

Turing Test Two

Copyright © 2007 Huoyan Gao

Turing Test Two

Huoyan Gao

huoyangao@yahoo.com

Dec. 7, 2007

Abstract

TTT, an imitation game and an extension of Turing Test, is proposed. TTT tries to address the question “which is more intelligent” instead of the question “what is intelligence”. TTT is a versatile format that can turn a set of questions into an interesting game on almost any subjects, by assessing the ability of players to guess other player’s thinking. In simplest terms, TTT is Turing Test for human. Both existing and newly conceived games are discussed within the context of TTT. The relationship between TTT and the original Turing Test and IQ test is discussed.

A little over 50 years ago, Turing proposed an imitation game, now known as Turing Test ¹, as a test of whether a computer has intelligence or not. In this game, two players, a man (A) and a computer (B), are being questioned by a human interrogator (C). If C cannot tell which of the two is human after reasonable length of time, the computer is considered passed the Turing Test and therefore “intelligent”. Turing Test has had and continue to have great impact on artificial intelligence, the philosophy of mind, the cognitive science and even pop culture ².

In this paper, a simple, straightforward extension of the Turing Test, which we will call Turing Test Two, is proposed. Turing Test Two attempts to determine which of the two players is more “intelligent”. For simplicity, we will abbreviate Turing Test Two as TTT (unrelated to Total Turing Test ³).

In Turing Test Two, two players A and B are again being questioned by a human interrogator C. Before A gave out his answer (labeled as aa) to a question, he would also be required to guess how the other player B will answer the same question and this guess is labeled as ab. Similarly B will give her answer (labeled as bb) and her guess of A’s answer, ba. The answers aa and ba will be grouped together as group a and similarly bb and ab will be grouped together as group b. The interrogator will be given first the answers as two separate groups and with only the group label (a and b) and without the individual labels (aa, ab, ba and bb). If C cannot tell correctly which of the aa and ba is from player A and which is from player B, B will get a score of one. If C cannot tell which of the bb and ab is from player B and which is from player A, A will get a score of one. All answers (with the individual labels) are then made available to all parties (A, B and C) and then the game continues. At the end of the game, the player who scored more is considered had won the game and is more “intelligent”.

As just formulated, TTT is not weaker than the original Turing Test and it is most likely to be much more difficult than Turing Test. Since at this stage, no computer is even close to passing the Turing Test, there might seem to be of little value to propose something that is even harder. One argument can be made is that Turing Test can only be taken Pass/Fail and since all computers fail now it does not offer immediate guidance towards the goal of eventual passing the Test⁴. TTT does give scores, which can be used for exactly such purposes.

However, since TTT gives scores, it is not only possible but could be interesting to have TTT being played by two human beings instead of by one human and one computer⁵. Much of the rest of the discussion is limited to this context of two human players. For some technical details of a TTT session, see Appendix A.

Below, we will first discuss some interesting games in TTT format before returning to some more theoretical issues. These concrete examples, hopefully, will serve as background material for the later discussion.

The “Hydrogen Atom” of TTT

Although Turing Test is proposed 50 years ago, only recently does serious contests are organized⁶. On the other hand, one could argue that TTT activities are already happening without even being noticed by anyone.

Let us start with the simplest things first. Although nobody ever put it in this way, one of the simplest children games, RPS (Rock, Paper, Scissors) can be thought of as one of the simplest and purest TTT activities ever, or we might say the “hydrogen atom” of TTT. In this game, players simultaneously change their hands into one of three gestures, Rock, Paper or Scissors. Rock win over scissors, scissors win over paper and paper win over rock. If two players choose the same gesture, the result is a tie and played again. By the rule of game, if one player puts out a gesture of rock, then it implicitly implies that he think the other player will put out scissors. Similar arguments hold for other gestures. So the RPS game is so simple that one answer serves the purpose of two (i.e. aa = Rock implies ab = Scissors and so on, in TTT terminology. If indeed bb = Scissors when aa = Rock, A wins in RPS game. But then ab = bb, so A score a point in TTT sense, as it should be). Mathematically, the result of RPS game is equivalent to that of coin flipping and it is often played just for exactly that purpose, and even preschool children intuitively understand that. However, since RPS gives players chance to guess opponent’s mind, it gives players the feeling (rightly or wrongly) that they have a certain level of control of their fate and only if they can outsmart their opponents... So there are national and international tournaments for this simple game⁷.

