Transformers and Philosophy 2/16/09 2:18 PM Page i

Transformers and Philosophy

Popular Culture and Philosophy®
Series Editor: George A. Reisch

VOLUME 1
Seinfeld and Philosophy: A Book about Everything and Nothing (2000)

VOLUME 2
The Simpsons and Philosophy: The D'oh! of Homer (2001)

VOLUME 3
The Matrix and Philosophy: Welcome to the Desert of the Real (2002)

VOLUME 4
Buffy the Vampire Slayer and Philosophy: Fear and Trembling in Sunnydale (2003)

VOLUME 5
The Lord of the Rings and Philosophy: One Book to Rule Them All (2003)

VOLUME 6
Baseball and Philosophy: Thinking Outside the Batter's Box (2004)

VOLUME 7
The Sopranos and Philosophy: I Kill Therefore I Am (2004)

VOLUME 8
Woody Allen and Philosophy: You Mean My Whole Fallacy Is Wrong? (2004)

VOLUME 9
Harry Potter and Philosophy: If Aristotle Ran Hogwarts (2004)

VOLUME 10
Mel Gibson's Passion and Philosophy: The Cross, the Questions, the Controversy (2004)

VOLUME 11
More Matrix and Philosophy: Revolutions and Reloaded Decoded (2005)

VOLUME 12
Star Wars and Philosophy: More Powerful than You Can Possibly Imagine (2005)

VOLUME 13
Superheroes and Philosophy: Truth, Justice, and the Socratic Way (2005)

VOLUME 14
The Atkins Diet and Philosophy: Chewing the Fat with Kant and Nietzsche (2005)

VOLUME 15
The Chronicles of Narnia and Philosophy: The Lion, the Witch, and the Worldview (2005)

VOLUME 16
Hip Hop and Philosophy: Rhyme 2 Reason (2005)

VOLUME 17
Bob Dylan and Philosophy: It's Alright Ma (I'm Only Thinking) (2006)

VOLUME 18
Harley-Davidson and Philosophy: Full-Throttle Aristotle (2006) Edited by Bernard E. Rollin, Carolyn M. Gray, Kerri Mommer, and Cynthia Pineo

VOLUME 19
Monty Python and Philosophy: Nudge Nudge, Think Think! (2006) Edited by Gary L. Hardcastle and George A Reisch

VOLUME 20
Poker and Philosophy: Pocket Rockets and Philosopher Kings (2006) Edited by Eric Bronson

VOLUME 21
U2 and Philosophy: How to Decipher an Atomic Band (2006) Edited by Mark A. Wrathall

VOLUME 22
The Undead and Philosophy: Chicken Soup for the Soulless (2006) Edited by Richard Greene and K. Silem Mohammad

VOLUME 23
James Bond and Philosophy: Questions Are Forever (2006) Edited by James B. South and Jacob M. Held

VOLUME 24
Bullshit and Philosophy: Guaranteed to Get Perfect Results Every Time (2006) Edited by Gary L. Hardcastle and George A. Reisch

VOLUME 25
The Beatles and Philosophy: Nothing You Can Think that Can't Be Thunk (2006) Edited by Michael Baur and Steven Baur

VOLUME 26
South Park and Philosophy: Bigger, Longer, and More Penetrating (2007) Edited by Richard Hanley

VOLUME 27
Hitchcock and Philosophy: Dial M for Metaphysics (2007) Edited by David Baggett and William A. Drumin

VOLUME 28
The Grateful Dead and Philosophy: Getting High Minded about Love and Haight (2007) Edited by Steven Gimbel

VOLUME 29
Quentin Tarantino and Philosophy: How to Philosophize with a Pair of Pliers and a Blowtorch (2007) Edited by Richard Greene and K. Silem Mohammad

VOLUME 30
Pink Floyd and Philosophy: Careful with that Axiom, Eugene! (2007) Edited by George A. Reisch

VOLUME 31
Johnny Cash and Philosophy: The Burning Ring of Truth (2008) Edited by John Huss and David Werther

VOLUME 32
Bruce Springsteen and Philosophy: Darkness on the Edge of Truth (2008) Edited by Randall E. Auxier and Doug Anderson

VOLUME 33
Battlestar Galactica and Philosophy: Mission Accomplished or Mission Frakked Up? (2008) Edited by Josef Steiff and Tristan D. Tamplin

