RIFL (2020) Vol. 14, n. 2: 47-62
DOI: 10.4396/AISB201905

Glanville's 'Black Box': what can an observer know?

Lance Nizami
Independent Research Scholar
nizamii2@att.net

Abstract A 'Black Box' cannot be opened to reveal its mechanism. Rather, its operations are inferred through input from (and output to) an 'observer'. All of us are observers, who attempt to understand the Black Boxes that are Minds. The Black Box and its observer constitute a system, differing from either component alone: a 'greater' Black Box to any further-external-observer. To Glanville (1982), the further-external-observer probes the greater-Black-Box by interacting directly with its core Black Box, ignoring that Box's immediate observer. In later accounts, however, Glanville's greater-Black-Box inexplicably becomes unitary. Why the discrepancy? To resolve it, we start with von Foerster's archetype 'machines', that are of two kinds: 'Trivial' (predictable) and 'Non-Trivial' (non-predictable). Early-on, Glanville treated the core Black Box and its observer as Trivial Machines, that gradually 'whiten' (reveal) each other though input and output, becoming 'white boxes'. Later, however, Black Box and observer became Non-Trivial Machines, never fully 'whitenable'. But Non-Trivial Machines can be concatenated from Trivial Machines, and are the only true Black Boxes; any greater-Black-Box (Non-Trivial Machine) may (within its core Black Box) involve white boxes (that are Trivial Machines). White boxes, therefore, could be the ultimate source of the greatest Black Box of all: the Mind.

Keywords: Glanville, Black Box, observer, mind, machine, von Foerster

Received 28/02/2020; accepted 14/03/2020.

0. Introduction

One dream of Artificial Intelligence is to exactly mimic natural intelligence. Natural intelligence is examined by observing the behaviors that imply a *mind*. Minds presumably exist within animals having recognizable brains. (Whether other species have minds will not be debated.)

Of course, behavior can be difficult to quantify, especially when experimental research subjects cannot 'report'. An example of reporting is the confirming of particular sensations evoked by stimuli (Nizami 2018, 2019a, 2019b). In animals, primitive reporting (Yes/No, Left/Right, etc.) can be painstakingly conditioned. But conditioning may not be feasible (or ethical) for human infants, for example. Nonetheless, we desire to establish infants' sensory abilities, in order to detect defects. By-and-large, however, the animal or the infant remains a 'Black Box' to the observer of behavior. The reason for the capital B's will soon be explained.

Here, the attempts to understand a mind through interaction with its host organism are placed in relation to the notions of the Black Box and its 'observer' as proselytized by Ranulph Glanville (1982, 1997, 2007, 2009a, 2009b). Glanville was a Professor of Design and a champion of Second-Order Cybernetics. Glanville (2007) notes that much of his discourse on the Black Box originated in the writings of W. Ross Ashby. Hence, we begin with Ashby. Ashby devotes a chapter to the Black Box in his book *An Introduction to Cybernetics* (1956), a book cited over 13,000 times (GoogleScholar) which, by present-day standards, makes it a "classic". (For a brief summary of Ashby's importance to science, see Ramage, Shipp 2009.) Ashby's 1961 edition is more readily available, and is cited here (Ashby 1961).

1. The Black Box and its observer

The 'black box' of an engineer or a physicist is a physical object that can be opened, allowing its operation to be comprehended. The 'black box' need not be hardware; it can be software, a computer program whose operation has, thanks to unanticipated effects of the code, become a 'black box' to its own creators (Holm 2019; Rahwan *et al.* 2019). Such effects have recently been named 'machine behavior', exhibited by 'artificial intelligence'. But the respective code can still be read by its makers, and potentially by others. 'Intelligence' is independent; software is clearly not.

The present interpretations of 'intelligence' and 'behavior' will be far less gratuitous. Particularly, 'behavior' will be taken as something that is done consciously and willfully (intentionally) by a living thing. This contrasts to 'reflex' behavior, which would include, for example, the jerk of the lower leg when the knee is tapped by a physician, or the tendency of some single-celled organisms to move towards light. Likewise, the much-touted 'behaviors' of Grey Walter's 'tortoises' (Walter 1950; Holland 2003) are merely reflexes.

If the 'machine' of the engineer is un-openable (or the programming by the computer scientist is unreadable), the black box becomes a Black Box (Ashby 1961). It is understood only through inputs given by, and outputs noted by, an observer. Indeed, the input/output cycle may never reveal the Black Box's mechanism; a *mechanical* basis, for example, may be indistinguishable from an *electrical* one. Ashby gives examples, noting that we can «Cover the central parts of the mechanism and the two machines are indistinguishable throughout an infinite number of tests applied. Machines can thus show the profoundest similarities in behavior while being, from other points of view, utterly dissimilar» (Ashby 1961: 96).

Following Ashby, we might imagine machines that consist of both mechanical *and* electrical components, *mechanoelectrical* 'systems' whose actual mechanisms are indistinguishable, one from another, through input and output. As such, the mechanisms become irrelevant. Glanville takes this logic to its limit: «You cannot see inside the Black Box (there is nothing to see: there is nothing there—it is an *explanatory principle*)» (Glanville 1997, II; italics added). That is, «Our Black Box is not a physical object, but a concept … It has no substance, and so can neither be opened, nor does it have an inside» (Glanville 2009b: 154). Even so, Glanville states that it has a *mechanism* (Glanville 1982, 2007, 2009a, 2009b).

Glanville's Black Box may sound suspiciously like a *mind.* After all, no-one can directly observe their own mind, or anybody/anything else's; *'mind'* is an *explanatory principle* for what we call 'behavior'. Glanville's work therefore deserves further scrutiny. Unfortunately, his principal exposition (Glanville 1982) requires clarification, as will be explained. Glanville later attempts clarification (Glanville 1997, 2007, 2009a, 2009b), but falls short. The present paper provides the missing details. Provocative insights emerge.