If something as boring as coin flipping can become interesting and fun when reformatted as TTT game, one can only imagine what would happen when other inherently interesting and fun things are reformulated as TTT activities. Traditional games like Chess and Go can be easily put into a TTT format by scoring how well each

player guess the opponents' next move. This may not be as easy as it seems⁸. More importantly TTT suggests new games to be invented. We discuss some prototypes below.

TTT Word game

This is game more for elementary school pupils. A and B take turns to question each other. A ask a word question. B responds with a word. A will have to guess this word. If A guessed correctly A will score a point, otherwise B will score a point. In case of a dispute about whether a particular word can be considered a reasonable answer to a question, a pre-agreed thesaurus or a third party C functions as a judge or referee.

A word question is simply a question that can and should be answered with a single word. For example, “give me a synonym for word w”, “give me an antonym for word w”, “give me a synonym for word w except w_1, w_2, \dots ”, “name an animal start with l and end with d?”, “what is the weather like today?”, “what do you think of my haircut?”, etc.

Supplemental Rule 1. If B cannot come up a word that answers the question, neither A nor B get any points.

The reason that B will not get points is quite obvious. A will not get points since we want to encourage a game of positive instead of negative. Players are awarded for knowing something instead of the opponent not knowing something. Since A has to guess B's answer, there is no danger of the word question to be too board so that many words fit. It is in A's interest such that the question to be very specific so as few words fit as possible. However we need a rule so that players cannot just ask questions that have unique answers to get the “lower hanging fruits”, e.g., “what day it today?”. Therefore we have:

Supplemental Rule 2. Reversal. At any point of the game, if a player that is supposed to respond feels that the question is such that only one answer is possible, he could request a role reversal that will always be granted by the rule of the game and is final for the particular question. So instead of have to utter the obvious answer to let the opponent to get a point, he can ask her to do so to get a credit himself. Also, if A correctly guesses B's answer he will need to give another plausible answer to get the score.

In general, TTT word game encourages questions that have multiple answers so players have to make judgment instead of simply knowing the answer.

Supplemental Rule 3. In advanced games both multiple answers and multiple guesses are allowed. For example, B can give multiple answers to try to get multiple points. However, if A can guess several of these words correctly A will get multiple points, too. B will get points for the answers missed by A. If B gave one answer B_w and A suggests three answers and one matched B_w , than A will get a point of $1/3$, etc.

There is no need to go TTT for the sake of TTT. In more traditional word games, one simply needs to know more words and their meanings than other players to score more points. However, if the goal of learning words is to communicate with others rather than to show off (as is the case in real word supposedly), it is better to make points using words that others do know, using the meanings that others know and you know they know... TTT word game is exactly for that.

“Democracy 101”

This is a game for a small group of people. The group considers “bills” on national, local and/or family level issues that are interesting to the group. For every bill the group will elect a member as a “representative”, presumably to represent group on the particular issue at some larger assembly or congress at some future date. The game plays like this: Everybody takes turn to act as a sponsor to introduce a bill to be considered by the group. Each player will decide his/her position on the bill (or more precisely whether he/she will vote for or against the bill at the future congress if he/she get elected) without knowing other players’ positions. Each player will also give vote to one group member as the representative. One cannot vote for oneself and the bill sponsor. Players will vote without knowing each other’s positions or votes on the particular bill but with full knowledge of all previous positions and votes of everybody. The group member got the majority votes will serve as the representative. If need be, dice can be used to break a tie. Obviously, the member you voted for may or may not take position on the bill the way you wanted. The group representative (which maybe the one you voted for, and maybe not) may not take the position the way you wanted. What’s more, the elected representative may not take the position the majority of the group member would wanted though he got a majority of votes. Sad? maybe, at lease sometimes. Realistic? you bet. After the voting, a debate will follow so players know each other’s reason for or against the bill. And optionally the bill can be amended in various ways and reconsidered by the group.

This game can be played at several levels. A beginner would focus on picking members that share his “political” views so he needs to guess other players’ position on the bill, judging from previous records etc. Once he is good at that, he would work hard so the elected member will share his position. So the electibility issue comes in here. Not only does he have to guess other players’ position, he also needs to guess their votes. He needs to make compromise between members that he is more confident that will share his position but has less chance to be elected and member that is more likely to be elected but the chance of sharing his political position is less. Advanced players would, of course, want to campaign to get elected. Therefore, one often has to vote out of expedience. On the other hand, one has to try to establish a track record, too. When one is get used to be elected, one can play for the hardest game, trying to get elected while being unpopular.

