumor has it, that a fundamental limitation exists within the English language - a limitation which impairs our ability to communicate abstract concepts, such as Justice, Liberty, Equality, etc. One may reasonably posit, that if different people are making different associations to the *same* words, these words will mean different things to different people. On this basis, one might logically argue against the possibility of developing universal systems of governance. In other words, the claim is that it's impossible to reach a consensus on how to carry out the design of a "perfect" society, or a *Utopia*, because we would never be able to unanimously agree on the abstract framework.

So, there's two ways of looking at this: either we've figured everything out and we know that it's impossible, or we haven't yet figured everything out, and that's why we think it's impossible. But how can we know what we don't know? Is there an absolute limit regarding what we can figure out? We've now arrived at the front porch of the study of epistemology, the branch of philosophy concerned with the nature and scope of knowledge. The central aim of this study is to relate the notions of belief, truth, and justification, in order to establish what can acceptably be regarded as justified true belief - knowledge.

Let's consider the following question: what does it mean that a belief is justified? Suppose I step in front of a light source and observe a shadow being cast behind me - should I believe in the existence of my shadow? Aristotle related the concept of existence to causality; let's examine this association. According to this view, the existence of my shadow is manifest in its effects. So suppose I used my body to block sunlight from reaching the ground, thereby causing a shadow to appear. Clearly, what we are observing as my shadow would be the effect

of my body's positioning, rather than vice versa. What may perhaps be inferred, then, is that it's more accurate to speak of an effect dependent on the position of my body, versus an effect dependent on the appearance of my shadow; though I perceive my shadow, its appearance alone should not be considered as existence. If Niels Bohr were present, he'd say: "How wonderful that we've met with a paradox; now, we have hope of making some progress". Considering that paradoxes make good riddles, see if you can answer these important questions: What is it which *can* be perceived to exist, yet should *not* be observed? Alternatively, What is it which *cannot* be perceived, yet whose existence *must* be observed? I have no clue whatsoever.

Anyways, supposing that knowledge is justified belief, the next thing to consider is the question: what is knowledge used for? Knowledge is used for responding. What connection does knowledge have with intelligence? In many AI textbooks, intelligent agents are considered to take actions which maximize their chances of success. Intelligence would therefore be related to knowledge of an "appropriately successful" response. We should like to know then, what determines the appropriateness or successfulness of a response? Logically, success depends on the relation of past and future events. Considering that knowledge of cause and effect is the key to understanding this relation, perhaps intelligence may be understood as knowledge of causality.

Causality is the relation between an event (the cause) and a second event (the effect), where the second event is understood as a consequence of the first. So, imagine we're observing some behavior or phenomenon and trying to explain it's occurrence by pointing to causes. The scientific strategy is to falsify hypotheses which are being tested under controlled environmental conditions. Now if we're trying to establish the causes of some phenomenon via the scientific method, we should do well to notice that we're implicitly controlling the phenomenon's occurrence (i.e. because we're trying to cause it on purpose). Intelligence, knowledge, and causality can therefore be seen as related through the concept of control, where control is defined as the power to influence or direct the course of future events.

Behold a thermostat. A thermostat is used to maintain the temperature of a room by stabilizing it at a reference value or set point. We may therefore surmise that the purpose of its behavior is to control the temperature; but how would we differentiate between the *purpose* and the *cause* of its behavior? The definition of a 'cause' is:
(1) a person or thing that gives rise to an action, phenomenon, or condition;
(2) a principle, aim, or movement that one is prepared to defend or advocate.

In contrast, a 'purpose' is defined as:
(1) the reason for which something is done or created or for which something exists.
(2) one's intention or objective.

Now, it seems to me that the difference between a cause and a purpose is not quite clear - in fact, the two definitions are arguably identical. Upon critical examination of the definition of a 'purpose', one might further argue that the first entry is a bit vague.

Perhaps an example will help us understand the ambiguity. Suppose you order someone to do something: you say, "drop to the floor". He responds, "give me a reason". Now, assuming you aren't going to simply respond, "because I said so", how would you explain the reason why? A 'reason' is defined as:
(1) a **cause**, grounds, basis, rationale;
(2) a **purpose,** motive, motivation, point, aim, objective, intention, goal;
(3) an **explanation**, justification, argument, defense, vindication, excuse, pretext.

Clearly, the choice of reason might involve a cause, a purpose, or an explanation of both.