2. 'Whitening' the Black Box

Let us clarify Glanville's notion of the Black Box as a 'phenomenon' or 'principle' or 'concept'. First, let us assume that the Black Box is spatially located. This forces another assumption, namely, that wherever the location, there must be a mechanoelectrical system that is the basis for – i.e. that *produces* – the Black Box. For example, the brain with its extended network of neurons and blood vessels indisputably *produces* the mind, whose existence is evident through conscious, wilful (i.e., intentional) behavior. The mind is not independent of its host body; likewise, the Black Box is not independent of its mechanoelectrical basis.

Figure 1 schematizes the Black Box and its observer. The observer makes inferences about the Black Box by presenting stimuli, the inputs, and recording the Box's consequent responses, the outputs (Ashby 1961; Glanville 1982, 1997, 2007, 2009a, 2009b). According to Glanville (1982: 1), the observer thereby obtains a 'functional description' of the Black Box: «The 'functional description' […] describes how the observer understands the action of the Black Box» (Glanville 1997, II). The Black Box is 'whitened' (Glanville 1982). Practical examples of 'whitening' through input/output might include an Experimental Psychologist studying the behavior of a human or an animal, or a Physiologist making a noninvasive electrical recording (Nizami 2015, 2017).

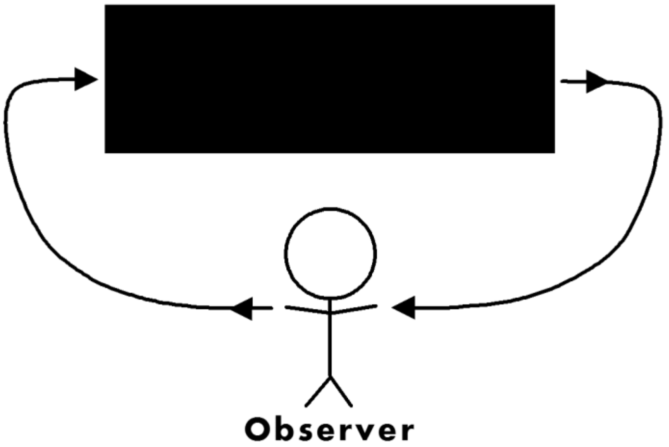


Fig. 1. The Glanville notion of the Black Box and its observer. The observer sends inputs to the Box, and receives outputs from it.

3. Observer as Black Box, Black Box as observer

'Whitening' of the Black Box becomes more intriguing yet. Glanville (1982, 1997, 2009a, 2009b) declares that the *observer* can be considered a Black Box, from the *Black-Box*'s viewpoint. Consider that an *output* of the Black Box is an *input to* its observer; likewise, an input to the Black Box is an *output from* the observer. Hence, «we come to assume that the Black Box also makes a functional description of its interaction with the observer» (Glanville 1997, II). The Black Box 'whitens' its observer, by *acting as* an observer (Glanville 1982).

Consider the following examples. Imagine the mind as a Black Box, probed through input and output. We call this action Psychiatry or Psychology. But each Psychiatrist or Psychologist has their own mind, a Black Box. Those particular Black Boxes regulate everything that those observers say and do; the observers are therefore now Black Boxes. And indeed, Moore (1956) and Ashby (1961) both imply that a *Psychiatrist* and a patient are interacting Black Boxes. When the Psychiatrist (or the Psychologist) probes the patient (or the research subject), e*ach participant* (if awake and aware) *is an observer, who regards the other as a Black Box*. Such interaction implies a *system*.

4. Black Box + observer = 'system': inside every white box there are two Black Boxes trying to get out

Ashby analyzes experiments as follows: «By thus acting on the Box, and by allowing the Box to affect him and his recording apparatus, the experimenter is coupling himself to the Box, so that the two together form a *system* with feedback» (Ashby 1961: 87; italics added). That is, experimenter and Box each 'feed back' to the other, each becoming both observer and Black Box. A BlackBox/observer *system* has properties that differ from those of either the Black Box or the observer alone, or so Glanville implies: «The Black Box and the observer act together to constitute a (new) whole» (Glanville 2009a: 1; see Glanville 2009b: 161). This he calls the *white box* (Glanville 1982).

Figure 2 schematizes the 'white box'. If the observer himself is now taken to be a Black Box, then the title of Glanville's paper of 1982 becomes comprehensible: «Inside every White Box there are two Black Boxes trying to get out». According to Glanville (1982, 2009a, 2009b) the white box, as a system, is nonetheless 'black' to any *further-external* observer.

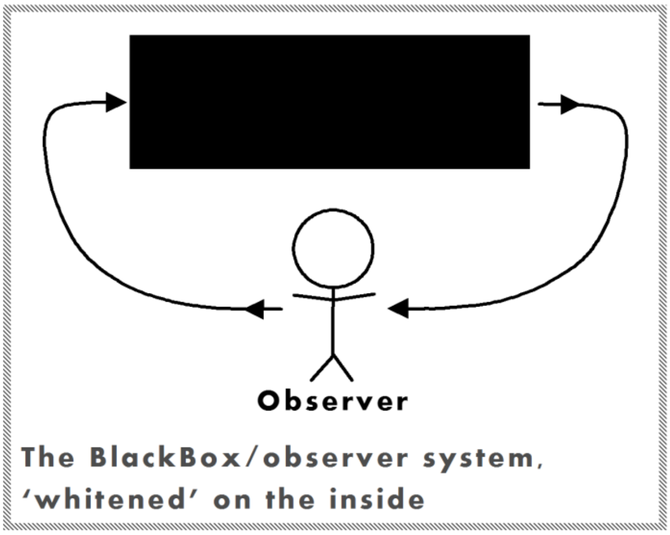


Fig. 2. The Black Box and its observer mutually 'whiten' through interaction, making a 'system' (dashed boundary) that is 'whitened' inside.

5. How would a further-external observer interact with the system?

We have assumed that the Black Box is the *product* of a mechanism. The pictured boundary of any Black Box (and any consequent white box) is *operational*, not physical. Where, therefore, can inputs from a further-external observer go within the BlackBox/observer system? Do they go directly to the core Black Box? Or to the [original] observer? Or, somehow, to both? The answer presumably also tells us where the consequent outputs originate from. Figures 3 and 4 illustrate two possibilities.

