

# God, Souls, and Turing

In Defense of the Theological Objection to the Turing Test<sup>\*</sup>

Selmer Bringsjord  
Department of Cognitive Science  
Department of Computer Science  
Rensselaer AI & Reasoning (RAIR) Laboratory  
Rensselaer Polytechnic Institute (RPI)  
Troy NY 12180 USA  
`selmer@rpi.edu`

version 5.19.09

## Abstract

It's widely known that when Turing first introduced his "imitation-game" test for ascertaining whether a computing machine can think, he considered, and found wanting, a series objections to his position. It seems safe to say that one of these objections, the "Theological Objection" (TO), was regarded by Turing to be positively anemic, and that ever since he delivered his rapid (purported!) refutation over half a century ago, the received view has been that, indeed, this objection *is* as weak as can be. The present paper is devoted to showing that TO is not the pushover Turing, and others since, take it to be.

---

<sup>\*</sup>I'm greatly indebted to AISB for allowing me to present at the University of Reading symposium on Turing and his test; the presentation gave birth to this paper. I owe a special debt in this regard to Mark Bishop, who arranged all the specifics of my participation, provided invaluable guidance, and showed supernatural patience in waiting for me to get this paper done. Thanks are due as well to local host Kevin Warwick, whose student Huma Shah was invaluable, to John Barnden, Chair of the AISB, and to my colleague Jim Fahey for helpful insights.

# 1 Introduction

I've been turning over in my mind, for more years than I care to admit, many facets of Turing's brief but remarkably fertile 1950 *Mind* paper (Turing 1950), in which he sets out and defends what is now known far and wide as the *Turing Test* (= TT). Despite all this reflection, and a number of publications that have resulted from it (e.g., Bringsjord 1995, Bringsjord & Noel 2002), I confess to having never directed my attention, in earnest, to the *Theological Objection* (TO) within it, and specifically to Turing's reply to this objection. Along with, I suspect, many other thinkers, I had assumed this objection to be too weak, and too easily dispensed with by Turing, to merit careful, time-consuming analysis. I now see I was wrong; this paper, devoted as it is to this objection, is intended to show why.

## 2 Preliminaries

We need some machinery to get smoothly off the ground.

I shall assume that Turing's main claim in "Computing Machinery and Intelligence" is  $\mathcal{T}$ :

If  $x$  is a TT-passer, then  $x$  thinks, or is a thinking thing.

We could of course spend much time discussing whether this ought to be regarded the main thesis of the paper; I have engaged in that discussion elsewhere (e.g., see Bringsjord 1995). But given that our focus is to be on a particular objection, and that that focus will require considerable space, a luxurious treatment of Turing's central claim is impracticable. I therefore simply make two short points.

First, I'm well aware of the fact that some will be inclined to say that Turing didn't have in mind a test the passing of which implies that the passer *in fact* thinks, but instead a main thesis such as this proposition ( $\mathcal{T}'$ ):

If  $x$  is a TT-passer, then  $x$  should be *considered* a thinking thing.

The chief problem with this interpretation is that even if one knows who is supposed to be doing the considering here (e.g., perhaps an ideal observer wholly free from bias?), a rigorous assessment of  $\mathcal{T}'$  will require that we

have on hand an epistemology explaining under what conditions one has an “epistemic obligation” to assent to some proposition  $\phi$ .<sup>1</sup>

The second point: Some will say that Turing meant to outright jettison such terms as ‘thinks’ and ‘thinking’ from the discussion, and thus to read him as advancing  $\mathcal{T}$  — a thesis that explicitly deploys the concept of thinking — is an error, and an ironic one at that. The problem with this response is two-fold.

One,  $\mathcal{T}$ , if affirmed, *would* in a meaningful sense replace “think” terminology. The sense is that such terminology standardly involves all sorts of concepts far removed from the steadfast empiricism Turing clearly takes himself to be promoting via the imitation game. Bound up with the ordinary, intuitive sense of ‘thinks’ are notions of consciousness, and indeed any number of internal, invisible, Cartesian properties. Given the truth of  $\mathcal{T}$ , all such notions are supplanted with a thoroughgoingly empirical outlook: You give a mechanical test, check the results, and if you observe a sufficiently high “convince rate,” the TT-passer is declared to be a thinking thing; end of story; no philosophizing about mysterious qualities having detained us. It seems quite reasonable to hold that we have here captured Turing’s perspective and position.

