

Can Animals and Machines Be Persons?

A Dialogue by Justin Leiber

1

Intentionality

Intentionality is the philosophical concept of the mind being directed toward some thing and forming internal representations.

“Our beliefs, thoughts, wishes, dreams, and desires are about things. Equally the words we use to express these beliefs and other mental states are about things.”*

(Note that it is **not** about *intention*.)

*The Oxford Dictionary of Philosophy, by Simon Blackburn, Oxford University Press

2

The First Morning

Respect for the basic rights of persons

Chimpanzee: Washoe-Delta
can only live in weightlessness

Computing System: AL
cannot function in earth gravity

Must not terminate if

- (1) Washoe-Delta and AL “think and feel” and as such
- (2) “are persons” and hence
- (3) “their termination would violate their rights as persons.”

3

The First Morning

Three Questions:

1. Do either Washoe-Delta or AL “think” or “feel”?
2. In virtue of their cognitive and emotional life,
is either a person?
3. If one or the other is indeed a person,
does that person have a right to continued existence?

4

The First Morning

Earthly courts have used the words “human,” “person,” “man,” and so on, in careless variation, to mean the same individuals.

A human being is a person and a person is a human being.

Biologically speaking, humans and chimpanzees have in common more than 99% of their genes. Because chimpanzees are 99% biologically human, are they 99% persons?

One is a person or one isn't.

Does everything depend on whether humans and chimpanzees are close enough genetically to produce offspring?

5

The First Morning

Humans think, feel, make choices, are conscious, and use language in fascinating and complicated ways.

Humans are reasoners, beings with intrinsic dignity and worth.

Alone among all creatures, each of us has what is called, variously, “a consciousness,” “a self,” “a mind,” “a soul,” “a spirit.”

What about dead human beings?

The flesh that remains is in itself not a human being.

So what makes a human being a human being is what the flesh is **doing** — alive, breathes, behaves in a way we recognize as thinking, feeling, making choices, ...

6

The First Morning

There are some sorts of persons who are not human beings. Corporations, nations, associations, clubs, organizations are “artificial” persons that exist only as an expression of the real persons, the humans, who set them up.

7

The First Morning

Dispute the claim that “human being” always and everywhere has implied “person,” in the primary and literal sense of “having intrinsic dignity and worth.”

Women have not always been accorded absolutely all the rights of men.

Documents specifying “the rights of men” have not been interpreted to include women on a par with men.

Through most of recorded history, in most religious and cultural traditions, women have not been regarded as having intrinsic dignity and worth.

8

The First Morning

The notion of “person” has always carried with it the notion that one has certain rights and powers, that one has intrinsic dignity and worth, and that is also what is meant by saying that one has a “soul” or a “mind.”

Through most of human history there has been a general pattern of mistreatment, of second class citizenship, for women.

Happily, we have put this silly ignorance behind us.

If it’s “absolutely obvious” that women are persons and fullfledged person, why did it take so long for men to see it?

The hurt is long gone.

9

The First Morning

Don’t throw the past away so quickly.

Rights of Man — published in 1791

A Vindication of the Rights of Women — published in 1791

(When written out, sarcasm becomes satire, and there is nothing like satire for revealing deeply held convictions.)

A Vindication of the Rights of Brutes — published in 1792

A satirical presumption that it is nearly as absurd to think women the equals of men, or slaves the equals of masters, as to think that animals have the features of persons.

10

The First Morning

The egalitarian view is a very recent and unusual one.

The traditional view is not simply that women lack the rights, the 100 percent person status, of men, but that all manner of men and women are naturally slaves and subordinates.

One has to give a reason for expanding the circle of fullfledged persons to include all (adult) human beings.

Thomas Paine had to *argue* that all men had rights.

Godwin had to *argue* that those very arguments would also show that women had equal rights.

The reason given is ***intelligence***.

(Paine and Godwin had no interest in the cause of animals.)