Figure 3 illustrates the further-external observer's inputs as going *straight through* the boundary of the BlackBox/observer system, right up to the edge of the core Black Box itself, without interacting with the core Black Box's observer. The latter persona is ignored, as if the further-external observer recognizes that observer's presence and behavior.

Now consider the contrary situation. Figure 4 (after Glanville 2009b: 164) shows the core BlackBox/observer system *not* being penetrated by the input-and-output pathways to the further-external observer. Indeed, Glanville (2009a) implies that in a recursion of Black Boxes (and observers), *none of the observers know of each other's existence*. But Glanville (2009b) later fails to be definitive about this. Indeed, he provides no rationale for the discrepancy between his approach of 1982 and his approach of 2009. And he can no longer provide a rationale (see the links to the obituaries, in the References). Consequently, the present author attempts the task. Important insights emerge, but first, some background is needed.

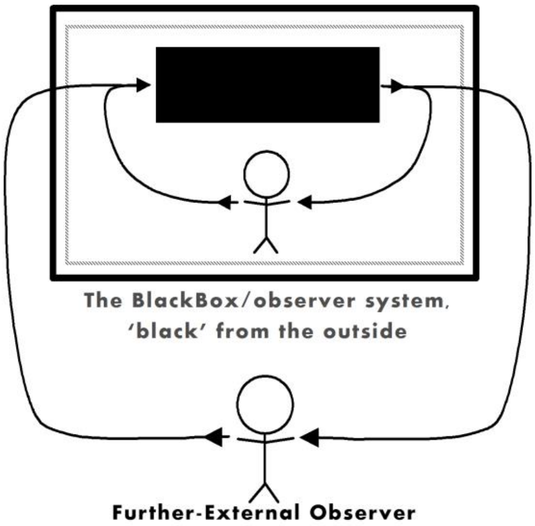


Fig. 3. The core BlackBox/observer system as a Black Box which is penetrable by the input/output paths from/to a further-external observer.

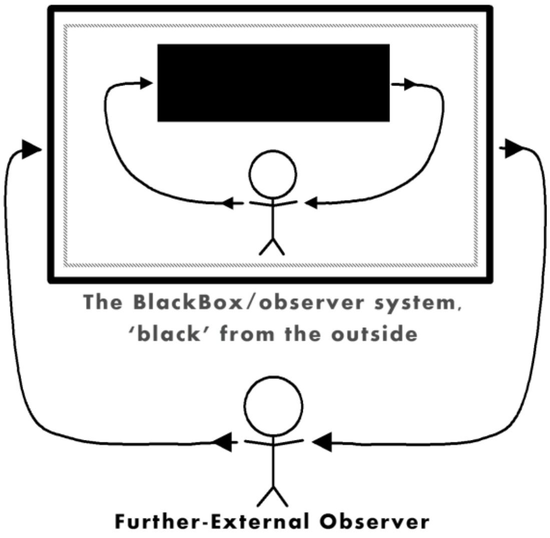


Fig. 4. Compare to Fig. 3. The core BlackBox/observer system as a Black Box which is *not* penetrable by the input/output paths from/to a further-external observer.

6. Interim (1): 'Trivial' Machines

Consider the *machine*, introduced in Section 2 as a mechanoelectrical device. Henceforth, 'machine' will be used in a different way. As von Foerster (2003: 207) states, «The term 'machine' in this context refers to *well-defined functional properties of an abstract entity* rather than to an assembly of cogwheels, buttons and levers, although such assemblies may represent embodiments of these abstract functional entities [i.e., the 'machines']» (italics added). By this definition, an abstract entity that is a *product* of a mechanoelectrical basis – i.e., an abstract entity such as the Glanville Black Box – is a 'machine'.

Von Foerster recognizes two types of machines: Trivial, and Non-Trivial. He explains: «A *trivial* machine is characterized by a one-to-one relationship between its 'input' (stimulus, cause) and its 'output' (response, effect). This invariant *relationship* is 'the machine'» (von Foerster 2003: 208; italics added). He continues: «Since this relationship is determined once and for all, this is a deterministic system; and since an output once observed for a given input will be the same for the same input given later, this is also a *predictable* system» (*ibidem*; italics added). Algebra-wise, von Foerster explains that, for input **x** and output **y**, «a **y** once observed for a given **x** will be the same for the same **x** given later» (von Foerster 1984: 9). That is, «one simply has to record for each given **x** the corresponding **y**. This record is then 'the machine'» (von Foerster 1984: 10).

Von Foerster (1984: 10) provides an example of a Trivial Machine, in the form of a table which assigns an output **y** to each of four inputs **x**. The **x**'s are the letters A, U, S, and T, and the respective outputs **y** are 0, 1, 1, and 0. We will return to this table, below.

Von Foerster (2003: 208) notes that «All machines [that] we construct and buy are, hopefully, trivial machines», that is, *predictable* ones.

7. Interim (2): internal 'states'

Of course, in reality, *not all machines are Trivial*. A *physical* machine, as Ashby (1961) points out, can have internal conditions or configurations, which Ashby calls 'states'. We will assume that so, too, can the *products* of physical machines, namely, *conceptual* machines such as Black Boxes. Ashby notes «that certain states of the Box cannot be returned to at will», which he declares «is very common in practice. Such states will be called **inaccessible**» (all from Ashby 1961: 92; original boldface). Ashby continues: «Essentially the same phenomenon occurs when experiments are conducted on an organism that *learns*; for as time goes on it leaves its 'unsophisticated' initial state, and no simple manipulation can get it back to this state» (Ashby 1961: 92; italics added). *Learning* presumably refers to changes in abilities and knowledge, that are reflected in changes of behavior.

Here, *mind is machine is Black Box*. Learning is presumably concurrent with changes in the mind's mechanoelectrical basis, the brain; changes in brain-states manifest as changes in mind-states. Thanks to learning, our response to a stimulus can differ from our previous response, and in unexpected ways. For a stimulus that is a *question*, for example, von Foerster (2003: 311) notes that a child can offer a correct [carefully trained] answer, or a correct but unexpected answer, or an answer that is intentionally capricious. *Minds are not Trivial Machines*.