VOLUME 34
iPod and Philosophy: iCon of an ePoch (2008) Edited by D.E. Wittkower

VOLUME 35
Star Trek and Philosophy: The Wrath of Kant (2008) Edited by Jason T. Eberl and Kevin S. Decker

VOLUME 36
The Legend of Zelda and Philosophy: I Link Therefore I Am (2008) Edited by Luke Cuddy

VOLUME 37
The Wizard of Oz and Philosophy: Wicked Wisdom of the West (2008) Edited by Randall E. Auxier and Phillip S. Seng

VOLUME 38
Radiohead and Philosophy: Fitter Happier More Deductive (2009) Edited by Brandon W. Forbes and George A. Reisch

VOLUME 39
Jimmy Buffett and Philosophy: The Porpoise Driven Life (2009) Edited by Erin McKenna and Scott L. Pratt

VOLUME 40
Transformers and Philosophy (2009) Edited by John Shook and Liz Stillwaggon Swan

IN PREPARATION:

Stephen Colbert and Philosophy (2009) Edited by Aaron Allen Schiller
Supervillains and Philosophy (2009) Edited by Ben Dyer
The Golden Compass and Philosophy (2009) Edited by Richard Greene and Rachel Robison
Led Zeppelin and Philosophy (2009) Edited by Scott Calef
World of Warcraft and Philosophy (2009) Edited by Luke Cuddy and John Nordlinger

Transformers and Philosophy 2/16/09 2:18 PM Page iii

Popular Culture and Philosophy®

Transformers and Philosophy

Edited by

JOHN R. SHOOK

and

LIZ STILLWAGGON SWAN

OPEN COURT
Chicago and La Salle, Illinois

Volume 40 in the series, Popular Culture and Philosophy®, edited by George A. Reisch

To order books from Open Court, call toll-free 1-800-815-2280, or visit our website at www.opencourtbooks.com.

Open Court Publishing Company is a division of Carus Publishing Company.

Copyright © 2009 by Carus Publishing Company

First printing 2009

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher, Open Court Publishing Company, a division of Carus Publishing Company, 315 Fifth Street, P.O. Box 300, Peru, Illinois, 61354-3000.

Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Contents

Making Proper Introductions
JOHN R. SHOOK and LIZ STILLWAGGON SWAN 00

Episode One: We Have Company 00

1. The Changing Shape of Things to Come
   J. STORRS HALL 00
2. First Contact
   JOHN SHOOK 00

Episode Two: Seeing and Believing 00

3. Object Lessons
   KEVIN S. DECKER, KARL ERBACHER, AND GABRIEL RYE 00
4. What Changes when Transformers Transform?
   MICHAEL SPICHER 00
5. In the Eye of the Beholder
   JOSEF STEIFF 00

Episode Three: Mind over Matter 00

6. Can Metal Be Mental?
   MATTHEW PIKE 00
7. Will We Meet Optimus Prime in Heaven?
   M.R. EYESTONE 000
8. Optimus Prime in Therapy
   ERIC SWAN AND LIZ STILLWAGGON SWAN 000

Episode Four: I and Thou 00

9. Morally Responsible Machines
   ROBERT ARP 000
10. Robots in Love
    JAMIE WATSON and ROBERT ARP 000

Transformers and Philosophy 2/16/09 2:18 PM Page vi

Episode Five: Good versus Evil 000

11. Is Megatron Really, Really Evil?
NICOLAS MICHAUD 000

12. Beyond Good? Beyond Evil? Beyond Your Wildest Imagination?
ADAM BARKMAN 000

13. Optimus Prime, Hero for Our Time
COREY NEIL 000

Episode Six: War and Peace 00

14. Megatron, Fascist Philosopher
JOHN SHOOK 000

15. Freedom Is the Right of All Sentient Beings
GEOFFREY ALLAN PLAUCHÉ 000

Author Bios 000

Index 000

Transformers and Philosophy 2/16/09 2:18 PM Page 67

6
Can Metal Be Mental?

MATTHEW PIKE

When we think of minds, or of 'events in the mind', such as thoughts or feelings, we think of something ephemeral, something wispy and hard to grasp. And yet as science has continued to march on, our views about what a mind is are now intimately connected with what goes on in the brain—a purely physical part of the body.