The second problem infecting the complaint that  $\mathcal{T}$  contains precisely what Turing desires to defenestrate is that, given Turing’s reply to the “Objection from Consciousness” in (Turing 1950), there is no denying his position to be that if the right verbal behavior is directly observed to be coming from the computer, we must regard it to be the bearer of *whatever* is denoted by such terms as ‘conscious’ or ‘thinks’ — which is to say that his position is in the end none other than  $\mathcal{T}$ .

To complete our preliminary work before turning in earnest to TO, it’s necessary to take account of one additional factor. It is that, in my opinion, as I rather long ago conceded (e.g., in Bringsjord 1992), there can be little doubt that TT will eventually be passed. That is to say, there is no question that, sooner or later, a suitably programmed computing machine, operating, mathematically speaking, at the level of a Turing machine (or conceivably below that level), will pass TT, where the time of the test is some reasonable

---

<sup>1</sup>I happen to have on hand such a thing, actually (viz., a modernization of Chisholm’s theory of knowledge (1966), as set out and implemented in the Slate system of Bringsjord et al. (2008)), but to invoke the scheme here would be beyond the scope and space of the present paper.

period of time  $\tau$ .<sup>2</sup> It might not happen within ten years, or fifty, or even one hundred, but surely at *some* point clever programming will yield a computing machine able to pass TT.<sup>3</sup>

Given this apparent near-inevitability, rigorous representation of the situation before us begins to suddenly look rather more challenging than one might have hoped at the outset. This is so because we have yet to invoke some formal language that would allow us to speak carefully about TT and temporal matters, and were we to do so we would probably need to employ representational machinery above what would be the norm in such dialectical contexts as that enveloping the present paper. For example, it would probably not sit well with readers if at this point I announced that to achieve precision in the debate into which we are about to plunge it's necessary to invoke the machinery of quantified tense logic.<sup>4</sup> Straightforward use of first-order logic, which is what *would* be the norm for attempts to analyze TO and related matters, allows us to represent  $\mathcal{T}$  as

$$\forall x(PassTT(x) \rightarrow Thinks(x)),$$

and my optimism about the future becomes simply

$$\exists x(PassTT(x) \wedge Computer_{TM}(x)),$$

where  $Computer_{TM}(x)$  iff  $x$  is a computational device operating at the level

---

<sup>2</sup>The adjective ‘reasonable’ is, yes, disturbingly indeterminate. Clearly Turing’s intention is that TT should be possible to administer in, if you will, “one sitting,” or “one session.” If one puts only a finite bound on  $\tau$ , identifying it with some natural number that denotes, say, number of hours, then some will insist that questions like “Can you please write a novel for me?” are fair game. But novels generally can’t be written and examined in one sitting. The human software engineers striving to build a TT-passer shouldn’t be saddled with having to figure out how to get a computer to write a novel in order to do their job. (Nonetheless, I have elsewhere discussed the possibility of a TT-like game that does involve a challenge to the computer to produce compelling *short* fiction; see Bringsjord 2000.)

<sup>3</sup>While not necessarily constituting *strong* evidence in favor of the inevitability of a TT-passing computer program arriving on the scene sooner or later, surely the fact that the form of TT in the Loebner Competition (the 2008 rendition of which was held in conjunction with the symposium at which I presented the kernel of the present paper) was nearly passed in 2008 provides significant evidence.

<sup>4</sup>A brief but informative and elegant overview of temporal logic can be found in the commendable (Goble 2001).

of a Turing machine.<sup>5</sup> Such a formula seems to be cast at about the right level of expressivity. But if time must be taken very seriously in the debate, and we pass to temporal logic, my optimism would be plausibly represented instead as

$$\mathbf{F}\exists x(PassTT(x) \wedge Computer_{TM}(x)),$$

Here, the operator  $\mathbf{F}$  can be (provisionally) read as ‘It will at some time be the case that ...’<sup>6</sup> If we started down the road to such fine-grained representation, we would soon find ourselves required to settle all sorts of niceties in order to adjudicate the central matter before us (viz., whether the theological objection succeeds). Such settling is beyond our scope, so, summing up, the final preliminary point is that our analysis will indeed for the most part be couched in nothing fancier than first-order logic.