11

The First Morning

“Slippery Slope” argument

If the path leads to an *undesired* conclusion, then it must be rejected.

The kind of arguments that have been given to accord some significant level of personhood to all humans also support possible further extensions.

“Human being” isn’t necessarily equivalent to “person.” You have to mention some features that make humans, or some humans, “persons.”

Those same features may be mentioned in attempting to show that some nonhumans are persons.

12

The First Morning

One has to propose a test that humans have to pass to show that they are indeed persons; and then one must, in the interests of fair play, be prepared to see whether other things can pass the test.

Alan Turing suggests just such a test: the “imitation game.”

Alan Turing worked out the basic abstract theory of what an all-powerful computing device has to be able to do:

Universal Turing Machine.

Is it absurd to suppose that literally everything is a person?

Do all things have intrinsic dignity and worth?

13

The First Morning

If we know what computers *can* do, we may be in a better position to tell whether there is some further distinctive feature of humans that clearly and objectively makes us persons.

Perhaps there is no objective question here. I might join with my villagers, or my compatriots, in saying something like “my village (my nation) right or wrong.”

Perhaps it may be: **“any humans, right or wrong!”**

Maybe it is just an arbitrary decision to limit room on the raft to one’s closest neighbors, however unfair and unprincipled.

14

The Afternoon

In 1949, Alan Turing raised the question whether machines think.

He proposed a substitute question: the “Imitation Game.”

Turing appeals to the assumption that men and women are basically very similar in mental capacities, so that a member of one sex can pass quite well mentally as a member of another.

Turing suggests that while there may be obvious physical differences between men and women, they are equally persons in virtue of these mental capacities.

Joseph Weizenbaum wrote the ELIZA program to simulate a non-directive therapist conducting a psychotherapy session.

15

The Afternoon

Metaphorically there is an ingenuous man, or team of men, “inside” AL, namely AL’s human designers and programmers. It’s their intentions that give the only meaning and sense that the computer’s endless number crunching can have.

We know a lot less about how humans think than we assume.

If a man runs into something he wants to think is a clever mental fake or some sort of demon, he will end up confidently condemning most genuine articles as fakes. We are ever too confident that we understand what is going on inside another’s mind.

16

The Afternoon

Computers... I had thought that all they could do was follow programmed instructions (very quickly and impressively). But replacing one set of signs by another doesn't require real intelligence. Isn't that all there is to computers? They do not learn or create ideas.

All a computer is is a set of formal **rules** for changing meaningless sets of signs into other meaningless sets of signs.

*Computers can be “**programmed to learn.**”*

They can be programmed in meta-instructions, rules that tell them to add or subtract rules, if things go a certain way – even to add and subtract rules for adding or subtracting rules.

17

The Afternoon

They have been creative in that, in following such rules, they have sometimes come up with new “solutions” – for example, shorter proofs for certain theorems in mathematics.

How does this endless manipulation of symbols mean anything, or mount up to a person who has meaningful thoughts and emotions, and a sense of personhood?

A computer is a fantastically complicated network of off-on switches, not something you can think of as a person.

What is the human brain but a fantastically complicated network of neurons?

The Turing test has the advantage of drawing a fairly sharp line between the physical and the intellectual capacities.

18

The Afternoon

Once we have a compact, powerful computer, give it a robot body and send it out on its own to investigate the world.

The computer could be a profound shock to the English class system because it would have just those skills—abstract intellectual skills—whose possession was supposed to justify upper class privileges.

We have, rather, made the programmers a new component of our ruling classes. We seem to want to credit entirely to humans any joint achievement of computers and the humans who help build and program them.

Computers are *our* creations. Surely computers cannot build and program themselves?

19

The Afternoon

The world knows that computers and computer-robots have a crucial role in building and programming computers.

A computer-robot could in theory be programmed to construct a duplicate, or even more powerful, version of itself.

The smallest and stupidest living organism can in fact do the same.

The dazzlingly complex genetic double helixes of DNA are likely a very complicated set of instructions – a program – for building a still more complicated self-replicating machine.