Perception, reasoning, and even consciousness are commonly traced to physical goings-on in our central and peripheral nervous systems. Is it possible then that something made not, like our brains, of carbon-based molecules, but rather of metal, silicon, or some other non-biological material, could possess what we would call mind?

Even our most powerful computers, despite their amazing capabilities, do not have *minds*. They have microprocessors that can do incredible things, ranging from creating immersive games like *World of Warcraft* to performing state-of-the-art climate modeling and DNA analysis. But they do not think for themselves, feel pain, or reflect on their lives. We might wonder then, whether it's even *possible* for machines to actually have thoughts or feelings.

If we encountered machines, say, from the planet Cybertron, that seemed to be having these kinds of experiences, how could we know that they actually had mental lives, and were not just programmed to act *as if* they did? The Transformers provide an excellent example of what machines with minds might be like. Thinking about the Transformers raises some fascinating questions about what it is to be a person and what it means to have a mind. It may also help prepare us to meet something that looks as if it's

Transformers and Philosophy 2/16/09 2:18 PM Page 68

intelligent—whether this is something we humans have built or something that shows up one day on our planet, looking for energon cubes or the All-Spark.

What is it about the Transformers that makes us more inclined to believe that they have thoughts and feelings than that our laptops or cell phones do? What evidence could humans have that Transformers actually do have *minds*? For that matter, can you ever be sure that *anyone* besides you has a mind? The problem presented by this last question, known to philosophers as 'the problem of other minds', is that it seems that any action, facial expression, or language use that another person or machine might exhibit *could* be nothing but the programmed response of an unthinking thing. So how do we tell the difference between something that does have a mind and something that doesn't? To answer this question, we need to have a sense of what a mind is, so that we can know what counts as evidence for or against its presence.

One of the first things that comes to mind (no pun intended) is to say that a thing has a mind if and only if it acts in certain ways. René Descartes, for instance, thought that mechanical things (things without souls, or minds) could do everything that things endowed with a mind can do, *except* for using language and solving unexpected or novel problems. Only a special sort of non-physical, God-given thing could do that.

A very different view of the mind became popular among certain philosophers, known as Logical Behaviorists, in the early and mid-twentieth century. These philosophers maintained that mental things (such as thoughts) just *are* doing certain things in certain circumstances, or being inclined to do do these things.. They argued that being in a certain mental state is nothing except for having a disposition to engage in certain observable behaviors when presented with certain sense input. So, being happy is the same thing as being likely to smile, and say that you are happy when someone asks how you are doing, and have a certain positive tone in your voice when speaking with other people, and so forth. The behaviors that you engage in when something causes you pain is all that pain is. These behaviors can be very complex though, and, as we know, they can be very unpredictable: different people respond differently, and even the same person responds differently to the same thing in different situations.

According to this theory, you know that your roommate has a mind because your roommate *does* certain things, like talking to

Transformers and Philosophy 2/16/09 2:18 PM Page 69

you about his or her favorite movie, choosing what to eat for dinner and preparing it, and saying "Ouch!" when stubbing a toe on the coffee table. In the same way, perhaps we can tell that the Transformers have minds because Bumblebee makes grunts and groans when being captured and tormented by the agents of Sector 7, Starscream repeatedly takes actions to try to gain more power and control, and Optimus Prime asks for help from humans in foiling the plots of the Decepticons. These all seem like behaviors that only things with minds can do.

Testing 1—2—3

As a result of the identification of the mind with behaviors, a test was suggested by Alan Turing (which is now unimaginatively called the "Turing Test") that aims at determining what it would take for an objective investigator to make the judgment that a machine was thinking. The test involves a human's asking questions of another human and of a computer, robot, or some other mechanical being, without knowing which is which (so, no peeking—the idea is to judge on verbal responses alone, as Descartes would have us do, not on looking like a person or otherwise observing them). If the human tester cannot reliably tell which is the human and which is the artificial system based on responses to a series of conversational questions, then, according to Turing the system is said to have demonstrated intelligence, that is, real thinking.