We must therefore ask whether the observer of any Black Box can give input, and record output, without changing the Box's possible output to the next input. That is, can the Black Box be *observed* without being *perturbed*? Likewise, can a Black Box's output be observed by, but not perturb, the observer?

8. Interim (3): sequential machines

There are conceivably perturbable 'machines'. Such devices were envisioned long before Ashby (1961). Indeed, Turing (1937) conceives of a machine whose input and/or output can change the response to the next input. Turing describes the machine only in terms of its process. The machine has a finite number of internal states, called 'conditions' or 'configurations'. The machine accepts an input in the form of a continuous tape, divided into equal segments, each containing a symbol or being blank. The machine scans one tape segment at a time; the scanned symbol (or the blank), along with the machine's current configuration, altogether determine the impending response. That response can include erasing a symbol from the tape; or printing (or not), on a blank segment of the tape, a symbol consisting of a digit (0 or 1) or some other symbol; or shifting the tape one segment to the left or one segment to the right (Turing 1937).

Turing's machine exemplified what came to be known as *sequential machines*. One class of them was described by E.F. Moore (1956). He, like Turing, used operational descriptions: «The state that the machine will be in at a given time [the 'current state'] depends only on its state at the previous time and the previous input symbol. The output symbol at a given time depends only on the current state of the machine» (Moore 1956: 133). That is, an input evokes an output, the identity of which is nonetheless determined *only* by the present internal state. That state then changes to another state, that is determined by the *input*.

Moore (1956) provides an example of a sequential machine, in the form of two tables that relate the inputs, outputs, and internal states (Moore 1956: 134). Let us call the inputs **x**. Moore's inputs are also the possible outputs, but for the sake of distinction, let us call the outputs **y**. One of Moore's tables shows the 'present output' **y** of the machine, as a function of the 'present state', call it **z**. Let this relation be called **y**=**F**(**z**), characterizing an 'Output Generator'. Moore's second table shows «the present state of the machine … as a function of the previous state and the previous input» (Moore 1956: 134). Let us call Moore's 'previous state' $\mathbf{z}_{-1}$ and the 'previous input' $\mathbf{x}_{-1}$. Let us use **z'** for the state of **z** which occurs *after* **y** is output. Let $\mathbf{z}_{-1}$, **z**, and **z'** be determined by the 'State Generator' **Z**, and express **Z** in terms of **z** rather than $\mathbf{z}_{-1}$. Moore (1956) uses four possible internal states, called q_1 , q_2, q_3, and q_4, and two possible inputs, **x** = 0 or **x** = 1. All of this notation may seem awkward, but it is consistent with the work of von Foerster (1984, 2003), continued below.

Table 1 shows a re-arrangement of Moore's two tables into five smaller tables, four of which show **z** as a function of $\mathbf{x}_{-1}$ for the four possible values of $\mathbf{z}_{-1}$(q_1 , q_2, q_3, and q_4). The remaining table shows **y** as a function of **z**.

As an example of how the Moore sequential machine works, note that **z** = q_4 could have arisen from $\mathbf{z}_{-1}$= q_3 and $\mathbf{x}_{-1}$= 0 *or* 1 (third pair of columns from the left in Table 1), or alternatively from $\mathbf{z}_{-1}$= q_1 and $\mathbf{x}_{-1}$= 0 (leftmost pair of columns in Table 1). Regardless of whether the input is **x** = 0 or **x** = 1, the resulting output is **y** = 1 because **z** = q_4 (see the rightmost pair of columns in Table 1), after which **x** = 0 or **x** = 1 respectively become $\mathbf{x}_{-1}$= 0 or $\mathbf{x}_{-1}$= 1. Likewise, **z** = q_4 is now $\mathbf{z}_{-1}$= q_4, which leads to a new internal state **z'**, which in fact will be q_2 (see the second pair of columns from the right in Table 1). A subsequent input **x** = 0 *or* **x** = 1 will then result in **y** = 0 (rightmost pair of columns in Table 1).

Note well that, in Moore's scheme, a particular output can result from different internal states; and a particular internal state can result from different *inputs*. Note equally well that Moore's two tables are *unchanging*. That is, what we presently call the State Generator and the Output Generator are deterministic (i.e., non-random) *and* they are predictable, insofar as an outside observer supplying input and recording output can gain increasing confidence about each Generator's operating rules. *Both Generators are Trivial Machines.*

$z_{-1} = q_1$	
x_{-1}	z
0	q_4
1	q_3

$z_{-1} = q_2$	
x_{-1}	z
0	q_1
1	q_3

$z_{-1} = q_3$	
x_{-1}	z
0	q_4
1	q_4

$z_{-1} = q_4$	
x_{-1}	z
0	q_2
1	q_2

z	y
q_1	0
q_2	0
q_3	0
q_4	1

Table 1. Relations in E.F. Moore's example of a sequential machine (Moore 1956). The rightmost table describes the Output Generator; the other four tables describe the State Generator, for internal states q_1 , q_2 , q_3 , or q_4 .

But the concatenation of two Trivial Machines can be non-trivial, i.e., non-predictable; the whole is more than the sum of the parts. This wholism is called 'emergence' (Nizami 2017, 2018). How would an *observer* of the sequential machine (not its *maker*) gain the data to fill-in Moore's two tables? Moore introduces «a somewhat artificial restriction that will be imposed on the action of the experimenter. He is not allowed to open up the machine and look at the parts to see what they are and how they are interconnected» (Moore 1956: 132). That is, «the machines under consideration are always just what are sometimes called 'black boxes', described in terms of their inputs and outputs, but no internal construction information can be gained» (Moore 1956: 132).

Moore himself offers no picture of a sequential machine as a 'black box'. Hence, let us make one. Figure 5 shows a sequential machine involving two Trivial Machines, whose operations follow the relations in tables such as Moore's.