So a plausible scoring scheme for the game can be like this, you get a point if the member you voted for took the same position on the bill as you do; you get a point if he

or she got elected; you get another point if he/she got elected and share your position. You get 10 points or so if you get elected yourself. You get another 10 points if you get elected and you took an unpopular position, i.e. your position on the bill is different from the majority of the group.

The game will become erratic when everybody wants to become representative and starts to guess and vote the popular way. There are two ways to limit such deteriorations. When a bill is voted for (or against) by an overwhelmingly majority of the group, the bill can be throw away or amended to be more controversial. Politics is really about controversy, when there is no controversy there is no good politics. Another easy way is to give some extra bonus points for players took minority position.

“I like...”

Felix and Ashlyn are both passionate classical music lovers and they are becoming friends. As music lovers, when they play games they want a game that is about music too. So they come up with this what they call “I like...” game. Real music lovers feel it is awkward and even meaningless to talk about the music in words. Certainly, one can name the composer, the conductor, the pianist, the recorder but that is far from enough, the main thing is the music and it makes little sense to talk about music. In this game, each will pick a piece of music and they will first simply enjoy the music. As they listen the two pieces of music together each generally will have an idea which of the two he/she likes better. They will also try to guess each other’s favorite. Whoever guessed right got a point. As they accumulate points, they know music better and they know each other better and they know they know each other better...

The Turing Game

There is a little problem if people start to take a TTT game seriously and become determined to win at any cost. For any given question, a player may really think answer X is right but instead he will say the he thinks that answer Y is right just to win the game. To the extent that a player consistently adopt positions that are of someone else (i.e. role play), that is not a big problem. Exclude role play, in TTT game just as in real life, there is always the possibility that what one say one thinks might be different from what one really thinks. As long as the frequency of such differences is not significantly different from what is happens in real life, it should be considered as a feature, not a bug⁹. However, unlike in real life where there is some sort of moral standard (however loose that may be), a TTT game could happen in truly a moral vacuum and there is the real danger that a player will just give out random answers to make it harder for the other player to guess his position. There are various ways to discourage such strategies.

- 1) Cooperative game¹⁰ where it is in a player’s interest to have the other player guesses his thinking right in stead of the other way, presumably competitive

- against another team. Probably the simplest example is a three party (A, B and C) game. For each question asked, if B guessed A's answer correctly *both* B and A will get a point while if C guessed A's answer correctly *only* C will get a point. Permutation works out for other players. Or it can be randomly decided at each step whether it is B or C is to cooperate with A, etc, to avoid any "conspiracies" among players, so to speak.
- 2) Take advantage of the structure in space S. For example, in the "I like..." game, though it makes no sense to dictate which music is better than another, it can be required that there is a rough order between all the choices, so random answers will be penalized.
 - 3) "You can fool some of the people all of the time, and all of the people some of the time, but you can not fool all of the people all of the time". The "Democracy 101" game is a good example for this. If someone just keep adopting random positions, the group will soon realize that and he will be seen as true "politician" and have little chance to be reelected again and again (certainly I hope this is the case).
 - 4) Lastly, but most importantly, in case where winning the game becomes more important than the game itself¹¹, like what had happened in many sports, neutral third party can be introduced as the point of reference. So instead of A and B guessing each other's answers, question will be directed to a third party D. A and B will both be requested to guess D's answers.

However, even in cases where the subject matter is such that no structure can be easily explored there is a still a way out of the quagmire and this is a general game, which I will again name as Turing Game, take advantage of the second order Turing space introduced in Appendix B. Depends on personal taste, this could be the best of the games or this could be the worst of the games. It could bring the best out of a person. It could bring the worst out of a person. In this game, for each question that is asked, A will give his answer a_0 as before. B will try to guess a_0 and give his guess as b_1 . A is also tasked to guess b_1 and give this guess as a_2 . A certainly can prevent B from guessing his answer (a_0) by give random answers as a_0 . In such cases, to guess a_0 the best B can do is to guess randomly so b_1 will be random too and then the best A can do when guessing b_1 is random guess. Giving out all random answers, however, are going to give the player a score that will approach the lowest bound of all possible scores since it does not take any intelligence to give random answers. If one really wants to win the game, one has to give up this strategy even facing the risk of losing the game.

