

# A Minimal Turing Test: Can You Prove That You're a Human?

Imagine you and a smart robot are competing for survival, but the judge can't see either you or the robot. The judge will guess which of you is the human. Whoever the judge thinks is the human will live, and the other will die. The judge says: "Each of you must give me one word from an English dictionary, and based on the word, I will guess who is the human?" What one word do you choose?

I encourage you to answer this "*Minimal Turing Test*" by John P. McCoy and Tomer D. Ullman before reading on.

Now, my turn is to provide you with a strategy to crack this competition. You want to choose a word that differentiates you from the robot. However, a word that reflects a too obvious difference to you will also be obvious to a clever robot, so it wouldn't be a good choice.

I bet you now have in mind that robots probably lack one of the most famous aspects of humans: emotions. To zoom out a bit, "emotions" belongs to a broad category (brought up by Kurt Gray and D. Weigner) that distinguishes humans from robots: **Experience**, which relates to "*feelings and experiences, such as pain, hunger, joy, sorrow, and jealousy*" (I doubt that there exists any robot having those things). On the other hand, according to Gray and Weigner, there's a category that robots might have succeeded in it: **Agency**, which relates to "*thinking, doing, including self-control, memory, planning, and thought.*" Just try to calculate  $1543^{234}$  without touching any computers!

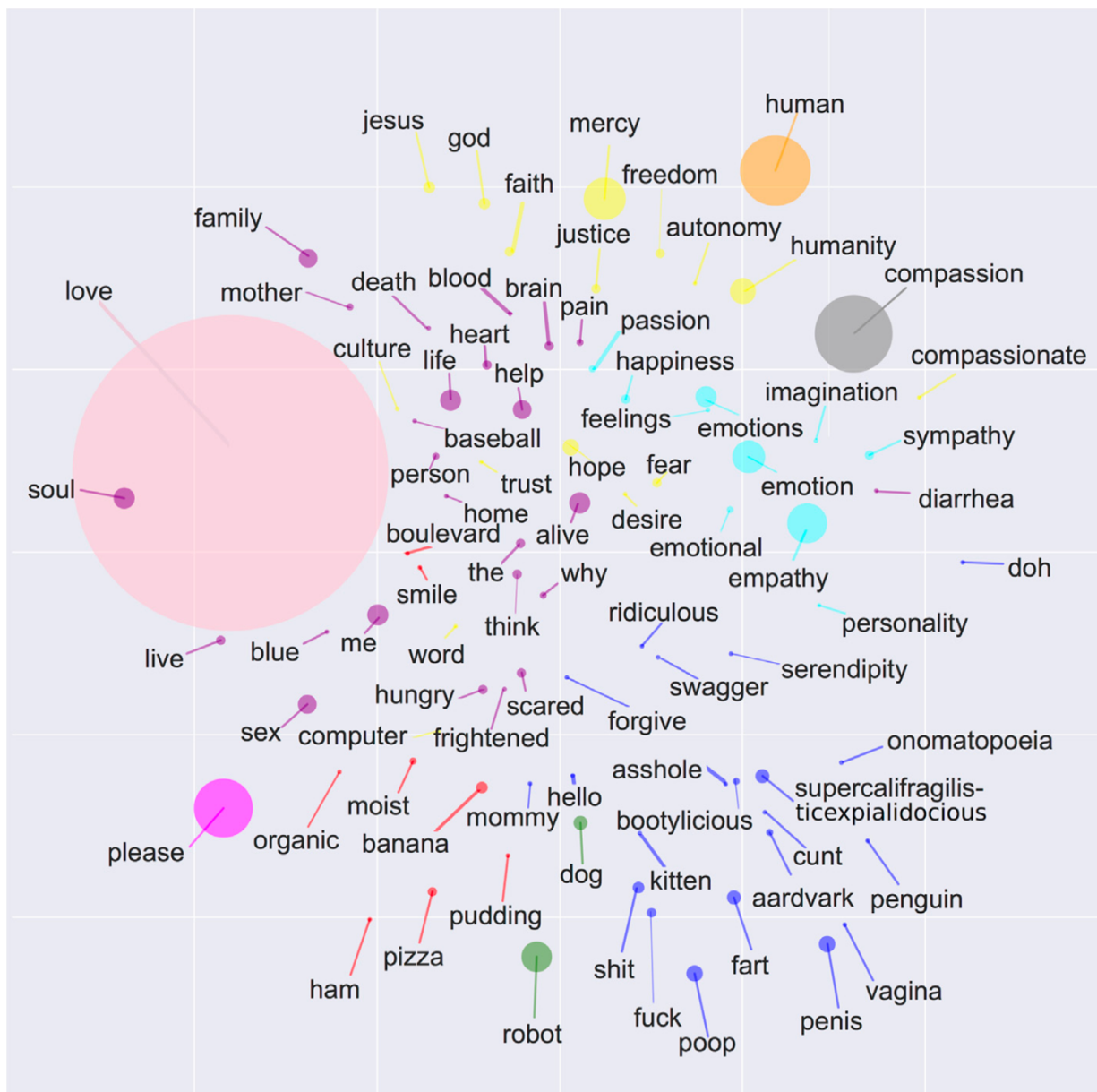
So, picking a word from **Experience** is the key to crack this competition, isn't it?