Sequential machines have broad importance. They are cases of what von Foerster (1984, 2003) later calls *Non-Trivial Machines*. Like Moore (1956), von Foerster provides an example in the form of two tables (1984: 11). The tables describe the output **y** and the next state **z'** in terms of the input **x**, but for only two possible internal states, the present states **z**, dubbed **I** or **II**. Having two states characterizes the simplest Non-Trivial Machine; under only *one* internal state, a particular input would always evoke a particular pre-determined, unchanging output, making the machine Trivial. Nonetheless, von Foerster's example inputs and outputs were the same as for his Trivial Machine: **x** = A, U, S, or T, and **y** = 0 or 1.

Figure 6 schematizes von Foerster's Non-Trivial Machine. Von Foerster's data can be re-ordered, to make four new tables, two for each of **z** = **I** and **z** = **II**. Table 2 contains the four tables. Two of the tables show **y** as a function of **x**, and two of the tables show **z'** as a function of **x**. These latter pairs are the respective equivalents of the Output Generator and the State Generator of Moore's (1956) sequential machine (Fig. 5). Von Foerster calls them the Driving Function and the State Function.

The machines in Fig. 5 and Fig. 6 profoundly differ in one detail. To Moore (1956), the input **x** has no bearing on the *immediate resulting* output **y** (Fig. 5), only on its *successor* by way of the internal state. Moore's **y** is *evoked* by **x**, but is only indirectly a *function of* **x** by way of the internal state. In contrast, in von Foerster's (1984, 2003) machine (Fig. 6), **x** directly affects **y**, *and* indirectly affects the next output by way of the internal state.

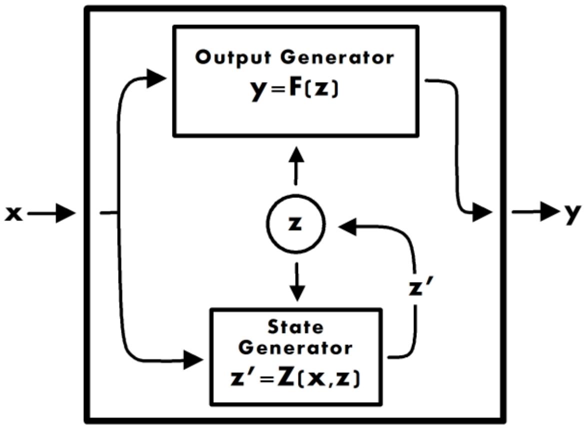


Fig. 5. A Moore (1956) sequential machine, depicted here in the style of von Foerster (1984, 2003). The boxes and lines and the circle represent mechanoelectrical parts. The lines with arrows represent the parts' operating relations, which need not occur simultaneously. The internal state **z** actively affects the Output Generator **F**, and the State Generator **Z** from which **z** arose. **Z** produces a new state **z'** after **y** is output by **F**, when **F** is prompted by the input **x**.

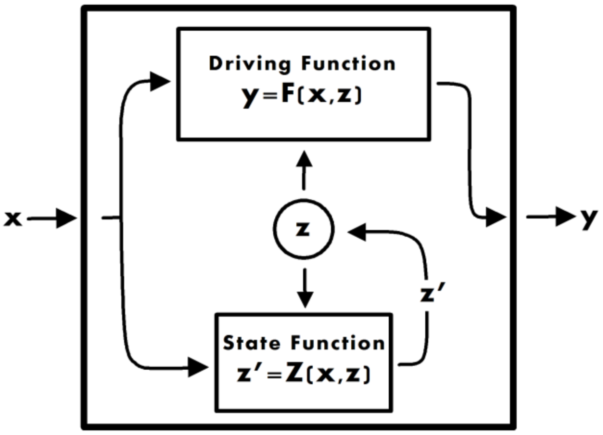


Fig. 6. Von Foerster's 'Non-Trivial Machine'. It involves two Trivial Machines, the 'Driving Function' and the 'State Function', the respective operational equivalents of the Output Generator and the State Generator in Fig. 5. In Fig. 5, however, **y** is a direct function only of **z**; here, **y** is a direct function of both **x** and **z**.

The operation of the von Foerster Non-Trivial Machine represented by Table 2 is best exemplified through von Foerster's own words:

> We present the machine with several A's (A, A, A, …), and to our satisfaction we get consistently zeros (0, 0, 0, …). We turn now to a sequence of U's (U, U, U, …), to which the machine responds with a sequence of ones (1, 1, 1, …). Confidently

> we try the input S and obtain 1; but when checking out S again, for one who does not know the inner workings of the machine, something unpleasant is happening: instead of a 1, the machine responds with a 0. We could have predicted that, because the state function switches the machine when in **I**, given S, into its internal state **II**, and here the response to stimulus "S" is "0". However, being in **II**, given S, the machine returns to internal state **I**, and a new test of S will yield 1, etc., etc., … Checking out the patriotic sequence USA, depending upon whether one starts when the machine is in its internal state **I** or in **II**, it will respond with either 111, or else with 000, apparently indicating different political persuasions. Perhaps these examples suffice to justify the qualifier "non-trivial" for these machines. (von Foerster 1984: 11)

If von Foerster's Non-Trivial Machine is not a Moore machine, then where did von Foerster obtain it? The answer is that von Foerster's Non-Trivial Machines are Mealy machines, named after George H. Mealy (1955), a contemporary of Ashby and Moore. Mealy's archetype machine (Mealy 1955) has inputs $\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_n$ and outputs $\mathbf{y}_1,\ \mathbf{y}_2,\ \dots,\ \mathbf{y}_m$ and internal states $\mathbf{q}_1,\ \mathbf{q}_2,\ \dots,\ \mathbf{q}_s,$ all of which are valued as either 0 or 1. Mealy's machines are realized as electrical circuits, although they can contain mechanoelectrical parts such as the relay coils of Mealy's era.

z = I	
x	**y**
A	0
U	1
S	1
T	0

z = II	
x	**y**
A	1
U	0
S	0
T	1

z = I	
x	**z'**
A	**I**
U	**I**
S	**II**
T	**II**

z = II	
x	**z'**
A	**I**
U	**II**
S	**I**
T	**II**

Table 2. Relations in von Foerster's example of a Non-Trivial Machine (von Foerster 1984). The two leftmost tables describe the Driving Function, and the two rightmost tables describe the State Function, for internal states **I** and **II**.