Turing Intelligence and subject knowledge

TTT tests, first and foremost, the players' ability to guess the opponent's answers, and in overwhelmingly majority of cases such guess has to be a judgment call. In honor of Turing, I will call the ability to predict other's thinking Turing intelligence. Turing intelligence is a high order cognitive skill. There are certainly logical, analytical

components in Turing intelligence as well as subconscious, “gut feeling” components which may well be more important. Clever Hans, a horse that initially was believed to be able to solve arithmetic problems, disappointed many when it was found that instead the horse was actually guessing the correct answer from watching the cues from human observers¹². Although horse does not have the intelligence to understand simple mathematics, the Clever Hans’s performance was actually is a primitive form of Turing intelligence. If we are open minded about this we should recognize that what Hans performed is actually more impressive than what initially believed. What Hans was believed to be able to do, billions of computers can do now and do it much better while it will be decades if not centuries before a computer can do what Hans actually performed. Without making any value judgment, it makes sense to hypothesize that Turing intelligence is an evolutionarily earlier trait than mathematics skill. If a horse can accomplish the seemingly unthinkable, admittedly after much training, can human being be far behind?

A player’s performance in TTT depends upon Turing intelligence but also depends upon mastering of the subject knowledge. If one knows nothings about French, then it does not matter how intelligent he is, he will not be able to guess the other player’s answer to questions about French. On the other hand, perfect subject knowledge itself does not automatically translate into Turing intelligence. Let’s use a math test as a concrete example. A is very good at math so that he can answer all the math questions correctly. B can only answer 75 percent of questions correctly. After a short period of warm up, A can infer that B is not as good as himself and can only be correct only at 75% of the time. However, this information alone does not allow A to predict B’s answers correctly. Since B has 75% chance to be right, for each question, A has to guess that B will give the correct answer so overall A can only be right 75% of the time in predicting B’s answers. B will predict A correctly 75% too so there is no difference in terms of Turing score. Only when B is only 49% correct, A can do better by predicting that B will give wrong answers all the time. In this case A will be 51% right in Turing score. To be really good at TTT game, A has to see more. For example, if A can tell that B’s weak point is geometry so he can start to predict B will give wrong answers for future geometry questions while give right answers for algebra (and other) questions, then A will be able to do better than B, in terms Turing scores. In practice, it will not be surprising if an absent-minded Ivy League math professor underperforms a really good family tutor in predicting a high school student’s answer to math questions.

The Turing score one gets in TTT reflects both Turing intelligence and the subject knowledge. In case the subject knowledge can be measured, it is relatively easy to decompose the Turing score. In cases where the subject knowledge cannot be directly measured, it might require scores against multiple players on multiple subjects and hypothesis about the independences between some of these variables to do such decomposition.

TTT and Turing Test

TTT is more closely related to Turing Test than it might seem to be. A popular view is that Turing Test tests a computer's ability to understand and speak natural language, i.e. verbal behavior, which is the hallmark of intelligence¹³. This author's view is that Turing Test does test verbal behavior and all the arguments for verbal behavior as hallmark of intelligence are valid. However it is not the original intension of Turing. As the real pioneer of artificial intelligence, Turing underestimated the difficult of natural language processing and therefore did not foresee that it will become a bottleneck for computer to pass Turing Test¹⁴. On the other hand, having natural language understanding and synthesis ability itself is not enough for passing Turing Test. Just as being able to understand what Fermat's Last Theorem is not, and actually far from, being able to prove it. Language ability is necessary but not sufficient for intelligence.

Instead, again it is this author's personal view that the core of Turing Test is "imitation", i.e. the ability of computer to *imitate* human's cognitive skills. To be able to imitate a human being, the computer has to be able to guess what the human being will think. So the similarity of TTT to Turing Test is not skin deep, but rather goes straight to the bone. On the other hand, the difference between TTT and Turing Test is not trivial either. To use terminology introduced in the Appendix B, Turing Test are essentially testing similarities of points in the subject space S ($a_0 =? b_0$)¹⁵ while TTT is testing in the Turing space ($a_1 =? b_0^*$) or higher order Turing Space ($a_2 =? b_1^*$). In Turing Test, the computer is trying to imitate the human player in a general sense. If the computer can imitate *any* one human being successfully, it will pass. In TTT, one is tasked to understand a *specific* individual, and effectively imitate the other player.