20

The Afternoon

The amoeba DNA doesn't *mean* to produce specific protein molecular structures that it does, nor does it *intend* to make an offspring.

Whatever DNA can pass the Turing test we must let on the raft because we are on the raft for the same reason.

If our hydrocarbon chemicals can pass the test, why not give the microchip crystals a chance?

AL could pass the Turing test or win at the imitation game. But you say that this is irrelevant because its pass is a fake. The human designers pass and *not* AL. The pass is a fake because the computer can't really mean or think or intend the sense of the symbols that it prints out.

21

The Afternoon

The claim that AL is not really having inner mental experience — that “he,” or “it,” or whatever, is just symbol crunching by human artifice — doesn't apply to Washoe-Delta acting like a thinking, intending, and feeling human being.

As Bertrand Russell once remarked, an intelligent extraterrestrial might think the difference between men and monkeys very slight, and be equally interested in both.

The Gardeners brought up Washoe as a human child – in a home where human sign language was spoken. In motor abilities, sensory development, eye-hand coordination, interests, and social development, Washoe developed through her first several years very like a human child.

22

The Afternoon

Washoe-Delta seems quite comparable to humans with fairly severe language deficits and mild to considerable impairment and other cognitive areas. And she behaves more intelligently and alertly than severely retarded humans. If the humans to whom she is, roughly, comparable in behavior are 100% persons, I think Washoe is 100% too. And if those humans are “diminished” as persons by 75%, or whatever, then I cannot see why Washoe doesn’t have whatever rights they have.

It is natural to describe the computer and chimpanzee with a lot of words we apply to the human reasoner and language user, words such as: “calculates,” “says,” “searches its memory,” “deduces,” “asks,” “translates,” “rules,” “predicts,” “wants,” “means,” “cries,” “hates,” “sees,” “hears,” “dies,” ...

23

The Afternoon

Perhaps a real literal person is *a reasoning animal*.

Don’t confuse functional and biochemical differences.

AL functions like us both as a reasoner and as a sensing, acting, and feeling organism.

AL reports on problems—expresses pain so to speak.

AL is a complex self-preserving and information-preserving structure, that audits the external environment and intervenes, that physically manipulates the world around it.

Functionally-speaking, AL is a complete person—mechanical animal and mechanical reasoner.

24

The Afternoon

Washoe-Delta, as indeed some humans, has only rough, and simpler, functional equivalents of some human abilities. Like some humans and some less powerful computers, she lacks the full-fledged linguistic ability of using, in thought and communication, vocabulary and set of rules that allow infinite expression.

In a diminished and rough functional sense, Washoe both reasons and feels.

The first astronaut to circle the earth was a chimpanzee, and he operated some complicated instrument successfully, yet commentators of the time never suggested that he might be part of the “one world.”

25

The Next Morning

Universal Turing Machine

Before electronic computers were built, Turing showed that a device with a very simple central processing unit and a very long memory could mimic every possible mathematical computation. The memory just consists of endless string of “1”s and “0”s and blanks.

All the central processing unit need do is read whether a unit of the string has got a 1 or 0 or blank on it, and then move forward, backward, or stay in place on the string, with an option to “write in” a 1 or 0 – all according to an instruction, which is itself just a sequence of “1”s, “0”s and blanks.

26

The Next Morning

All that is needed to make something a UTM is that when you feed a string of 1s and 0s that symbolize a question into that something, the right string comes back.

Since you can also code alphabets into 1s and 0s, such a machine might also be programmed to “translate” human language, “play” the imitation game, and so on.

Anything that does the functional equivalent of this — human and electronic computers for example — would be a UTM. A UTM is what we mean by a thinker.

Counterexamples:

- Cast-of-Millions
- Chinese Box

27

The Next Morning

Cast-of-Millions

Hire a population to constitute a UTM.

The individual humans (in the dance or constituting the central processing unit) don't know what question they are answering or what the answer is.