So, what is the reason an apple falls from a tree? To give a mathematical explanation of its behavior, we would invoke the laws of physics: we say, the gravitational potential energy of the apple is minimized as it falls, thus providing an understandable reason for the apple to fall.

But instead of passively waiting for an apple to gravitate, suppose you *commanded* the apple to fall. If the apple, for some reason or another, obeyed your command and fell, would the sight of the falling apple immediately make you happy? or would you wonder: why did the apple not question my order? If, instead, the apple steadfastly refused to obey your instruction without provision of a reason, perhaps you'd pugnaciously retort, "How *dare* you question my order, apple?! Off with your skin!!" Supposing that coercion was not an option, however, would you need to reason and argue with the apple until it believed a certain truth value, at which point you could only hope that it would logically fall of its own accord? If so, what reason would you provide? Mathematically, we know the apple seeks equilibrium. Maybe, you could try to explain to the apple that equilibrium is not on the branch. But, then again, maybe the apple would insist on hanging because it perceives a different equilibrium position.

In reality, of course, an apple doesn't perceive anything - it doesn't hear arguments, nor does it respond to commands - it just feels forces. But let's imagine we're talking about a *special* apple, which, in addition to feeling the forces upon it, also harbored its own intention. If we needed to figure out the intention of this apple, how could we do it? First, suppose that the apple did not actually have an intention. Since this apple has no intention, if it were to fall, there would be no purpose to its behavior. Even if we observed that when it fell, it knocked over a stack of dominoes, which in turn flipped a switch, which turned on a computer, there would still be no purpose to its behavior. Now, whoever put the dominoes there must have known the apple would be arriving. We could ask: who is this man, and what does he want? But, before concerning ourselves with the wants and desires of men, let's first guess the intentions of apples.

So assuming an apple has a certain intention in mind, how could we ascertain its intention? Consider, how would the apple respond if our actions were to negatively interfere with what it intended? Imagine an apple, at rest, intending to remain stationary. We push it and it starts sliding across the floor. Now, if the apple does *not* intend to remain stationary, we should expect it to keep sliding until friction eventually halts it motion. But, if the apple *does* intend to remain still, we might expect that it would bring itself to rest independently of friction. Of course, apples don't usually do this sort of thing - i.e. they don't behave intentionally. This raises the following question: if someone or something is doing something intentionally, then what's 'causing' their behavior? Clearly, their behavior stems from some desire to achieve an intention.

Let's take a moment to ponder, then, how intentions might originate. To address this question, let's return consideration to our old friend, the thermostat. So, the purpose of the thermostat's behavior is to control the temperature. *Why?* Because the thermostat maintains the temperature at a reference value. *How?* Inside the thermostat is a mechanism which exists at equilibrium only when the temperature is at a reference value, such that if the temperature were *not* at this reference value, then the mechanism would generate a response which brings the temperature back to this value. Thus it makes sense to say that the purpose of the thermostat's behavior is to control the temperature. *But what causes the thermostat to do so?*

Imagine a thermostat in a room, with the room temperature matching its reference value. When the temperature falls, an electronic circuit generates the thermostat's response. How might the behavior of charged particles moving through an electronic circuit be correlated to the purpose of the thermostat's behavior? To aid us in this inquiry, let's imagine an epic dialogue between two different types of controllers: one is a thermostat and the other is an autocrat.

Epic Dialogue

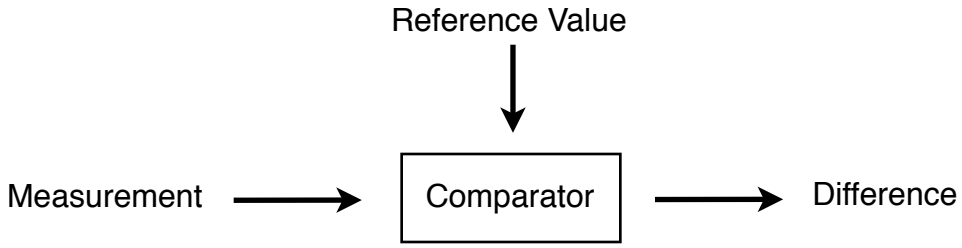
King: Squire, are you aware of the purpose of your behavior?

Thermostat: Indeed Sire, the purpose of my behavior is to control the temperature of the air.