The Turing Test works on the assumption that indistinguishable input and output patterns can be safely presumed to indicate roughly indistinguishable degrees of mind. In other words, if a system acts *as if* it has a mind, then it must have a mind. This test is by no means universally accepted as satisfactory, but as a quick justification of its value, we should note two things. First, we all seem to intuitively employ what amounts to a modified version of the Turing Test when determining whether we are dealing with other minds. This seems especially likely during our childhoods, in which we at some point learn to distinguish between objects and persons, and then devise (albeit unconsciously, most of the time) a series of conditions that a thing must meet in order to be identified as a person. These conditions will be relative to a prototypical person (most likely one's mother or father) that then serves as the point of comparison for all future encounters. Granted, by the time that we have performed this evaluation several hundred times, and have

Transformers and Philosophy 2/16/09 2:18 PM Page 70

noted that each and every "object" we encounter with human shape, appearance, and movement has met the conditions, this process becomes increasingly automatic, but it seems that, at bottom, we are performing something like Turing's test all the time.

Second, Turing competitions are still held annually, because this behavior test is quite a challenge, and the possibility of writing a program that can pass it continues to absorb many an Artificial Intelligence investigator. No machine so far has yet been able to convince anyone that it is a person, based on its linguistic behavior, although they have come closer over time. Although they have become good at chess, and at other "rational" kinds of thinking, no machine so far has convinced humans set on making the distinction that it is in fact thinking. If we don't consider the Turing Test to be a satisfactory test of whether there's a mind, then we ought to be able to point to some other way of deciding that our fellow humans have minds. It seems unlikely that we will be able to come up with a method that does not depend on observing their behavior.

In earlier eras, the most popular approach which was to judge whether something has a mind based simply on appearance. Only beings of human form, and this very narrowly defined, were believed to possess minds. Famous disputes over "changelings" in the medieval period and indeed into the Enlightenment period were waged over whether babies born with severe deformities were in fact human, and the treatment of individuals with such diseases as neurofibromatosis (Elephant Man's disease) reveals that the pervasive attitude was that individuals who did not look sufficiently like the prototype human surely could not think or feel. Remnants of such views indeed still remain, with many people continuing to maintain that primates other than humans, regardless of their problem solving abilities, tool use, or other complex behaviors (never mind animals such as octopuses, that look so very different from us), cannot really be thinking.

Surely at least part of the reason for our willingness to accept that Transformers are thinking beings is attributable to the fact that they do have roughly the same physical shape as us, speak our language, engage in similar interactions, and even appear to have the same emotions as us. Someone even apparently felt the need to give Optimus Prime lips in the 2007 movie so that his facial expressions could more closely resemble ours. Although this is a natural and common association, we can see that physical appearance is the least sophisticated criterion that one might use for determining

Transformers and Philosophy 2/16/09 2:18 PM Page 71

whether something can think. At least observing a thing's behavior broadens the pool of those who might be considered to have minds; under this idea, even entities who do not resemble our human beings could be understood to have minds, if they act as though they do.

The Chinese Room

There seems to be a rather obvious problem, though, with trying to decide whether something has a mind merely by watching the way it behaves. Despite the fact that it hasn't yet been done, can't we conceive that as technology improves, a computer could be specifically programmed to do things that would give every appearance of its having a mind?

If some brilliant and compulsively industrious programmer were to program in every conceivable bit of information, say on a particular topic (this is how the Turing competitions are currently run—the computers being tested are "interviewed" on only a select area of concern, such as baseball, or the stock market, since the more general challenge of attempting to build a machine that can converse intelligently across areas of discourse has been suspended for the time being), and successfully anticipate every question that an examiner might ask it, then the computer would be able to pass the test—but it wouldn't actually be thinking. The reason that we would be inclined to this opinion is that we can see that all this inputting and outputting was being done without the computer understanding what any of it *meant*.

This is the argument made by philosopher John Searle.[1] He illustrated his point by way of a clever thought experiment: he imagined a human being locked in a room where the only opening is a slot through which written messages are passed. The human inside does not speak, read, or understand any components of the Chinese language, and yet is able to successfully process messages passed into the room that are written in Chinese through the use of a collection of Chinese characters (a database) and a rulebook (a program) which, written in a language that the person *does* understand, gives step-by-step instructions for what output to

[1] John R. Searle, *Minds, Brains, and Science* (Harvard University Press, 1984), pp. 28–41. Searle had earlier stated this argument in "Minds, Brains, and Programs," *Behavioral and Brain Sciences* 3:3 (1980), pp. 417–457.