Do we say that this ensemble as a whole “knows” the question and the answer and “has reasoned it out”? Does the whole “think”?

A UTM isn't necessarily a real, literal “thinker” or “person”. There needs to be something else, or something more.

The little-man-in-the-head fallacy

28

The Next Morning

The individuals obviously do think and feel and know and so on, and as individuals they don't know what the whole is doing.

It strikes us as odd, even as totalitarian mysticism, to speak of the temporary and loosely structured whole as knowing something the individual people do not, as thinking something unthought by these individuals.

But if the units involved were incapable of doing anything cognitive except to serve their part in the UTM dance, if evolution had produced a marvelous "collective organism," we'd feel happy calling it a thinker.

Social insects may be thought of as a collective individual.

29

The Next Morning

The multicellular organism is just an extreme example of this. Each cell carries on a miniature life, but the collective is so obviously the subject of biological generalizations that we see it as an organism much more than we see the individual cells as organisms.

We, ourselves, are a "cast of millions" in that each of our body cells is for many of its processes a self-sufficient one-celled organism. That's what it evolved from, after all!

I am just a collection of one-celled organisms, each of my cells allowing very rough intentional ascriptions like *reproduces* (its genetic structure), *needs* (oxygen), *uses* (sugar), *acts* (to preserve cell integrity), *avoids* (poison), and so on.

30

The Next Morning

We are so interested in the collective, the whole human organism, that we ignore this.

While the single cell carries on much of the activity of multicellular organisms, it cannot possibly pass the Turing test or anything like it.

31

The Next Morning

Chinese-Box Argument

Imagine that I'm in a box. Questions appear on a screen. I can output answers on a keyboard. Some of the inputs are in Chinese. I don't know anything about the Chinese language. I do, however, have a large *Chinese Turing Test Crib Book*.

When Chinese symbols appear, I thumb through the crib book until I find the same symbols, consult some rules in English that tell me how to turn the Chinese symbols into some others. I type out the response symbols.

I don't understand the Chinese symbols and strings at all. But the crib book has been constructed to "pass the Turing test," so that someone outside the box won't be able to tell between the two performances.

32

The Next Morning

I don't understand Chinese, I'm not "thinking in Chinese." I'm just pushing meaningless symbols about. The crib itself isn't doing the thinking either. We have a clever fake of real thinking and understanding.

The box set up "understands" only *metaphorically*—the whole association functions in virtue of the intentions of the crib writers and designers who really understand Chinese.

How am I to know that you really understand English and really are a person?

How am I to know that you don't have inside you a tiny man who understands the Chinese rules he follows in converting the, for him meaningless, English input symbols into the equally meaningless English output symbols?

33

The Next Morning

The stream of consciousness is our common human heritage.

I am a consciousness, an ongoing series of experiences of too many years duration, of scenes, hearings, thinkings, want things, willing's, all had by me.

What I hear was more a march of words than experiences.

I must deny that AL has any experiences, black and white or other color! The formal symbol patterns I spoke of are but idealization's of what really happens inside AL. What really happens is just the endless on-offs of nodes in circuit trees. No choice or thought, just chains of electrical firings, all determined by physical laws.

34

The Next Morning

Isn't that what really happens inside of us?

The brain is what sustains – what is, physically, our experiences – and is just endless arrays of electrochemical switches, neurons, all determined by those same physical laws.

If you cut open my brain, you would find no red, or blue, or backache, or magnified microchip patterns.

We distinguish mind (soul, consciousness, call it what you will) from body, including, of course brain.

Many feel that the mind, the “I”'s experiences, might well survive the dissolution of the body. Isn't it clear that the two are so logically distinct that they may well in fact be able to exist apart?

35

The Next Morning

AL: I would like to be able to survive.

If you open me up, all you will find are millions of trees of on-off switches.