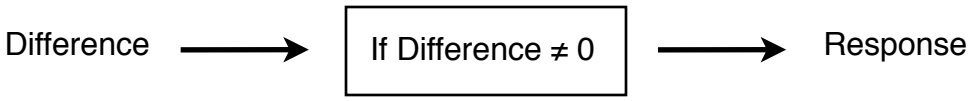
King: Very good. And, how do you go about doing so?

Thermostat: Um...well, the first thing I always do is measure the temperature value...

...and after I measure the temperature value, I compare this measured value to my reference value...

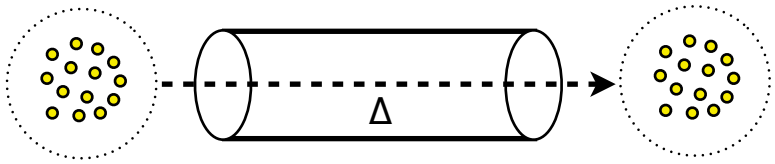


...and if the temperature is not what its supposed to be, I respond.



King: So, could you explain to me then, what causes your behavior?

Thermostat: Well, I know that my actions are being generated by signals composed of electrical currents, which are being generated by the displacement of electronic particles through a wire.



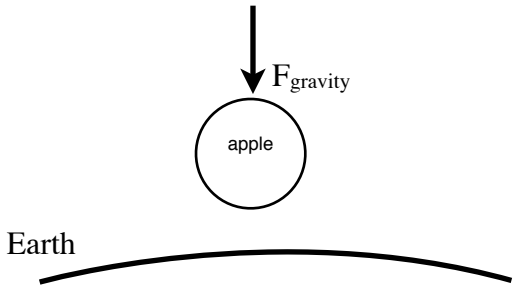
I suppose the causes of my behavior, then, should be related to the causes of these particular electron-displacement currents.

King: And if these electronic currents caused your actions, then would you also suppose that the temperature is controlled by the behavior of these electrons passing through your wires?

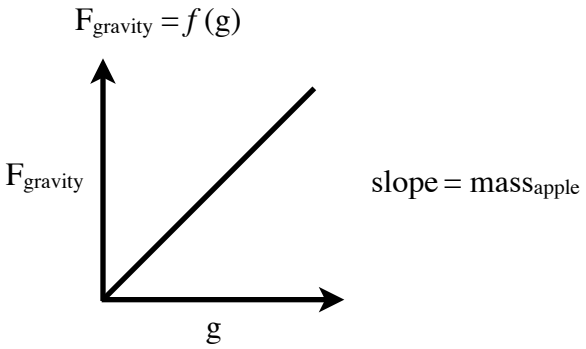
Thermostat: Well I don't really see how *their* behavior could actually control the temperature, because these electrons are not really aware of the temperature value - at least not in the same sense that *I* am aware. I mean, if one can say that an electron is "aware" of anything, I suppose it is only aware of the forces that act on it, at the place where it is located at any instant.

King: That's very interesting, Squire, because in many respects, the behavior of an electron is very similar to the behavior of an apple falling from a tree. You see, a falling apple is

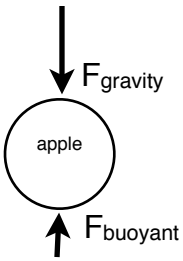
also “aware” only of the forces in its immediate vicinity. For instance, it feels the force of the Earth’s gravity.



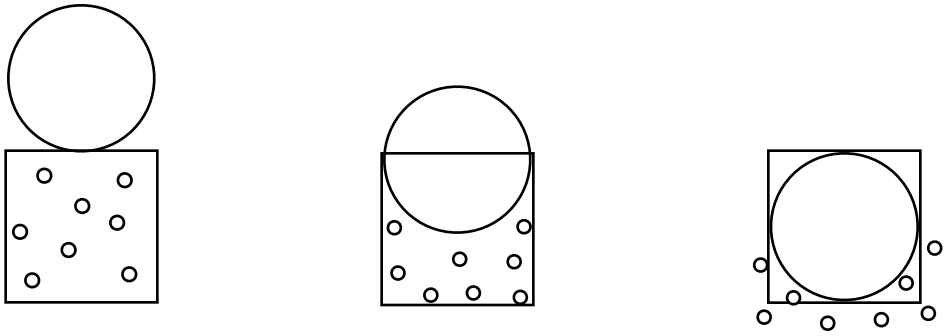
And if gravity were the only force acting on the apple, then the force which the apple experiences will be a function of a single variable, *g*, which refers to the acceleration that the Earth imparts to the apple.