Preparation now complete, we turn to TO.

### 3 The Theological Objection

As is well known, after asserting  $\mathcal{T}$ , Turing considers and replies to a series of objections; the first happens to be TO itself, which runs, verbatim, as follows.

Thinking is a function of man’s immortal soul. God has given an immortal soul to every man and woman, but not to any other animal or to machines. Hence no animal or machine can think. (Turing 1950, 433)

It’s easy enough to unpack this laconic objection into a first-order proof  $TO^p$  with relation symbols that correspond in the obvious way to the natural-language adjectives and verbs in the just-quoted passage (and with use of specific rules in some standard deductive calculus left to the reader to select):

---

<sup>5</sup>There are all sorts of information-processing machines more powerful than Turing machines. (For examples, see Bringsjord & Arkoudas 2004 and Siegelmann 1995.) At least for now, we can leave aside consideration of these devices.

<sup>6</sup>Since it’s doubtful that, *at present*, there is on our planet a computer which will eventually pass TT, it would be unwise to instead go in the direction of  $\exists x(\mathbf{F}PassTT(x) \wedge Computer_{TM}(x))$ .

## TO<sup>p</sup>

∴	(1)	$\forall x(Thinks(x) \leftrightarrow Soul(x))$	premise
∴	(2)	$\forall x(Machine_{TM}(x) \rightarrow \neg Soul(x))$	premise
∴	(3)	$\forall x(Machine_{TM}(x) \rightarrow \neg Thinks(x))$	(1), (2)
∴	(4)	$\neg \exists x(Machine_{TM}(x) \wedge Thinks(x))$	(3)
∴	(5)	$\exists x(Machine_{TM}(x) \wedge PassTT(x))$	premise
∴	(6)	$\neg \forall x(PassTT(x) \rightarrow Thinks(x))$	(4), (5)

As this is indeed a certifiable proof, the issue of course is whether the premises of the argument are true. I have already laid my cards on the table with respect to (5): It seems to me quite undeniable that TT will eventually be passed; and in my experience, while there is considerable disagreement as to *when* this passing is likely to occur, there is widespread agreement that eventually it will indeed come to pass.

Two premises, (1) and (2), therefore remain for discussion. In his reply, Turing targets (2); to that reply we now turn.

## 4 Turing's Reply

Turing's reply begins with his confession that he is "unable to accept any part of this, but will attempt to reply in theological terms." I assume this means that instead of attacking directly, by for example arguing that God simply doesn't exist, he intends to refute TO under the supposition that God does in fact exist. Turing then says that he finds "the orthodox view" arbitrary, because a member of some other religious community would find it unpalatable. By 'orthodox view' I assume he means something like the orthodox Christian view: minimally, that the God of Judeo-Christian monotheism exists, and human persons, as described in both the scriptures and orthodox doctrines derived therefrom, have physical bodies, and are immortal, thinking souls. This interpretation is supported by the next sentence: "How do Christians regard the Moslem view that women have no souls?" For Turing here seems to be making the trivial point that even some who in general accept the existence of a supernatural realm reject the orthodox Christian view that you and I have souls, and are immortal.

After his preliminary comments, Turing clearly sees himself as getting down to the business of eviscerating TO:

But let us . . . return to the main argument. It appears to me that the argument quoted above implies a serious restriction of the omnipotence of the Almighty. It is admitted that there are certain things that He cannot do, such as making one equal to two, but should we not believe that He has freedom to confer a soul on an elephant if He sees fit? . . . In attempting to construct [machines suitable for God to confer a soul upon them] we should not be irreverently usurping His power of creating souls, any more than we are in the procreation of children: rather we are, in either case, instruments of His will providing mansions for the souls that He creates. (Turing 1950, 443)

However one does the exegesis, it seems undeniable that, with respect to the proof-theoretic version of TO laid out above, only premise (2) is targeted by what Turing says here; premise (1) appears to be conceded. That is, Turing appears to tacitly admit in this exchange that thinking is indeed a function of man’s immortal soul. The idea behind the attack on (1) would presumably consist in Turing’s giving a counter-example — in which (1)’s antecedent is satisfied, but not its consequent. Thus he asks us to imagine an elephant that receives a soul from God, in order to support the claim that, contra (2), a computer (at the level of a TM) can have a soul as well. The inference appears to be that since such an elephant is possible, by analogy so is a God-ensouled computer, and such a computer is intended to be a counter-example that overthrows (1).