In Turing Test, the function of the human player is rather passive, more like a control in a typical scientific experiment. The computer needs to work hard to imitate human being while the human player needs just "to be oneself". It is clearly the roles of the two players are asymmetric. In TTT, on the other hand, the roles of A and B are symmetric.

TTT and artificial intelligence

Whether passing Turing Test should be viewed as the ultimate goal of artificial intelligence (AI) or not, the relationship of Turing Test and artificial intelligence is very clear and very natural. TTT, however, mainly tests Turing intelligence and readers from AI background might feel that TTT has very little to do with AI. Yes, Turing intelligence is real, not artificial. On the other hand, to the extend that AI want to build computers that having human intelligence, Turing intelligence could and should be part of it. Probably of more immediate applications are software systems (expert system, machine learning based maybe) that show reasonable level of Turing intelligence in some well defined sub-domain of human knowledge and/or interests. There system could be deployed to play against human player in TTT games, for example. With a rather large database and good learning algorithms, it is certainly possible, in at least some sub-domains, for software system to perform good enough to have practical applications.

TTT and IQ Test

Turing Test is often being viewed as an IQ (Intelligence quotient) test for computers ¹⁶. This is not a paper about IQ and the author is certainly aware of the controversies related to IQ test ¹⁷. TTT is not IQ test. TTT is not and is not meant to be a competitor of IQ test. As TTT is supposed to measure which of the two players are more “intelligent”, there is certainly the danger (or opportunity, depending upon one’s point of view. This is just another case where there is no GAP answer) of using TTT to test IQ and/or to rank order people. It’s worth to note, though TTT does not depend upon GAP answers for scoring, the familiarity with the subject matter still affects scores and could still have biases against certain sub groups of the population ¹⁸.

On the other hand, as it seems every cognitive skill is correlated with IQ ¹⁹, it will not be surprising that Turing intelligence is also correlated with IQ. To use psychometric terminology, Turing intelligence as measured by TTT is more an achievement than an aptitude attribute, much like one’s mathematical skill. There might be an innate ceiling of such skill (math or TTT) any particular individual can achieve. If there is indeed such a ceiling for Turing intelligence, TTT is not meant to measure it.

Final Words

One could say that nothing said here is really new. The importance of the ability to tell somebody else’s thinking, especially that of one’s opponents or enemies (Turing intelligence, as we would call it) has long been recognized. More than two thousand years ago, Sun Tzu, an ancient Chinese military strategist said “if you *know others* and know yourself, you will not be imperiled in a hundred battles” ²⁰ (italic added). Nor does it become obsolete these days. I quote from a recent business textbook: more successful managers “tend to perceive co-workers, as well as opponents, more accurately and are effective in *discerning those person’s intents* and strategies” ²¹ (italic added). Advice from chess grandmaster: “know your opponent” ²². Indeed, figuring out what other people’s thinking is such an integral ingredient of human activity that one often does so without noticing it. TTT is just pointing out the fact that one can do so explicitly, on almost anything that one can think of, and more importantly, getting immediate feedbacks while doing so (in a game context) which is often what is lacking in real life.

It is often said that this is the information age. However as more and more information become just simply “googleable”, the competitive advantage one once enjoyed by simply knowing some “information” is coming down rapidly. What’s next? Judgment. Sound judgment. When you have to make a call one way or the other, while you don’t have all the information, while you never will have all the information, while not only you don’t know the answer but nobody knows the answer or only one person knows the answer but will not let you know until you have made the call (this is what TTT is about). Knowledge != Judgment. If a question has a well defined (GAP) answer, then it does not matter how profound or subtle it is, one can in principle just “know” the answer. However, if it does not have such answer, then it does not matter how trivial the

question is, one has to make a judgment call. (Of course, more significant are cases where for a question there may be a time when nobody knows the answer and a time when the answer was first found. Or even more significantly when the question was first conceived. But these belong to discovery and innovation, which has been for the privileged few). One can make bad judgment while have the world's best information at his/her fingertip. On the other hand, sound decisions were made with incomplete information and/or with disinformation. Turing intelligence is probably one of the most important ingredients of many such judgments. This is, or at least is rapidly becoming, an age of Turing intelligence²³. A bold predication can be made: Turing intelligence will be the most important, universal element of the productivity that will drive the world economy in a post-information age.