Generally, however, in addition to the gravitational force, a falling apple will also feel a force in the *opposite* direction of gravity, called the buoyant force.



This force, which is a result of the air pressure along the apple’s path, is related to the motion of the apple falling through the air. Now, in case the air is initially motionless, the falling apple will cause the air in its path to move with a speed related to the speed of the falling apple itself - the faster the apple is falling, the greater the speed with which the air is made to move out of the space being occupied by the apple.



So consider, if the motion of the apple through the air causes the motion of the air itself, and the motion of the air affects the pressure along the apple's path - which, in turn, influences the motion of the apple via the buoyant force - then the motion of the apple is automatically effecting a change to itself via the buoyant force. See, whenever there's a difference between gravity and buoyancy, this results in a net force on the apple.

$$F_{net} = F_{gravity} - F_{buoyant}$$

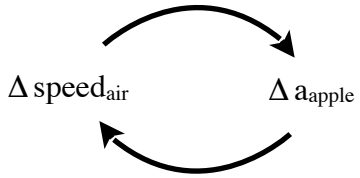
This causes the apple to accelerate with a magnitude inversely proportional to its mass.

$$a_{apple} = F_{net} / m_{apple}$$

This acceleration defines an immediate change in the apple's velocity, but also translates into an effect on the speed of the air moving immediately along the apple's path.

$$a_{apple} = \Delta v_{apple} \rightarrow \Delta speed_{air}$$

This change of air speed causes the pressure being felt on the surface of the apple to change, which generates a different value of the buoyant force, and a different acceleration.



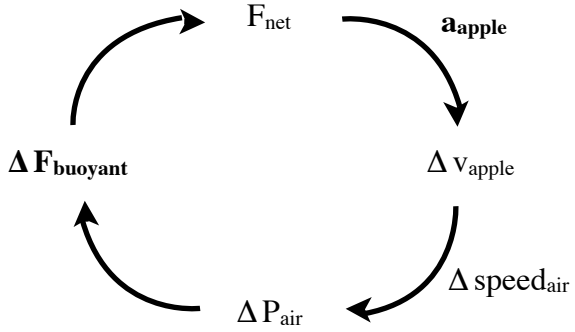
Accordingly, with the effect of the buoyant force causing a change in the motion of the apple, the effect of which is causing a change in the buoyant force itself, a complete description of this force requires us to incorporate a circular model of causality known as closed-loop feedback. Again, this need arises because the motion of the apple through the air generates a changing effect on its own motion. Are you following me so far, Squire?

Thermostat: Indubitably, Sire, I shall relate to you what you have described: the apple's motion causes the displacement of the air, generating the buoyant force; the feedback emerges because this force changes the apple's rate of motion, which causes a different amount of air displacement, generating a different amount of buoyant force, and therefore a different change to the apple's rate of motion; and the process thus repeats because a change in the buoyant force causes a change in the acceleration of the apple, and a change in the acceleration of the apple causes a change in the buoyant force.

King: Majestic Squire, you are a splendiferous hear-sayer, indeed!

Thermostat: Bless you, my Liege, you have generously provided me with invaluable insight concerning the behavior of falling apples.

King: Kindly think naught of it, good Squire. Now let us resume our analysis. Thus far, we have established that a falling apple experiences a continuously changing buoyant force, and is consequently subject to a continuously changing acceleration. So, we have defined two changing variables - acceleration and buoyancy - which are mathematically related through closed-loop feedback.



Logically, a change in either of these values will establish a self-propagating condition that persists until the buoyant force assumes the value of the gravitational force. Thus, as long as the force of gravity is greater than the buoyant force, the apple will continue to experience a changing acceleration and buoyancy.

$$F_{net} \neq 0 \text{ -> } \Delta F_{buoyant} \neq 0 \text{ -> } \Delta F_{net} \neq 0 \text{ -> } \Delta a_{apple} \neq 0$$

As soon as these two forces acting on the apple equilibrate, however, they will do so constantly - and the apple will no longer accelerate.

$$F_{net} = 0 \text{ -> } a_{apple} = 0 \text{ -> } \Delta F_{buoyant} = 0 \text{ -> } \Delta F_{net} = 0$$

