"The Argument from Jonah", as presented in Selmer Bringsjord's *What Robots Can and Can't Be*, presents a case for the impossibility of "Strong" AI. The argument is based on the idea that all computers really do, and all they could ever do, is manipulate symbols, but with no true understanding of the symbols. The argument centers on a mono savant by the name of Jonah. Jonah does not know any language other than English, but Jonah has the capability to perfectly visualize, in his head, a register machine running a computer program written in some high-level programming language. The premises of the argument can be expressed as follows:

(1_v) If Strong AI will succeed, then there is a program, such that when it is run on a register machine, the machine understands Chinese.

(2_v) If there is a program such that when the program is run on Jonah's register machine, the machine understands Chinese, then Jonah understands Chinese.

(3_v) It is not the case that if Jonah runs the program, Jonah understands Chinese.

From these premises, it is trivial to show that Strong AI can not succeed. However, I believe that the truth of these premises is far from proven.

To start, I would like to present a formalization of the premises given above, to make clear all of the assumptions that are being made:

(1_V) $AI \rightarrow (\exists x \forall y(Rxy \rightarrow CMy))$

(2_V) $(\exists x(Rxj \rightarrow CMj) \rightarrow Cj)$

(3_V) $\neg Cj$

AI = Strong AI will succeed
Rxy = x is run on y's register machine
CMy = the r-machine on y understands Chinese
Cy = y understands Chinese
x = a program
y = the hardware for a register machine
j = Jonah

Premise (1_v) follows from the definition of Strong AI. Premise (3v) can be shown true by a simple sub-argument. Since Jonah can not translate between his native language (English) and Chinese, it is obvious that Jonah does not understand Chinese. I believe that the formal argument given for this is sound, and that there is no need to look at it further. However, the seemingly simple statement in (2_V) contains several hidden premises. Premise (2_V) is actually derived from (1_V) and the following premises:

(2_{Va}) $CMj \rightarrow Cj$ (2_{Vb}) Rxj

The thought experiment requires that (2_{Vb}) be true, so the only premise left unproven is (2_{Va}) – and this is where I think that the problem lies. Nowhere in the argument is (2_{Va}) – the idea that the register machine understanding Chinese is the same as Jonah understanding Chinese – proven. In (2v), it is simply taken for granted that the register machine and Jonah are in fact the same entity. However, there is no justification presented to support this claim. There appear to be two possibilities for (2_V):

(1) (2_V) is true, and the argument is valid

(2) (2_V) is false, because running the r-program creates another person (Cole's objection)

In the book, (1) is never explicitly proven. Instead Bringsjord attempts to disprove (2), so in order to disprove (2_V), all one needs to do is show that (2) is logically possible.

Let us take a closer look at option (2). How can we tell if we have one person, or two? To answer this, let us look back to the definition of a person, as presented earlier in the book:

> "Persons are bearers of psychological, Cartesian, or self-presenting properties … and when a person has such a property she is said to be in a mental state."

When he is not running the program, we know that Jonah has mental states, so when he is running the program, he can clearly still have mental states ("I'm tired of running this program",

etc.). Yet the program (or the second personality), if it understands the Chinese, must also have mental states. If it didn't, it could never express feelings - but without feelings or other mental states, how could it be said to understand Chinese? If it did "understand" Chinese, it would be in a far different sense that what we are concerned with. Therefore, the program, or second personality must have mental states. Yet these could be, and in most cases probably are, different mental states. Since one person clearly can not be in two mental states at the same time, there must be two people.

A possible objection to this might be that, for example, a person could be both happy and sad at the same time. Yet, in this case, they would then be in a mental state that involved both happiness and sadness - not two completely different mental states.

The main objection presented in the book, and in class, to this argument is that it is completely absurd to think that simply by running a register program, a new person pops into existence. While this seems to me to be a rather weak argument to begin with, I do not think that it would necessarily be regarded as absurd by many people. Presented with the question "Could a computer ever be a person?" many people would not think it absurd – suggesting that the phrasing of the question has far more to do with it's absurdity than the contact to the average person. So, let us instead consider someone educated in philosophy. Many such people would not find the idea of creating new people absurd – in fact, this is exactly what Strong AI says! So by labeling (2) absurd, we are really saying that the entirety of Strong AI is absurd. This, however, is begging the question; attempting to prove Strong AI wrong by assuming that what it says is absurd – which certainly doesn't make for a sound argument. In fact, as shown above, we have good reason to believe that (2) is true. This makes (2_{Va}) false, and invalidates the "Argument from Jonah".