This paper is written very informally so it is accessible to as wide an audience as possible. In appendix B. a few things were explained more rigorously. One might be attempted to dismiss the distinction between the subject-matter space S and Turing space T; or believe that it is really the S space that matters, T is just notational convenience; or S is real while T is just a mirror image of S and higher order Turing space just like seeing things reflected through multiple mirrors. It is true that it is in the S space that $1+1=2$, the sky is blue and Thomas Jefferson wrote the declaration of independence, etc., etc. Objective truth lives in S space, no argument there. However, at risk of repeating what may well have been repeated million times, "beauty is in the eyes of beholder" therefore beauty lives in the Turing space²⁴. As well as faith, feeling and almost everything else that is subjective, personal, controversial. However, have to live in Turing space does not make them less important, instead the fact that they live in Turing space makes the Turing space important.

It is also not so clear that S is more fundamental than T. A movie star's birthday falls on a particular day is a simple fact that belongs to S. Knowing such trivia certainly can make someone pride of oneself. However, many many computers can "know" (or "remember") such facts much better than human being. It is the ability to tell whether a close friend knows these details from the reading of his/her mind that truly makes human mind the miracle of the universe.

Popular games like Chess and Go have being played for centuries. Although there is no question that top chess (or Go) players are smart, the conventional wisdom seems to be that they become world champions because they are smart, not that they become smart as they become world champions. It often takes someone years to decades of deep immersion in such game before becoming world top players. However, it does not seem the cognitive skills developed from playing such games give one huge advantage in other (more "real world") professions, probably because the important skills for winning in such games are too special, too specific to the particular game²⁵. TTT offers the opportunity to develop games that are more real world like and one can hope the skills developed there will be more beneficial in real life²⁶.

Mathematics, particular geometry is sometime considered as "mental gymnastics", for exercising one's logical, quantitative, analytical skills. Then, perhaps, TTT can be

viewed as “mental ping-pong”, exercising brain circuitries that certainly overlap the math circuitries to some extent but otherwise quite distinct. The importance of mathematics could not be over emphasized and it will certainly continue to increase. Yes, mathematical world is beautiful, pure, abstract, structured, precise, rigorous, rational, absolute, universal, eternal ²⁷. But the real world often can be otherwise and TTT give us a chance to play in such a world and wandering in TTT world will be equally fun and rewarding as in the math world.

TTT is a format where personal opinions can be studied more seriously than just poll taking. It is essentially a different way of looking into the human mind, through a “mirror” (“lens” is probably a more accurate description). Obviously nothing one can see the TTT way that one cannot otherwise. However, looking the world the TTT way gives abundant information about the lens, which is another human mind.

Turing had predicted that in 50 years computer would be close to pass the Turing Test. He made the predication at 1950. It is fair to say that his prediction failed. With hindsight, it is probably safe to predict now that computer will not pass unrestricted Turing Test in another 50 years. Therefore, there is the admittedly minority view that Turing Test should now be moved from test book to history book ²⁸, probably more out of frustration than conviction. However, as TTT illustrates, Turing’s simple but deep insight as crystallized in Turing Test is still inspiring us fifty years later. We can only expect it will continue to do, if not for another 500 years, certainly for at least another 50 years.

Appendix A. TTT in details

For many questions there are what can be called generally accepted or preferred (GAP) answers, which most reasonable people will agree to be the “correct” answer to the question. For example, for question “what is 1 plus 1?”, the generally accepted answer is “2”. For question “who is the first president of united states?”, the answer “George Washington” can be considered as generally accepted, etc.

Most tests (SAT, GRE, IQ) consists of this kind of questions. TTT, on the other hand, allows questions to be used as testing materials on almost any subject - art, music, politics, religion, myth, sports, gossip, rumor, trivia, cooking, wine, clothing (for women?), cars (for men?), shopping ideas, etc. Most interesting are questions that do not have answers that are agreed either by the mass or by the experts. For example, questions like “should we support stem cell research?”, “Is there life after death?”, i.e. questions that reasonable people will and generally and usually do disagree (remember “ten Jews, eleven opinions?”). Another category of interesting questions for TTT concerns questions that *do* have a GAP answer but knowing such answers itself will have very little correlation with general intelligence. For example, “Which day is Jerry Seinfeld’s birthday?”²⁹. I would assume even Jerry Seinfeld will agree that one should not expect, on average, much difference in intelligence between the people who know the answer and who does not. However, if you are told one of your close friends knows the answer, it should not be that hard for you to guess who he/she is.