Let’s now see why Turing’s reply is unsuccessful.

## 5 Refuting the Reply

The first fatal problem infecting Turing’s reply is that he seems to conceive of souls rather like a finger or a mustache, or even like property outside our bodies that we might own, such as a shoe, a vehicle, or a computer. I have two index fingers, yes; I could make it true that I soon enough have a mustache; and I have a number of computers, all of which are a good deal less intelligent than Turing (1950) predicted they would be,<sup>7</sup> and one of which I’m typing into now, at 37,000 feet on the way to Japan, well aware of the fact that this paper is overdue. But when by the orthodox view it’s said that “Selmer has a soul,” something very different is asserted: something in line

---

<sup>7</sup>He predicted in (Turing 1950) that by 2000 a TT-passing computing machine would arrive.

with what any number of illustrious defenders of the “soul persuasion” have taken themselves to be holding.

The list of soul-persuasion thinkers is of course quite lengthy; it includes Jesus, St. Paul, Augustine, Descartes, Leibniz, Pascal, Bolzano, and — more recently — Geach, Chisholm, and Swinburne. One commonality among those on this list is the idea that when it is said that Selmer “has” a soul, this is elliptical for saying that Selmer, fundamentally, is a rational, thinking thing, a mind, non-physical in nature, but nonetheless connected to extended objects (e.g., Selmer’s brain) in various ways, and such that he can be can shed his current terrestrial body in favor of *reconnection* to another one in the afterlife.<sup>8</sup> There is insufficient space for a sustained and systematic survey of how thinkers like those just listed understand what Turing calls “the orthodox view.” We shall need to rest content with painfully brief references to some of the proponents of the view, and a few germane quotations.

Given that Turing evidently takes the “orthodox view” underlying TO to be the Christian one, it should come as no surprise that the originator of this view would appear to be Jesus himself. For example, the following is recorded in Matthew 16:

“If anyone would come after me, he must deny himself and take up his cross and follow me. For whoever wants to save his life will lose it, but whoever loses his life for me will find it. What good will it be for a man if he gains the whole world, yet forfeits his soul? Or what can a man give in exchange for his soul? For the Son of Man is going to come in his Father’s glory with his angels, and then he will reward each person according to what he has done.” (Matt 16: 24–27; NIV)

One doesn’t need to be a systematic theologian to grasp that part of Jesus’s point is that if Jones is a follower of Jesus, then he, himself, that is, Jones, qua soul, will persist into the afterlife and be rewarded there. There are any number of additional passages in which what Jesus and other New Testament writers (e.g., St. Paul<sup>9</sup>) say commits them to the orthodox view

---

<sup>8</sup>We shall surely have to leave aside coverage of the various proposals for the nature of the “connection.” Descartes, as is widely known, was a proponent of *interactionism* (for a defense of this view in the 20<sup>th</sup> century, see e.g. Popper & Eccles 1984), while Leibniz rejected this view in favor of a “connection” based in his famous doctrine of “pre-established harmony.” We shall also need to leave aside longer analysis of the difference between *having* a soul and being a soul. This issue, and others, is discussed in a paper by Philip Quinn (1997).

<sup>9</sup>E.g., most prominently, his first epistle to the Corinthians, chapter 15.



— a view in which, again, one doesn't have a soul as one has a nose, but in which one *is* a soul that is embodied and can be re-embodied.

Here is Leibniz in *Monadology*, a work that can be viewed as a philosophically sophisticated statement and defense of (among other things) the orthodox view, expressing the position that we are souls of a special sort:

[T]here is, as far as minds or rational souls are concerned . . . , this thing peculiar, that their little spermatic progenitors, as long as they remain such, have ordinary or sensuous souls, both those of them which are, so to speak, elevated, attain by actual conception to human nature, and their rank of sensuous souls are raised to the rank of reason and the prerogative of minds. Among the differences that there are between ordinary souls and spirits, some of which I have already instanced, there is also this that, while souls in general are living mirrors or images of the universe of created things, minds are also images of the Deity himself or of the author of nature. They are capable of knowing the system of the universe, and to imitate it somewhat by means of architectonic patterns, each mind being like a small divinity in its sphere. (Leibniz 1991, 269–270)

Turing's elephant is thus quite beside the point when it comes to the orthodox view. Were Leibniz here to defend TO against what Turing says, it's clear he would simply point out that as thinking things we *are* souls (of an elevated sort), and that for God to ensoul a body is to bring into existence a thinking thing that now has use of that body.