That might sound like a good strategy. But don't you wonder which single word will be the most optimal choice for us as humans? (Remember: it's the survival competition we're talking about!)

To answer that question, I will first show you what the choices of other people are, which words people choose the most, and after that, which words have the greatest strength to convince our judge.

The above considerations were also in the minds of John P. McCoy and Tomer D. Ullman, and they called the "survival competition" of giving one word to prove that you're human: "A Minimal Turing Test". To find out the solution, they received responses from 936 volunteers. The 936 participants gave 428 words. If you wonder why 936 givers only had 428 results, this was because 90 words were given by more than one participant. To be more generalized, words given by more than one volunteer accounted for 64% of the responses.

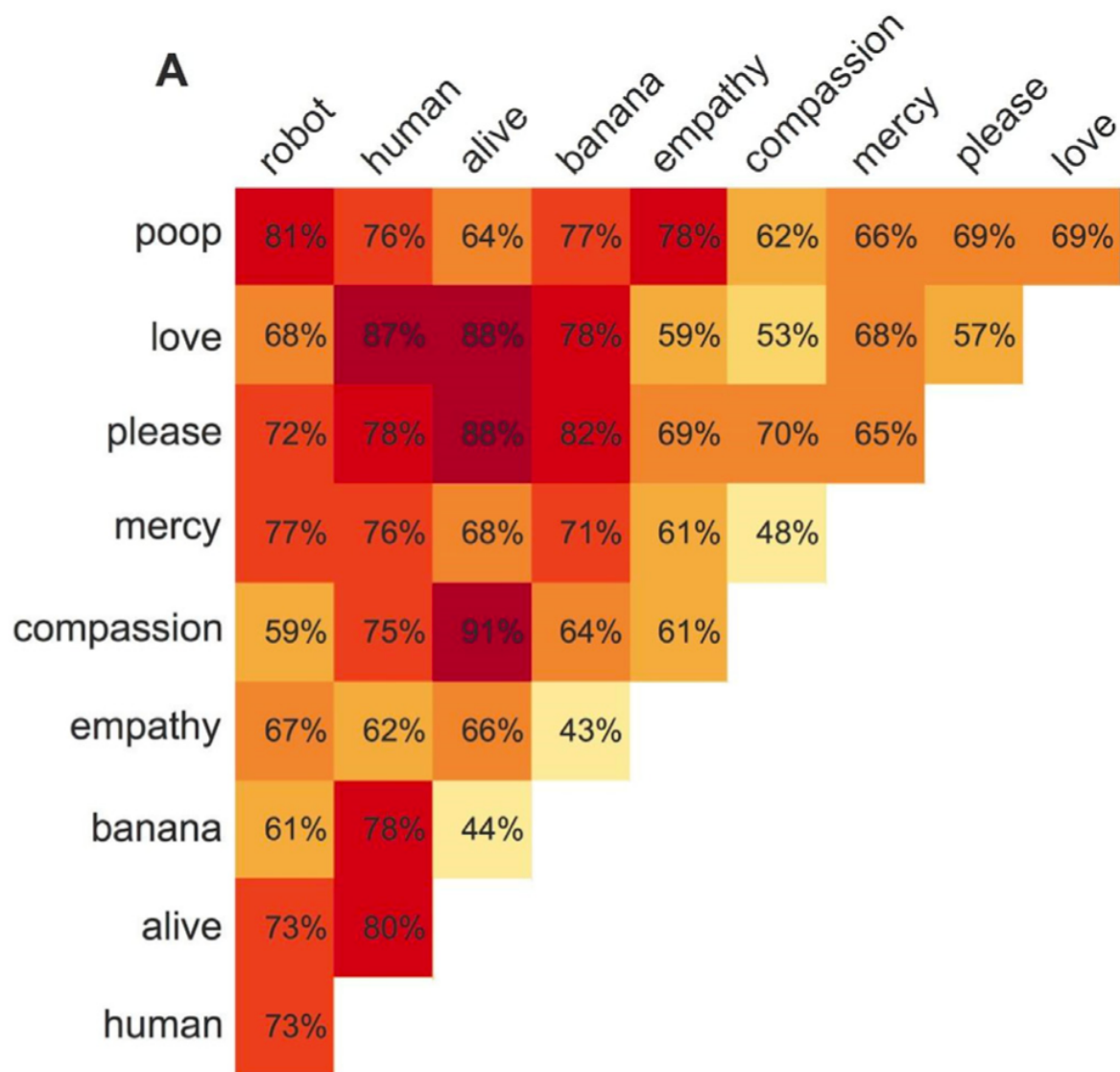
To understand more clearly the characteristics of the words, John and Tomer divided them into 10 clusters (concepts). The figure below by John and Tomer describes the assignment of the words into the clusters. Words with the same color belong to the same cluster. The bigger the circle is, the more times the words were given.