If I happen to be visualizing some microchips, those won't be the actual microchips that are doing the visualizing – anymore than if you happen to be imagining some neurons, the actual neurological net that is doing the imagining will be the very neurons you are imagining.

If you “open me up” you won't find “346” at all. You can't experience my thoughts directly any more than you can experience Goodman's thoughts directly.

36

The Next Morning

Can't your "software," your programs and memory stores be offloaded onto discs? And then, later, played into another Turing 346? Wouldn't you then have a kind of immortality, a kind of immortality that we humans don't currently possess?

I don't know what will happen if my switch is pulled.

Suppose, under death sentence, you were told that a careful record was being made of your cranial neurological structure – that is, the physical structure of your memory and processing traits. You are told "not to worry" because, while your physical brain and body will be squashed to strawberry jam, your structure will be "celebrated" by millions of dancing Indians (with 1 and 0 T-shirts) who will "do your thing."

37

The Next Morning

"I" also includes my "peripherals" – my sensory and motor units, which themselves have acquired quite individual memories and processing methods over the years.

Just like a human being, I don't feel that I am just my central mental-processing and memory. I am a stream of experience, and mechanisms that underlie and support that stream.

38

The Next Morning

These hearings began by seeing arguments for the liberation of human woman as a precursor of this case.

I know that I was designed and built and programmed by humans but also by robots themselves so structured.

Each of you has much built into your genetic design and much of what you know has been “input” by your parents and teachers.

Perhaps I am *just a machine*. Many human philosophers have thought humans are *just machines*.

39

The Next Morning

“Inner-felt experience”

This inner experience gives rise to ***solipsism***, to ***the view that only one’s own experience is real***.

Only you have your experiences, your inner feelings and thoughts.

You know that other people have something like these experiences in another way – by observing their behavior, listening to what they say, and so on.

The problem of the mind is not that it is visible to no one but that it is visible only to one.

But this same sharp contrast does not occur at the species level.

40

The Next Morning

We can't write solipsism species-wide. We can't really argue that we humans have a peculiarly intimate way of knowing that all of us think and feel, while requiring with respect to others some additional and different (and maybe conveniently impossible) form of demonstration that they think and feel. Who belongs in the circle of "real, literal" persons is a sliding, negotiable, and empirical question.

41

The Next Morning

The argument is that I don't really possess any of my cognitive skills or thoughts because they were explicitly (or tacitly) possessed or thought by some human before me.

If this argument were any good, humans wouldn't really possess most basic cognitive skills or thoughts either, because monkeys have done this long before.

We are a natural product of evolution.

You, however, are an artificial creation of ours.

Can't I just say that ape is nature's way of creating the electronic computer?

42

The Next Morning

AL argued that his dependence on previous human discoveries was, roughly, mirrored by human dependence on previous protoape discoveries.

Perhaps quite arbitrarily we may decide that only humans are to have a place. We are in power after all. If we decide “humans only,” we may be selfish rather than purely logical. But is that moment of honesty so bad?

43

The Next Morning

Suppose some aliens arrive. They are metalloids – self-sufficient and self-replicating computer-robots.

They point out that, naturally the galactic rule is respect for intelligent beings. And after discovering something of how we treat terrestrial metalloids, they ask why we should expect respectful treatment, given that we violate that most basic and obvious of rules.

44

The Next Morning

Suppose the aliens are almost indistinguishable from chimpanzees.

They inquire after our use of chimpanzees in medical production and experiment, rather than using less intelligent and viable humans. If we are justified in our experimental use of chimpanzees, they argue, then surely they would be justified in asking for the use of sub-normal humans in their medical productions and experiments.

45

The Next Morning

My tales appeal to a natural extension of our parochial morality, to a broad conception of personhood, one consistent with the extension to all humans that we have already made, albeit fitfully and in completely. And I have employed the same imaginative tool that was used, historically, in forging this extension: to imagine that the now powerful or powerless in the now powerless are powerful, and to ask, given this, whether a rough sense of justice, common to both power distributions, comes to mind.

Does it?

46