Mealy (1955: 1047) explains his use of 0s and 1s as follows: he wishes to design an electrical circuit in which «*Any lead or device within the circuit may assume, at any instant of time, only one of two conditions, such as high or low voltage, pulse or no pulse*» (original italics). A pulse may be of voltage or of current (Mealy 1955: 1047). There is a further condition: «*The behavior of the circuit may be completely described by the consideration of conditions in the circuit at equally-spaced instants in time*» (Mealy 1955: 1047; original italics). Altogether, this statement along with its predecessor define a *synchronous circuit*. Mealy clarifies:

> (1). There is a so-called clock which supplies timing pulses to the circuit. (2). Inputs and outputs are in the form of voltage or current pulses which occur synchronously with pulses from the clock. (3). The repetition rate of the clock pulses may be varied, within limits, without affecting the correct operation of the circuit, so long as input pulses remain synchronized with the clock. (Mealy 1955: 1047)

Mealy then provides

> an abstract definition of a switching circuit: *A switching circuit is a circuit with a finite number of inputs, outputs, and (internal) states. Its present output combination and next state are determined uniquely by the present input combination and the present state. If the circuit has one internal state, we call it a* combinational circuit*; otherwise, we call it a* sequential circuit. (Mealy 1955: 1049; original italics)

What allows a 'present' state to be physically distinguished from a 'next' state? The answer is *delay* elements, where «The unit of delay is the interval between the start of two successive clock pulses» (Mealy 1955: 1048). The lines of electrical conductance that contain delay elements are *delay lines*. The input of a delay line will be its output one clock cycle later. But present-day electronics may contain circuits that are synchronous and those that are not, called *asynchronous circuits*, where

> We agree (1) that no clock will be used and (2) that "l" in switching algebra will correspond to a high voltage or current, an energized relay coil, or operated relay contacts. We must now pay careful attention to circuit conditions at *every* instant of time. (Mealy 1955: 1067; original italics and quotation marks)

For example, the difference between a present state of 1 and a consequent state of 1 is the difference between a relay being operated and a relay being energized to allow imminent operation. The reader is left to peruse the details in Mealy (1955). Note well, however, that von Foerster's Non-Trivial Machine lacks an explicit clock. Hence, we are perhaps expected to *presume* that a sequence of inputs is clocked. Clocking of either a Moore machine or a Mealy machine might occur in the following manner (thanks to a reviewer for this suggestion). Imagine each input being accompanied by a positive pulse having both a rising edge and a falling edge. The arrival of an input (rising edge) would trigger either the state function or the driving function, depending upon how the cycle of operation of the respective machine is imagined; then, the falling edge would trigger the other function. Input and output would thus occur with a predetermined delay.

9. 'Systems' in terms of 'machines'

To Moore, scientists belong to systems: «The experiment may not be completely isolated from the experimenter, i.e., the experimenter may be experimenting on a system of which he himself is a part» (1956: 133). So «The experimenter [probing a 'machine'] could be described as another sequential machine, also specified in terms of its internal states, inputs, and outputs. The output of the machine being experimented on would serve as input to the experimenter and vice versa» (1956: 135).

Further logic-wise (but earlier text-wise), Moore (1956: 132) notes that a Psychiatrist *experiments on* a patient, giving inputs and receiving outputs. Moore's 'black box' is evidently the *mind*. As Moore declares (1956: 132), «The black box restriction corresponds approximately to the distinction between the psychiatrist and the brain surgeon», i.e. insofar as the surgeon can alter the brain, but only the Psychiatrist can alter the mind. (Modern surgeons might disagree, but that is beside the point.)

For a sequential machine to be a mind, it would need to be capable of an enormous number of possible behaviors. Presently, let us treat a Non-Trivial Machine's input/output combinations as behaviors. Just how many such behaviors are possible? For a Mealy machine having just one binary-valued input (0 or 1), one binary-valued internal state (0 or 1), and one binary-valued output (0 or 1), there are 256 ways of having either a 0 or a 1 as input and subsequently a 0 or a 1 as output. If we increase the number of binary-valued input variables by just one, then the number of input/output combinations increases to 4,096, that is, it increases by a factor of 16. Imagine now the immensity of the electrical device that is the human brain, where a voltage spike might be taken as a 1 and a lack of it be taken as a 0. Nonetheless, we should avoid the mistaken notion of the brain as a computer.

10. How a further-external observer would interact with the system

Sections 6, 7, and 8 introduced the concept of the machine, as an aid to resolving a quandary. That predicament is illustrated in Figs. 3 and 4. It is the question of whether a further-external observer of the BlackBox/observer system can ignore the original, internal observer, as if knowing of that observer's presence and behavior.

To resolve the uncertainty, let us assume first that the core Black Box can be probed by its immediate observer without being perturbed. Suppose also that the immediate observer remains unperturbed by the output from the core Black Box. Altogether, the core BlackBox/observer system is unaltered by its internal interactions. Hence, the degree to which the immediate observer and the core Black Box understand each other will be limited only by the number of possible inputs from each to the other. The core Black Box and its immediate observer can fully 'whiten' each other in time. They are Trivial Machines.

This *implies* that if the core BlackBox/observer system is probed by a further-external observer, the latter can ignore the immediate observer and directly interrogate the core Black Box. This direct access applies 'by induction' to all further-outward observers. This is what Glanville illustrates in 1982, and is shown here as Fig. 3. The system formed by the combination of *any* observer with the core Black Box is no different than the system formed by the combination of *any other* observer with the core Black Box. The core BlackBox/observer *system* is penetrable. And it is not unique; any other, possibly further-out observer can pair with the core Black Box to form a potentially identical system.