Transformers and Philosophy 2/16/09 2:18 PM Page 72

send back, given any particular input. The rulebook does not in any way reveal the meanings of any of the Chinese characters; it simply specifies which set of symbols, meaningless to the man inside the room, are to be returned, based on the shapes of the written lines, curves, dots, and squiggles.

Searle points out that the room in this thought experiment may be able to pass the Turing Test *without* the human inside understanding any Chinese, or being aware of any of the content of the "discussion" that has taken place. Searle takes the conceivability of this situation to show that no simple rule-governed manipulation of symbols (as a computer is restricted to doing) is sufficient for *meaning*.

It's the lack of *meaning*, of true *understanding*, in this kind of blind instruction following, that accounts for the failure of the machine to think. As Searle famously put it, "syntax is insufficient for semantics." No computer program, regardless of whom it can fool, Searle insists, is sufficient to give a system the *understanding* that is essential to having a mind. So, even though a Transformer might be able to perform all the actions and provide all the responses we might expect from a minded being, that fact would, according to Searle, be no proof that it actually possessed a mind.

The situation that Searle lays out points to one of the reasons why it's difficult to work with the concepts involved in talk about minds: there are two different perspectives from which we look at minds. First, there's the external, third-person perspective from which we see other people. We see the way they behave, hear them give verbal reports about what they are thinking, feeling, and so forth, and from this we come to the conclusion that they have minds. We don't directly experience their minds. All we actually experience are the external results that would come from their having minds (in other words, evidence *consistent* with their having minds).

Even when we examine state-of-the-art brain scan results, we are only looking at neurons in action—we are not looking at the mind of the subjects. All of this is very different from the way we come to know that we have minds ourselves: we seem to know this directly. We have an internal, first-person perspective on our own mental states. We seem to have direct, privileged access to our own thoughts, feelings, desires, hopes, fears, beliefs, and consciousness, in a way that no other person can have. It seems that I cannot experience what it feels like for you to smell a rose, or

Transformers and Philosophy 2/16/09 2:18 PM Page 73

enjoy your favorite Transformers episode. I may have very similar experiences, but they are *my* experiences, not yours.

So, since none of us can *be* a Transformer, we are left with the question of whether we can decide whether Optimus Prime has a mind based on our own view of him (what philosophers call an epistemological question), or based on whether, regardless of whether we can ever know it or not, he has the right internal experience (which is what philosophers call a metaphysical question).

The Spark Within?

As we have seen, Descartes thought that only beings with minds could perform certain kinds of feats, including using language and solving novel problems. In order to rise above the capabilities that simple mechanical nature accounts for, to the level of language use and reason, Descartes thought that we must have some special element bestowed by God—an immaterial mind, or a soul.

In the opinion of Descartes, Searle, and countless others, it's not enough to behave in certain ways because what makes us willing to say that a human or a Transformer has a mind is the idea that something is going on *inside* of them. When we see someone in pain, we observe their behaviors, such as taking pain medication, or pulling their hand away from a hot stove, and from this we infer that they are *feeling* pain. And we seem to know from our own internal experience that there is a host of internal feelings that go along with our behaviors. When you see someone you are in love with, you do not just automatically start behaving in certain ways; rather, you have thoughts and powerful feelings, and these thoughts and feelings *cause* you to act in the ways that you do.

One of the reasons that Behaviorism became popular is that it seemed to offer an appealing alternative to the view put forth by Descartes, which came to be known as Substance Dualism. The problems with Substance Dualism were many, and were noted from the moment, back in the early seventeenth century, when Descartes published his *Meditations on First Philosophy*. For one thing, there appears to be no way for mind, a completely non-physical thing, to interact with physical things (like one's own body), with which it has absolutely nothing in common. For another, there appears to be no way for one to get a mind other than to have it bestowed on him or her by God. This was objectionable for many thinkers in the scientifically revolutionary late

Transformers and Philosophy 2/16/09 2:18 PM Page 74

nineteenth and early twentieth centuries, when the theory of evolution, along with developments in mathematics and physics, were becoming popular.