Now, since the acceleration of the apple equilibrates to zero, it follows that the changing buoyant force minimizes the changing acceleration of the apple. That is, the changing acceleration of the apple establishes a damping or *negative* feedback effect on itself via the changing buoyant force. But when we trace the chain of causalities in reverse, starting from a change in the apple's acceleration, we may derive the following relations:

the changing acceleration depends on the changing buoyant force

$$\Delta a_{apple} = f(\Delta F_{buoyant})$$

the changing buoyant force depends on the changing air pressure

$$\Delta F_{buoyant} = f(\Delta P_{air})$$

the changing air pressure depends on the changing speed of the air

$$\Delta P_{air} = f(\Delta speed_{air})$$

the changing speed of the air depends on the changing velocity of the apple

$$\Delta speed_{air} = f(\Delta v_{apple})$$

the changing velocity of the apple depends on its acceleration

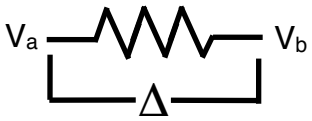
$$\Delta v_{apple} = f(a_{apple})$$

We may thus see that, whereas the apple's changing velocity is a function of its acceleration, and the apple's *changing* acceleration is a function of the changing buoyant force, *there is no explicit function of the apple's changing acceleration*.

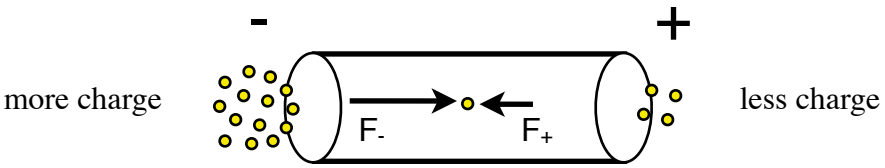
Thermostat: Sire, what an esoteric mention. What significance does this entail?

King: It signifies, Squire, that the value of the apple's changing acceleration does not immediately depend on the value of the changing buoyant force in the same manner as the value of the changing buoyant force immediately depends on the value of the apple's changing acceleration. In a sense, therefore, the changing acceleration can be understood as a side-effect of the changing buoyant force. Now tell me, Squire, perhaps you fancy a discussion of axiomatic set theory and the foundations of mathematics?

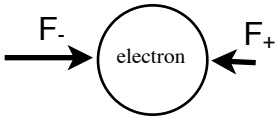
Thermostat: Not really.
King: Very well then, we shall return our attention to the issue of the temperature.
Thermostat: I shall have no qualms.
King: Now, you say you are to control the temperature value, correct?
Thermostat: Yes.
King: So, if the temperature were to drop below your reference value, you would need to generate heat in order to raise the temperature to your reference value, correct?
Thermostat: Indeed.
King: It is known that heat may be generated by an electronic resistor connected across a voltage difference.



This voltage difference results from a difference of charge density across the two ends of a wire generating an imbalance of forces on the electron and causing it to move down the wire.



Now, the end result of the electron's behavior would simply be to balance this difference of charge - similar to the manner in which the apple balances the buoyant force against gravity.



But a complete treatment would involve a discussion of what is known as the Lorentz force, which is the effect of the accelerating electron's own electromagnetic field on its motion. However, these forces do not concern us right now. The main point we are concerned with here is to understand the distinction between the *cause* of the electron's behavior and the *purpose* of your behavior.

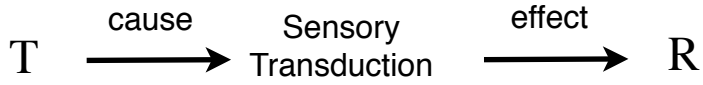
Thermostat: It seems you wish to distinguish, then, between behaviors which result from 'causes' and those which result from 'purposes'.
King: That is correct, Squire.
Thermostat: If I may, I would like to advance my humble opinion.
King: Please do so.
Thermostat: It seems to me, for reasons that are not yet completely understood, that perhaps the *only* thing which may properly be understood as a 'purpose' is something which would cause a particular perception, rather than something which would cause a particular response. In lieu of this consideration, however, I notice that it would be intuitively difficult for an observer to establish the purpose of a behavior, if only because the intended perception must be observed from the point of view of the behavior itself.

King: That is interesting, Squire. Would you elaborate on your reasoning?