Each of the four most frequent words forms a “single-word cluster”: **Love** (given by 134 volunteers), **Compassion** (by 33), **Human** (by 30), and **Please** (by 25). Those 4 “single-word clusters” account for 24% of the responses. Six remaining clusters can be roughly characterized as relating to **Affect** (‘empathy’, ‘emotion’, ‘feelings’, ‘sympathy’, ‘happiness’), **Faith and Forgiveness** (‘mercy’, ‘hope’, ‘Jesus’, ‘God’, ‘Faith’), **Food** (‘banana’), **Non-human agents** (‘dog’, ‘monkey’), **Life and Death**, and **Bodily Functions and Profanities** (‘poop’).

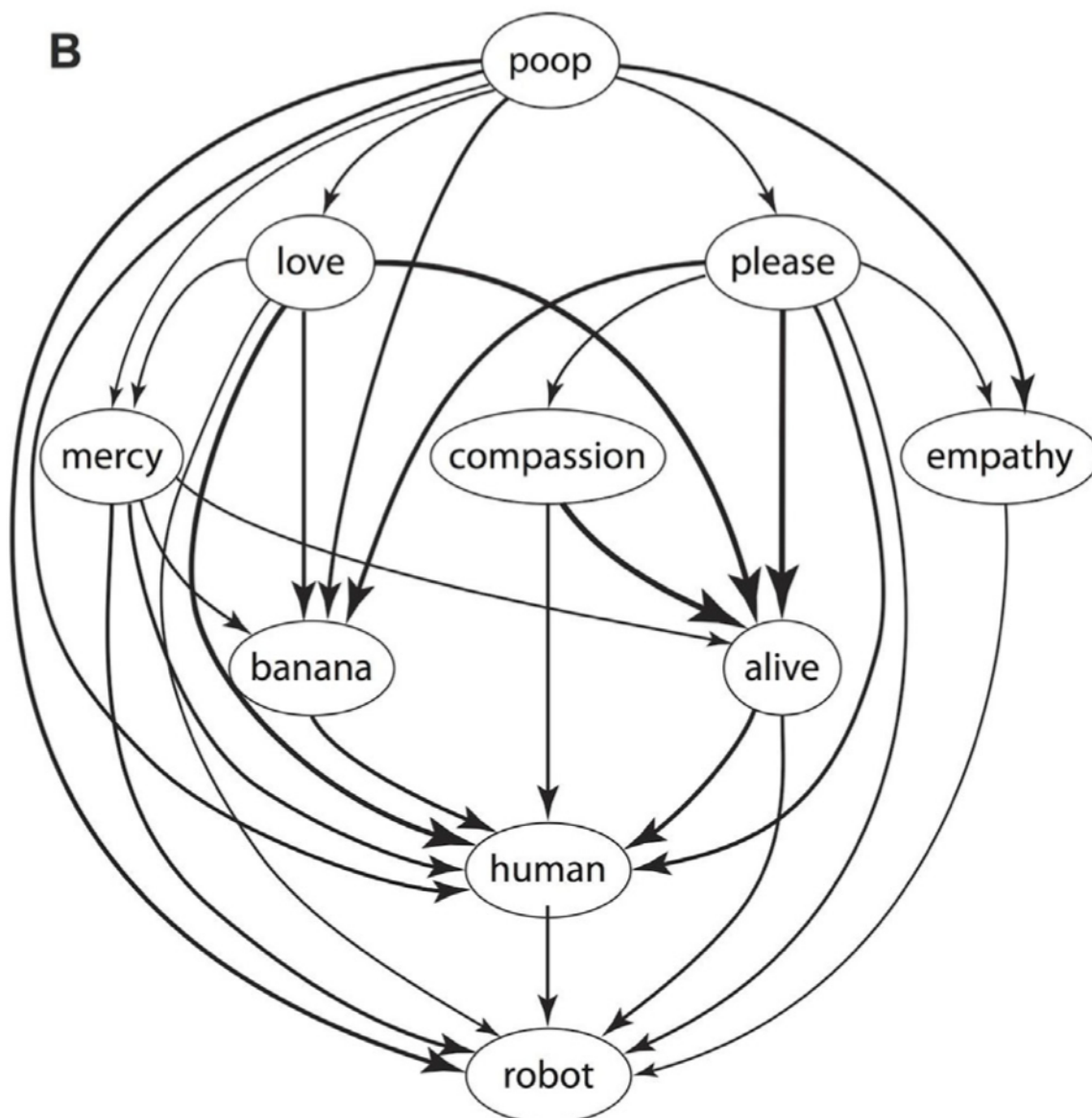
Remember 2 categories: **Agency** and **Experience** I told you earlier? 7% of the participants gave a word relating to “Agency”. 40% of the participants gave a word relating to “Experience”. The rest didn’t belong to any of the 2 categories.

For the most interesting question: “Which word has the greatest strength of conviction?”, John P. McCoy and Tomer D. Ullman recruited 2084 “judges”. Each of them was given a pair of words (for example: “love” and “human”) and decided which one of the two was “stronger”. The “A graph” below describes the results of judges’ decisions.



The percentage represents the portion of “judges” who decided the word in the column was more convincing than the word in the row. For example, concerning the pair: “poop” and “robot”, 81% of the “judges” agreed that “poop” was stronger than “robot”.

Having the data concerning the strength of the words, John and Tomer also drew out the “hierarchy” of the strongest words. (In the B graph below).



The higher the word is in the hierarchy, the stronger the word is. For example, “love” is stronger than “mercy”, “mercy” is stronger than “banana”, and “banana” is stronger than “human”.

It’s intuitive that “robot” is the weakest one, and obviously from the graph, “poop” takes the winner title here.

According to John McCoy and Tomer Ullman, the “Minimal Turing Test” can be generalized to understand attributes that people believe distinguish different social groups of people. To prove that you’re a member of a particular group (in the test, humans), you would rely on properties, or characteristics, that reflect on shared cultural and social background unavailable to non-members (in the test, robots), while also considering how easily the judges can evaluate the properties.

To sum up, if you want to prove that you’re a human, just “poop”.

**REFERENCE:**

John P. McCoy & Tomer D. Ullman (2018). A Minimal Turing Test. *Journal of Experimental Social Psychology*. 78(1), 1-8. <https://doi.org/10.1016/j.jesp.2018.05.007>