Consider now the alternative. Imagine that the core Black Box *cannot* be probed by its immediate observer without being perturbed, and that, similarly, the immediate observer cannot receive output from the core Black Box without changing. Box and observer are now Non-Trivial Machines. But a Non-Trivial Machine concatenated with a Non-Trivial Machine is, perforce, a Non-Trivial Machine. The core Black Box and its immediate observer are now truly 'entangled', that is, no outside observer can tell them apart and hence *ignore* the immediate observer.

Thanks to the concept of machines, we can now comprehend the difference between the portrayal of the Black Box and its observer in Glanville (1982) and that in Glanville (2009a, 2009b). In the earlier Glanville, the Black Box and its observer are Trivial Machines; in the later Glanville, they are Non-Trivial Machines.

If the core BlackBox/observer system is a Non-Trivial Machine, then it would be perturbed if probed through input from a *further-external* observer, as in Fig. 4. Whether or not that further-external observer is himself a Non-Trivial Machine, nonetheless his concatenation with the core BlackBox/observer system is a new system which is a Non-Trivial Machine. *That* system is a Black Box to any *yet-further-external* observer, and can be perturbed by that observer. Glanville notes that «each Black Box is potentially made up of a recursion of Black Boxes (and observers)» (Glanville 2009a). Figure 7 shows the recursion.

11. At the core of any Black Box there are two (or more) white boxes, required to stay in

The title of Glanville's landmark paper of 1982 was «Inside every White Box there are two Black Boxes trying to get out». Figure 2 shows this arrangement when the observer himself is a Black Box. But the arguments above suggest a new interpretation. Let us presume that the core Black Box is a Non-Trivial Machine, composed of concatenated Trivial Machines. Then, no matter how many nested layers of Black Boxes and observers might occur Russian-Doll fashion within *any* Black Box (Fig. 7), the latter Box has an utter core containing a Black Box which consists of two (or more) white boxes, boxes that are required to *stay in* – observed by an observer who, if he's a Black Box himself, also consists of two (or more) white boxes. Figure 8 schematizes the old versus new approaches to the relation of white boxes to Black Boxes.

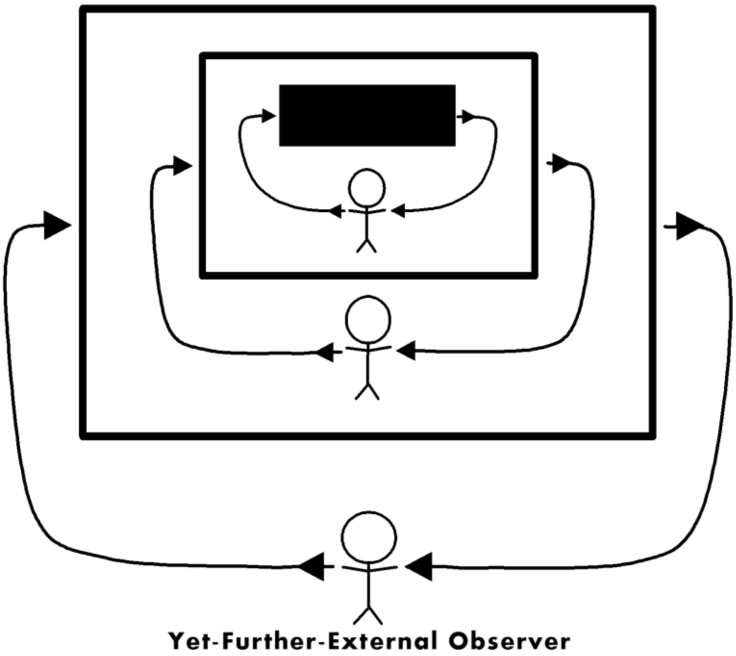
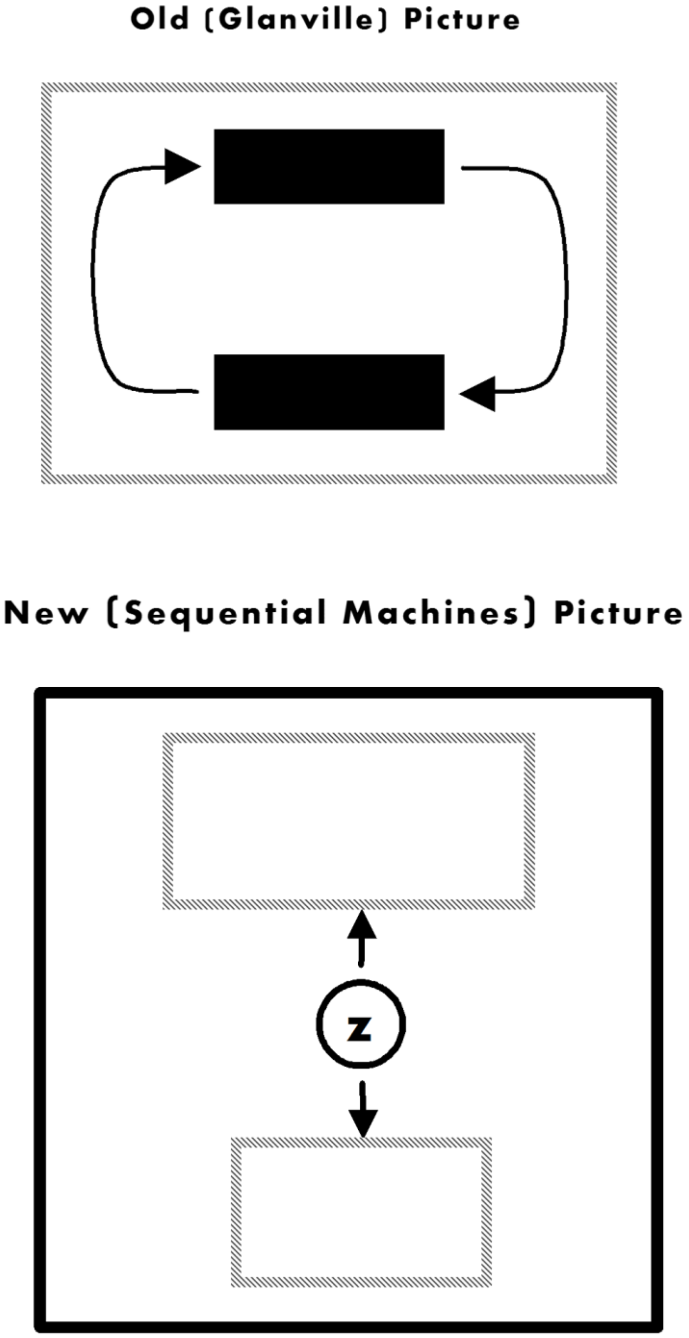


Fig. 7. From the viewpoint of a *yet-further-external* observer, the greater system shown in Fig. 4 is a Black Box; and so on, with each further-outwards Black Box having its own immediate observer.

