R09/Sarthak Mittal/200050129

March 5, 2023

This paper studies the relationship between **intelligent behaviour** and the required **mental qualities**. The viewpoints taken into consideration involve the **"Turing Test"** (by Alan Turing) and the **"Chinese Room"** (by John Searle). Using a different "room" setting, they show that the behaviour is **too complicated** to be produced without mental qualities. The experiment tries to find out **conclusions** that can be drawn about mental qualities (knowing how to speak, understanding speech, or being intelligent) **given** the **right** external behaviour (linguistic responses in a conversation).

The Turing Test claims that if the responses by an **entity** are **indistinguishable** from what it should be from a **person** in the long run, then the entity can be said to have the **same** mental qualities as that of the person. Searle claims via his Chinese Room that he could be using a **large book** that tells him what to do with a particular Chinese message, and he gives **congenial responses**, which he does not understand, thus showing that correct behaviour does not imply mental qualities. Although, they both agree that the "conversation" needs to be **unrestricted** and **natural**.

Restricted linguistic behaviour has been simulated using uninteresting means, but this does not reflect on the unrestricted Turing Test. For the Chinese Room, the discussions refuted it saying that the system of Searle along with the book knew Chinese, and Searle was essentially **executing a program** without understanding it. Although, if Searle **memorizes** the book's contents, he can claim to generate the correct responses without knowing Chinese. This paper shows that "memorizing" makes some **assumptions** that do not seem justified.

The paper argues about two **types** of books, one that makes the reader learn Chinese (type 2) and the other that does not (type 1). Searle claims he uses a type 1 book. The authors argue that an extensively large "English-Chinese-English" manual is an example of a type 2 book, but there **cannot exist** a type 1 book for Chinese. The past arguments about Chinese Room were not convincing because no one knew what such a book would be like, which taught Chinese as a **first** language. To prove this claim, they use another more straightforward experiment.

The **Summation Room** has a person (who does not know addition) with a book. Given a list of twenty 10-digit numbers, the person needs to respond with a 12-digit number (the sum). Consider **Book A**, which has 10 billion

chapters, each chapter has 10 billion sections, and so on till depth twenty. The instructions are as follows: (1) go to chapter number, which is the first number in the list (2) go to section number, which is the second number in the list ... (21) the number on the page reached is the answer. This book shows that the person can **look up** numbers and produce the right answers without knowing addition if the book is structured accordingly. Although, Book A **cannot exist** due to space requirements. It would need $\mathbf{10^{200}}$ **entries**, and the physical universe has about 10^{100} atoms. The maximum possible is 10^{20} entries, which would allow adding **two** 10-digit numbers. The argument is that even if the number of conversations is bounded, a "sufficiently large" lookup table might be **impossibly large**.

They justify the existence of the "Summation Room" by showing that Book B, a type 2 book, does exist for addition. By breaking the process down into **4 procedures**, which carry out addition using lookup for adding two single-digit numbers, and dividing the input into smaller numbers to handle them using simpler procedures, they claim that the person who memorizes the process **actually learns** to add. To further argue that type 1 books do not exist, they allow **large memory aids** and store sums of all pairs of 10-digit numbers. Book C then uses these to get the sum of twenty 10-digit numbers. **Book C is also type 2** because addition can be performed with any **base**. Book B uses base-10, while Book C uses base-10^{10}.

The key here is that Book A cannot be adapted to **any** set of numbers, while Books B and C can be. Within a **restricted** setting, Book A seems to work because it can achieve the correct behaviour within the allowed **page limit**. Generally, Books B and C might require a few pages more to **adapt**. The authors claim that if a **procedure** produces the correct answers for addition without using lookup (due to memory constraints), it works for any list of numbers and only operates on the numbers. The procedure is **"addition"**.

In conclusion, they use their proof of the **non-existence** of a type 1 book for "Summation Room" to argue that an unrestricted Chinese conversation would surely be **more complex** than adding twenty 10-digit numbers, and thus not possible to achieve via any trick, simulation or mere lookup. Essentially, they claim that the proper behaviour of **responding in Chinese** cannot be **easier to fake** than **adding twenty 10-digit numbers**, and just exploiting the absence of a clear definition of a type 1 book for Chinese does not justify Searle's Chinese Room. Turing's belief that "simple-minded tricks" **cannot scale up** to "human intelligence" is further strengthened.