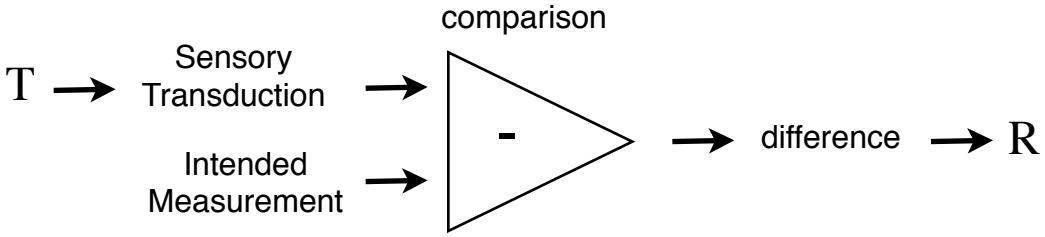
Thermostat: Of course. So, imagine if a cold breeze were to cause a drop in the temperature and my response were to then occur. To an observer, I notice, it would appear that a temperature change was causing my response.

King: Yes, it would look like this environmental variable was somehow directly related to your response.

Thermostat: Indeed, but this would merely be an illusion; the environment would only *appear* to be causing my response. In reality, my response to the temperature would not be mechanistically caused by a measurement of its value.



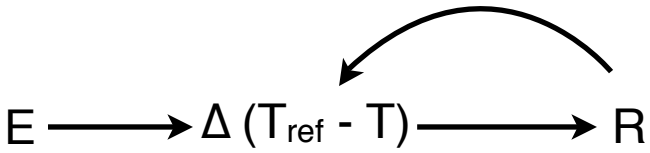
Rather, my response would have occurred only because the environmental temperature did not match my intended measurement.



As such, my actions would be properly understood only if they were viewed as being caused by the changing state of a *difference* between my perception and my reference value. Now, the reason for this is not complicated. It's quite simply because my responses have an effect on the temperature value.

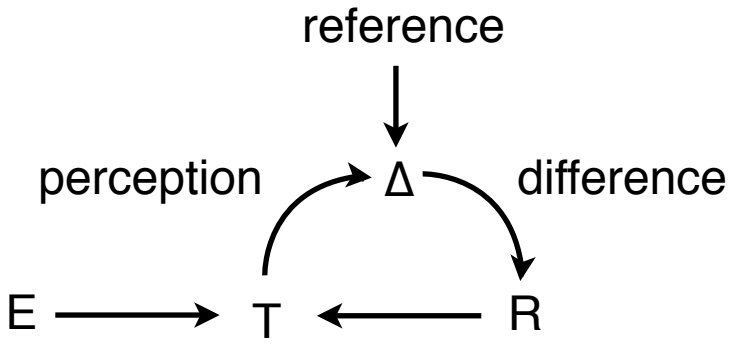
And so, as a result of the influence of my own actions upon my perceptual state, we are forced to dismiss the impression that my actions are being entirely caused by an environmental process.

King: Hmm. So let me hear you straight: the cause of your response is the *difference* between the temperature and your reference value, which is being influenced simultaneously by the environment and by your response.



And since your response alters this difference, the cause of this response at any given moment is the result of a circular law of cause and effect. To state it forwardly, with the observable consequences of your actions directly affecting the perceptual variable being

compared to your reference, the perceptual effects of your actions are feeding back onto their perceptual causes.



As a result, the temperature value which you perceive at any given moment is fundamentally a controlled perceptual input.

Thermostat: My Lord, you are precisely correct! My behavior is *fundamentally* a controlled perception, the purposive result of which is to immediately minimize a perceptual difference with respect to a defined - yet intrinsically hidden - reference value.

King: This is true, Squire. In fact, we may very well be logically forced to define the purpose of *all* behavior as the control of perception. But before anything else, allow me to point out to you the significance of this model of behavior. The operating mechanism of your response actually provides a model for the quantum mechanical wave function collapse.

Thermostat: My God, Sire, that sounds like a bodacious claim! What exactly are you saying?

King: Well, you see, there's this interesting concept in quantum mechanics known as "the observer effect," which refers to changes which the act of observation will make on a phenomenon being observed. Evidence from experiment has led many a scientist to the realization that what we perceive as reality in the present necessarily depends on our earlier decision of what to measure. Therefore, it's absolutely wrong to assume the observable features of a system exist prior to our measurement. So, consider now that you've decided to measure a particular temperature at a certain time in the past by defining your reference value; this decision would alter the difference causing your response. What you would perceive as the temperature at a future time would therefore depend on what you've decided to measure at an earlier time. As a result, by strategically defining your behavior as the control of perception, it is in fact possible to meaningfully interpret your current perceptual state as the result of a previous action which was dependent on your earlier choice of reference value - which would have been your decision of what to measure.

End of dialogue