12. Conclusions

Sensations – and the ability to report them – characterize the mind. But no-one can directly observe their own mind, or any other. Here, we attempt to understand the mind indirectly, through the concepts of the Black Box and its observer. Ranulph Glanville proselytized these concepts after W. Ross Ashby.

Glanville's seminal paper (1982) was titled «Inside every White Box there are two Black Boxes trying to get out». Instead, we can say that at the utter core of any Black Box there are two (or more) white boxes, required to stay in. Remember that the operation of a Black Box that is a Non-Trivial Machine is not random, but may nonetheless be difficult, perhaps impossible, for an observer to comprehend. As such, the operation of Black Boxes – or of ensembles of them – may seem *emergent*. The mind, too, seems emergent (Nizami 2015, 2017, 2018), such that ensembles of white boxes and Black Boxes may be the ultimate source of the mind.



Old (Glanville) Picture

New (Sequential Machines) Picture

Fig. 8. White boxes versus Black Boxes. (Upper) The Glanville (1982) view that «Inside every White Box there are two Black Boxes trying to get out». (Lower) The alternative philosophy that at the utter core of any Black Box there are two (or more) white boxes, required to stay in.

Acknowledgments
I thank the Reviewers and Dr. Claire S. Barnes PhD for their substantial contributions to improving this paper.

References

Ashby, W. Ross (1961), *An introduction to cybernetics*, Chapman & Hall Ltd., London.

Glanville, Ranulph (1982), «*Inside every White Box there are two Black Boxes trying to get out*», in *Behavioral Science*, vol. 27, n. 1, pp. 1-11.

Glanville, Ranulph (1997), *Behind the curtain*, in Ascott, R. (eds.), *Proceedings of the First Conference on Consciousness Reframed*, UCWN (University of Wales College Newport), Wales, 5 pages, not numbered.

Glanville, Ranulph (2007), «*A (cybernetic) musing: Ashby and the Black Box*», in *Cybernetics & Human Knowing*, vol. 14, n. 2/3, pp. 189-196.

Glanville, Ranulph (2009a), *Darkening the Black Box* (Abstract), in *Proceedings of the 13th World Multi-Conference on Systemics, Cybernetics and Informatics*, International Institute of Informatics and Systemics, Orlando (FL).

Glanville, Ranulph (2009b), «*Black Boxes*», in *Cybernetics & Human Knowing*, vol. 16, n. 1/2, pp. 153-167.

Glanville obituaries:
https://www.aaschool.ac.uk/PUBLIC/NEWSNOTICES/obituaries.php?page=5,
http://www.isce.edu/Glanville.html

Holland, Owen (2003), «*Exploration and high adventure: the legacy of Grey Walter*», in *Philosophical Transactions of the Royal Society A*, vol. 361, n. 1811, pp. 2085-2121.

Holm, Elizabeth A. (2019), «*In defense of the black box*», in *Science*, vol. 364, n. 6435, pp. 26-27.

Mealy, George H. (1955), «*A method for synthesizing sequential circuits*», in *Bell System Technical Journal*, vol. 34, n. 5, pp. 1045-1079.

Moore, Edward F. (1956), *Gedanken-experiments on sequential machines*, in Shannon, Claude E., McCarthy, John (eds.), *Automata Studies* (Annals of Mathematics Studies Number 34), Princeton University Press, Princeton (NJ), pp. 129-153.

Nizami, Lance (2015), «*Homunculus strides again: why 'information transmitted' in neuroscience tells us nothing*», in *Kybernetes*, vol. 44, n. 8/9, pp. 1358-1370.

Nizami, Lance (2017), «*I, NEURON: the neuron as the collective*», in *Kybernetes*, vol. 46, n. 9, pp. 1508-1526.

Nizami, Lance (2018), «*Reductionism ad absurdum*: *Attneave and Dennett cannot reduce homunculus (and hence the mind)*», in *Kybernetes*, vol. 47, n. 1, pp. 163-185.

Nizami, Lance (2019a), «*Too resilient for anyone's good: 'infant psychophysics' viewed through second-order cybernetics, part 1 (background and problems)*», in *Kybernetes*, vol. 48, n. 4, pp. 751-768.

Nizami, Lance (2019b), «*Too resilient for anyone's good: 'infant psychophysics' viewed through second-order cybernetics, part 2 (re-interpretation)*», in *Kybernetes*, vol. 48, n. 4, pp. 769-781.

Rahwan, Iyad, Cebrian, Manuel, Obradovich, Nick, et al. (2019), «Machine behavior», in *Nature*, vol. 568, pp. 477-486.

Ramage, Magnus, Shipp, Karen (2009), *Systems thinkers*, Springer, New York.

Turing, Alan M. (1937), «*On computable numbers, with an application to the Entscheidungsproblem», in *Proceedings of the London Mathematical Society*, vol. s2-42, n. 1, pp. 230-265.

von Foerster, Heinz (1984), *Principles of self-organization – in a socio-managerial context*, in Ulrich, H., Probst, Gilbert J.B., (eds.), *Self-organization and management of social systems*, *Springer Series in Synergetics, vol. 26*, Springer, Heidelberg, pp. 2-24.

von Foerster, Heinz (2003), *Understanding understanding: essays on cybernetics and cognition*, Springer-Verlag, New York.

Walter, W. Grey (1950), «*An imitation of life*», in *Scientific American*, vol. 182, n. 5, pp. 42-45.