It's easy enough to focus the problem Turing faces, by returning directly to premise (2). This premise encapsulates the view that as a simple and mere mechanical device, as nothing more than a body, a computing machine is not an incorporeal and everlasting thinking thing (= not a soul). It does no harm to this view whatsoever to imagine God connecting a soul to a body (whether that body is elephantine, pongid, or humanoid). If instead of the relation symbol *Soul*, we talked of *being connected to a soul*, symbolized as *C2Soul*, and asserted a variant of (2), for example

$$(2') \quad \forall x(Machine_{TM}(x) \rightarrow \neg C2Soul(x)),$$

then, yes, Turing's imagination would have traction. But of course TO has (2) as a premise, not (2').

We come now to the second problem plaguing Turing's response to TO.

The second problem is that in his response to TO Turing promotes a radically deflationary view of the contribution of human engineering to the production of TT-passing machines. He does so by granting premise (1) and giving God, not the human engineers, credit for bringing a soul into the picture. Premise (1), recall, captures “thinking is a function of man’s immortal soul,” which Turing appears to grant for the sake of argument. As to credit for souls arriving on the scene, there can be little doubt (as our above analysis has made clear) that God, in the scheme Turing describes, is to get that credit. But then since thinking is a function of what *God* does, what does the engineer do? No much. Turing’s response to TO entails the evisceration of the fields we today call ‘artificial intelligence’ (AI) and ‘artificial life.’

This is easy to see. It’s not unreasonable to hold that driving is a function of the cooperation of our visual system and some key effectors (hands, arms, and feet). Let’s call this composite system simply *D*. Clearly, driving is by definition a function of man’s *D*. Now, imagine that some AI team has been working for years to engineer a robot able to automatically drive a car on our roads. While they have made considerable progress, their ultimate objective appears to be many years off, at best. But then suddenly God shows up, takes a look at one of the inadequate robots produced by our team, and magnanimously imbues this robot with its own system *D*, which immediately enables the robot to be a superb driver. Suppose in addition that the leader of the engineering team, upon witnessing the marvelous performance of the formerly inept robot, declares: “Colleagues! We have succeeded! We have reached our engineering goal, and we will be heroes!”

While it may indeed happen that these engineers become heroes, it is certainly not the case that they succeeded. While perhaps it can be truly said that this engineering team provided a “mansion” for the receiving of *D* from God, their purpose was to engineer *D* itself. Likewise, AI engineers who only provide a mechanical “mansion” for thinking have not succeeded, and hence in his response to TO Turing concedes all that the proponent of this objection could wish for.

The flaw in Turing’s response can be expressed in more general terms, as follows. Consider the case not of someone specifically aiming, via computational techniques, to engineer a robdriver or a TT-passer, but rather of someone *S* aiming to produce a computer program  $\mathcal{P}$  that computes some number-theoretic function *f*. (In this scenario, it isn’t just a computer program that would count as success: any number of other things to compute

$f$  would be sufficient — an algorithm, a Turing machine, a register machine, and so on. But we shall speak specifically of computer programs for convenience.) To make things concrete, let's suppose that  $f^*$  is the particular function which, when applied to the Gödel number  $n^\phi$  of some formula  $\phi$  in the propositional calculus, returns 1 if  $\phi$  is provable, and 0 if it isn't. And let's suppose as well that the particular program  $\mathcal{P}$  produced by  $S$  is one that *only* renders a verdict as to whether or not  $\phi$  is grammatically correct; that is, whether or not  $\phi$  abides by the grammar for the propositional calculus. In this situation, does  $S$  get credit for having produced a computer program able to decide theoremhood in the propositional case? Obviously not. Moreover,  $S$  is no better off if some other person  $S'$  shows up and hands to  $S$  a semantic tableau algorithm, and corresponding code, to compute  $f^*$ , where one early step in this algorithm is a check that, where  $n^\phi$  is some input,  $\phi$  is at least a well-formed formula, and hence at least a candidate for being provable. Yet by Turing's logic in response to TO,  $S$  has succeeded because of what  $S'$  provides.