To some philosophers, Behaviorism looked like an improvement over Descartes, but the problem with Behaviorism is that it leaves out altogether the immeasurably important internal mental states that we all recognize in ourselves. By defining mental states as nothing but behaviors and dispositions to behave, it completely failed to recognize the internal thoughts, feelings, and beliefs that many see as central to the nature of mind.

These apparently non-physical elements that we tend to attribute to minds lead many people to think of something very similar to a "soul", regardless of their enthusiasm for the developments of science. According to this common view, there is something else besides the physical body that composes the individual, and it is that which makes people different from inanimate and mindless things like tables and chairs.

This special internal something is a candidate for what the Transformers have that our current computers do not. The 2007 movie, *Transformers*, seems to portray something similar to this special something. When the All-Spark inexplicably brings cell phones, steering wheels, and vending-machines to life with a single *zap,* something happens. What is it, and how does it happen? Just what does the *zap* do that instills non-living, non-mental gadgets with life and the will to fulfill something like personal desires or goals? Is it some kind of divine magic that gives the Transformers souls and minds, or is it some combination of energy and programmatic information that accomplishes this feat? Whatever it is, the All-Spark obviously imparts something more than being able to perform as the actor in Searle's Chinese Room does. Once hit with that force, machines are able to 'think for themselves'.

Physicalism

Some of us would prefer the explanation that the All-Spark imparts some combination of programmatic information and physical energy to get things started to the idea that it bestows something like a soul on the machines it empowers, because at least with the first option there is some hope of finding an explanation for how minds arise. In the second case, the existence of a mind remains a mystery, something about which no more can be said. For that rea-

Transformers and Philosophy 2/16/09 2:18 PM Page 75

son, many of us tend to think that the view known as "Physicalism" makes better sense as an explanation of what minds are than does the Cartesian soul interpretation of mind.

According to Physicalism, having a mind is *nothing but* having the right kind of physical system (which in our case is our brain, together with other parts of our central nervous system, extending to our sensory receptors). This Physicalist view comes in many different forms. For instance, some philosophers who hold this view think that it doesn't matter so much what the physical system is made of, or how it is set up, but rather *that* it is set up so that it performs certain functions. According to these philosophers, it is the function that matters, not the particulars of the physical system. So, whether it is instantiated, or "realized", in organic matter, silicon, or some as-yet-undiscovered substance, as long as it does the jobs that minds do (believing, perceiving, intending), it is a mind.

Others, though, such as John Searle, think that only biological systems could ever develop minds. For those in Searle's camp, it's the organic nature of biological beings (having evolved from and in interaction with their environments), and the very chemistry of the nerve cells themselves, that is necessary for the emergence of mind. So, while on Searle's own view Transformers couldn't have minds, on other views similar to his, it might be possible to say that as long as they were in some sense biological beings, they could have minds. The mere fact that they are not organic, not having evolved on our planet, wouldn't automatically preclude their being biological entities, one might argue. Some physicalists, though, disagree with both these claims, and maintain that neither is there anything magical about biology, nor is function alone sufficient to identify mind. They think that Searle's biological account is too specific (and thus chauvinistic toward other kinds of beings), while functionalism isn't specific enough.

This last group of physicalists, (known as reductive physicalists), think that our scientific research into the way that the brain and nervous system work can eventually provide a complete explanation for our mental life, and that there is nothing else needed, because physical explanations can tell the whole story. They maintain that biology is reducible to physics and chemistry, and functionalism only points to something it could label as mind, but does not explain what makes it function the way that it does. These philosophers lean heavily on developments in the last decade or two in neuroscience, which have been impressive, in the way in

Transformers and Philosophy 2/16/09 2:18 PM Page 76

which they have come to show how certain processes in the brain allow for perception, memory, association, emotion, and other mental activities.

Perhaps unsurprisingly, though, some philosophers want to have their cake and eat it, too. They think *both* that minds depend upon brains, *and* that mind is irreducible—that it is something purely subjective, something decidedly different from the physical. While there might be neural *correlates* of mental activities, these philosophers say, those correlates are not to be *identified* with mind. Nevertheless, these philosophers maintain, there is no second substance, no soul or mental 'stuff' in addition to the physical stuff of the universe. So they remain physicalists.