## 6 Conclusion

I conclude, then, is that Turing's response to TO fails. Note that my conclusion is *not* that TO is victorious, and that Turing's main thesis,  $\mathcal{T}$ , is therefore false. It's true that proponents of the orthodox view who assimilate  $\text{TO}^p$  would no doubt feel entitled to reject  $\mathcal{T}$  outright — but that is only because they are persuaded of the soundness of reasoning that would be a key supplement to  $\text{TO}^p$ . For example, in the case of Leibniz and others (e.g., Descartes), the supplementary reasoning would for example include a version of the ontological argument for God's existence, combined with various further arguments (e.g., those given by Pascal) for the trustworthiness of the Judeo-Christian scriptures. And surely additional ancillary reasoning would include arguments dedicated to showing that we are indeed souls (e.g., the one provided by Chisholm 1991). In short, in seeking to move beyond the result that Turing's response fails, we must rest content with a conditional: *if* the “orthodox view” is correct, Turing is wrong.

## References

- Bringsjord, S. (1992), *What Robots Can and Can't Be*, Kluwer, Dordrecht, The Netherlands.
- Bringsjord, S. (1995), Could, how could we tell if, and why should—androids have inner lives?, in K. Ford, C. Glymour & P. Hayes, eds, 'Android Epistemology', MIT Press, Cambridge, MA, pp. 93–122.
- Bringsjord, S. & Arkoudas, K. (2004), 'The modal argument for hypercomputing minds', *Theoretical Computer Science* **317**, 167–190.
- Bringsjord, S. & Ferrucci, D. (2000), *Artificial Intelligence and Literary Creativity: Inside the Mind of Brutus, a Storytelling Machine*, Lawrence Erlbaum, Mahwah, NJ.
- Bringsjord, S. & Noel, R. (2002), Real robots and the missing thought experiment in the chinese room dialectic, in J. Preston & M. Bishop, eds, 'Views into the Chinese Room: New Essays on Searle and Artificial Intelligence', Oxford University Press, Oxford, UK, pp. 144–166.
- Bringsjord, S., Taylor, J., Shilliday, A., Clark, M. & Arkoudas, K. (2008), Slate: An Argument-Centered Intelligent Assistant to Human Reasoners, in F. Grasso, N. Green, R. Kibble & C. Reed, eds, 'Proceedings of the 8th International Workshop on Computational Models of Natural Argument (CMNA 8)', Patras, Greece, pp. 1–10.  
**URL:** [http://kryten.mm.rpi.edu/Bringsjord\\_etal\\_Slate\\_cmna\\_crc\\_061708.pdf](http://kryten.mm.rpi.edu/Bringsjord_etal_Slate_cmna_crc_061708.pdf)
- Chisholm, R. (1966), *Theory of Knowledge*, Prentice-Hall, Englewood Cliffs, NJ.
- Chisholm, R. (1991), On the simplicity of the soul, in J. Tomberlin, ed., 'Philosophical Perspectives 5: Philosophy of Religion', Ridgeview, Atascadero, CA.
- Goble, L., ed. (2001), *The Blackwell Guide to Philosophical Logic*, Blackwell Publishing, Oxford, UK.
- Leibniz, G. W. (1991), *Discourse on Metaphysics • Correspondence with Arnauld • Monadology*, Open Court, LaSalle, IL. This is the twelfth printing. First published in 1902. Translated by George Montgomery; translation modified by Albert Chandler.
- Popper, K. & Eccles, J. (1984), *The Self and Its Brain: An Argument for Interactionism*, Routledge, New York, NY.
- Quinn, P. (1997), Tiny selves: Chisholm on the simplicity of the soul, in L. E. Hahn, ed., 'The Philosophy of Roderick Chisholm, The Library of Living Philosophers Volume XXV', Open Court, Chicago, IL, pp. 55–67.
- Siegelmann, H. (1995), 'Computation beyond the Turing limit', *Science* **268**, 545–548.
- Turing, A. (1950), 'Computing machinery and intelligence', *Mind* **LIX** (59)(236), 433–460.