What the physicalist philosophers of mind all have in common is their faith in the value of physical explanations and in the achievements of science. Science has proven to be tremendously successful at unraveling apparent mysteries and discovering straightforward (although not always simple) explanations for the things being investigated. Research into the functioning of the brain has provided explanations for innumerable questions about how perceptual processing occurs, about how different kinds of perceptual processing is integrated, and even about how consciousness—or mind—can arise.

If a physicalist view is correct (with the possible exception of the non-reductive physicalists, who will have to determine mind on other grounds), then the question concerning whether Transformers are mental beings is actually just a question concerning the specifics about what constitutes their information systems, how the different mechanisms for perception, self-regulation, and self-movement are organized, how they interact with each other, and how they are tracked by further mechanisms of the same nature. If a 'brain' that is sufficiently functionally similar to ours exists in Transformers, then there's no reason not to credit it with having a mind.

And so, if an entity is made out metal, silicon chips, or something completely unknown to us is hooked up in the right way to whatever perceptual, visceral, and posture and movement systems it has, and if that entity is capable of processing all this information, as well as the information *that* it is doing this processing then we might be willing to accept that that this entity had a mind. But would it be a mind *like ours*? Well, that would depend on what you think 'like ours' means.

Transformers and Philosophy 2/16/09 2:18 PM Page 77

Many animals on this planet have minds, at least on the views of lots of people, but they differ from our human minds, given, for instance, that they don't encode information into language and use that extensively in their thinking, as we do. But then many people (look at babies!) don't do that, either, and yet we think of their minds as being like ours. So, one might say that Transformers, given the conditions described above, do indeed have minds, both as similar to and as different from our minds as our minds are to each other's, and to those of other animals with which we are familiar.

So if minds are just some kind of physical information processing system, then why do Transformers have mental lives and our computers (so far) do not? The answer that many physicalists would give is that the difference is just a matter of complexity and function. Our current computers are nowhere near as complex and capable as the human brain, with its average 10^{11} (one hundred billion) neurons, each of which in turn possesses an average seven thousand synaptic connections to other neurons. This means that by some estimates, the average human brain at its peak (our brains begin to degenerate after we are about three years old) has about 10^{15} (1 quadrillion) synapses. Add to that the over four hundred chemical transmitters, peptides, hormones, and large variety of other modulating chemicals that can radically influence the environment in synaptic gaps, and you can see that computers are not even on the same playing field as brains. If we eventually manage to build computers with enough complexity, and with the right kind of structure, they may well turn out to have minds too—according to some of the physicalist views.

Mind—More than Meets the Eye?

If some version of physicalism is right, then there may come a day when our scientific understanding has advanced to the point that we have identified which structures in the human brain are responsible for our mental experiences, and it may be that from this knowledge we can identify which attributes a computer or robot would have to possess in order to have a mind. In this case, definitively identifying whether the Transformers have minds may be as simple as a scan by a hand-held gadget. Until that time, however, or if the most straightforward reductivist version of physicalism is wrong about what constitutes the mind, then we must look for

Transformers and Philosophy 2/16/09 2:18 PM Page 78

other ways to decide which entities have minds and which don't. So, if you can't know that an entity has a mind by looking at the way it operates (and we can't know that even about human beings, with our current state of knowledge about the brain), how *can* you know whether it does?

It seems that the best candidate for an approach to making this determination may be what philosophers call an abductive inference, or an inference to the best explanation. While this kind of an inference does not provide us with complete confidence that we have reached the truth, it can provide pretty good reason for believing that something is true, and so could justify our belief that the Transformers have minds. This kind of reasoning generally starts from a set of observed facts about the world, and then tries to identify the best theory or explanation of why they occurred. It is a way of reasoning from effects back to causes. Frequently, many different explanations are possible, and so we tend to try and pick the best of them, using several different criteria. We all do this constantly throughout the day, even if we are not consciously aware of it. If you're walking along and encounter a big hole in the ground, you will likely start thinking up possible explanations for it. Maybe this is a construction site, where the hole was made in preparation for a building foundation. Or maybe the hole is a crater that resulted from a Transformer who crash-landed following a journey from Cybertron. Or maybe the hole happened because in exactly this spot, gravity completely ceased to function and the dirt in this specific area all floated into space. All three of these scenarios can explain the observed hole in the ground, but some of them seem less probable than others. We tend to think that the building construction hypothesis is a simpler, less outlandish, explanation that fits better with our other observations in the past, and so is more likely to be true.

Lacking definitive evidence that others have minds, we presumably do something similar to the process illustrated in the example above. If we encounter a robot performing fairly basic tasks, we may formulate several theories about how it does them. We may theorize that the robot is being remote controlled, or that it has been programmed to autonomously perform certain functions, or that it is a robot with a mind. What observations could we make that would lead us to conclude that it actually has a mind? We will have to base our conclusion on the external evidence that we can observe, as well as our past experiences with similar things.

Transformers and Philosophy 2/16/09 2:18 PM Page 79

Given that all of our past experiences with mechanical devices have led us to believe that they do not have minds, and being somewhat familiar with the current state of technology available on planet Earth, we would sensibly conclude that the robot was either being controlled by someone, or had been programmed by someone to perform specific actions. But what if it engaged in behaviors that were very different from the other kinds of machines that we had encountered? What if the robot started up a casual conversation about the weather, and started to complain about how stiff its robotic joints were? This could just be a programmed speech, or even a recording. But what if we started to ask it questions on a wide variety of topics? As the robot's behaviors become more and more complex, adaptive, and versatile, it seems that we might become less and less convinced that it was just a preprogrammed machine that was only following the instructions it had been given.

What if there were some reason to think that the robot had originated on a different planet, in which case there might have been a considerably more advanced level of technology available for its construction? This is one of the important differences in encountering a Transformer versus a human-made system, because, if there is good reason to think that the robot is not limited by the present state of our technology, then the possibility that it could be sufficiently complex to have a mind is no longer effectively ruled out. Observing the robot doing very complex things that are sufficiently different from the machines we routinely encounter (such as smoothly transforming from an automobile into a roughly-human shape) will further support this possibility. After an extended period of observing the robot, and seeing it engage in things like learning, attempting to avoid hazards, pursing preferences and goals, making plans to defend against other robots, mourning the loss of its fallen comrades, and making decisions based on what seem to be moral values, we could very easily come to the conclusion that the behaviors it exhibited were too complex to originate from anything without a mind.

It would be nice if the determination of whether something had a mind could be made simply by comparing its actions and characteristics to a check-list of conditions that are both necessary and sufficient for mind. It seems though, that there is a very wide range of characteristics that beings with minds can have, but don't necessarily have. Our understanding of mind is a cluster concept, and some but not all of the characteristics included in that concept

Transformers and Philosophy 2/16/09 2:18 PM Page 80

belong to anything that we would be willing to call a mind. Persons with certain neurological disorders, for example, can navigate and catch a ball, but they insist that they cannot see; others can function normally in most contexts, but cannot process language. They know what they are doing and show purposive action and even normal intelligence, but they cannot communicate via language. Other kinds of brain damage result in some people's having a complete inability to remember things for more than a minute, while others smell colors or taste sounds. Some people hear music when none is present, while still others can hear individual notes, but cannot put together a melody. In none of these cases do we say that the people in question do not have minds; rather, we say that they have deficiencies or gifts. Identifying the key marks of the mental is thus surprisingly difficult.

While not so many people these days are likely to insist that having a mind is somehow tied to human form, the other proposed criteria seem to also run into problems. We saw already that the Turing Test and its emphasis on language use does not seem to be as good of a criterion as first thought. And as computer programming becomes more advanced, many other suggested test criteria will likely also be met by unthinking robots. This leaves us in the unfortunate position of having to admit that specifying exactly how we decide if other things have minds or not, may not be possible.

Perhaps the best we can do is give a list of things that *some* things with minds do, and if we encounter something that does *enough* of them, then we should be willing to conclude that it does have a mind, although perhaps of a fairly different kind than ours. But having enough of these properties would not guarantee that the thing in question has a mind and can *actually think*. Advanced robots could be programmed to mimic human facial expressions, behaviors and actions, and could *seem* to be thinking, feeling, and acting like humans do, without their ever having minds of their own. Since (at present) we can only observe other people and things "from the outside", there does not seem to be any way to be *sure* that even our friends and family have minds of their own!

It does seem, though, that even if we could never be *completely* sure that a Transformer had a mind, we could have *as much reason* to think that the Transformer had a mind as to believe that any other human being had one.