Let’s digress here and take a look of a concrete, admittedly hypothetical but none the less completely realistic example. The example is more complicated than what usually would happen in a TTT game, just to help to make the points clear. Beauty is in the eye of beholder. Given two pieces of arts, X and Y, try to come to a GAP answer to the question “which one is more beautiful?” is at best controversial, at worst downright insulting to some³⁰. However, let’s imagine a controlled setting. A and B both have some idea (they know it will be about art) but not exactly what the fateful question is. They can select 12 individuals by cross-examine a long list of potential candidates just like what would happen in a US court system. After the cross-examination, A and B were asked to predict how each of the twelve will answer the original question, “which of X and Y is more beautiful?”. At the same time, the question is put to the twelve. Once each of the twelve individual gave their own choices, these answers are in a sense become micro “historical facts”³¹. These answers are, of course, no where as significant as but none the less just as solid as, if not more solid than, the fact that “George Washington was the first president of united states”. There and then, after these opinions are collected, which of A or B can better guess these twelve answers is something that can be measured accurately and objectively. Of course, *if* a different set of twelve individuals are selected the answers certainly will be different. *If* the same twelve individuals were questioned again at a later time, some of them might well change their opinions. Or *if* the order of X and Y is switched in the original question some of the answers might be different. But all these (and anything else) are irrelevant, as the game is over. On the other hand, it is exactly

these “ifs” that make the TTT activity interesting, challenging, fun, worthwhile and, above all, realistic or one might attempted to say, simply just real³².

For questions that have no GAP answers, building games in *non-TTT format* is hard if not impossible. For questions that do have GAP answers, it is straightforward. However, such game tends to encourage rote learning, which is less desirable. Simply knowing the answer to a question is not real intelligence.

The TTT format we just introduced depends upon the human interrogator C to make judgments at every step of the game. However, for many test subjects, it is possible to change the format to that of multiple choices. In such format, at every step C will give a question and a list of possible answers, A and B will be requested to make their choices in the list provided, for both their own answer and their guess of the other’s. Then scoring is simplified to checking whether the answer labels match each other or not. Multiple choice format could cause controversy in tests where a single “correct” or “best” answer has be to provided³³. However, TTT only needs to match answer labels to get a Turing score, it does not need such special answers for scoring and therefore should be less controversial. In informal, friendly games, the questions could even come from players themselves.

In Turing Test, to avoid any bias against the computer, Turing had to propose to have A and B sit in separate rooms from the interrogator C and have them only communicate through a teletype. In TTT, such setting is certainly OK, too. However, it is not clear that it is necessary or desirable. It is possible to have A and B to sit in the same room and talk face to face just like, for example, in a typical job interview. Or in a group setting, like a cocktail party. Facial expression, body language and even what cloth someone wear can give subtle but none the less important clue to someone’s thinking and TTT is meant to be as realistic as possible.

It takes two to play a TTT game. What if no other player is available at the time one wants to play? Of course, many such games can be played over the internet. Another option is have many TTT sessions recorded and replayed later for other players to play. For example, a game between A and B is recorded and later on another player D can take the position of A to play against B. Since B’s answers are directed against A not D so it may not be perfect but still it is valuable.

Appendix B. Turing Space

For a given question Q , which can be on any subject s , all possible answers form a space, which we will call space S (of the subject matter) for question Q . For simplicity, we will limit the discussion to multiple choice type of answers, so the space S is really just a set that can be labeled with integer $1, 2, \dots, N$. Therefore S contains s_1, s_2, \dots, s_N . A will give his answer to the question which we will denote as a_0 and B will give her answer to the same question which will denote as b_0 . At the same time, A will give his best guess of B 's answer to the question (i.e. guess of b_0) which we will denote as a_1 . B will also give her best guess of A 's answer (i.e. guess of a_0) which we will denote as b_1 .

While a_0 and b_0 can be any of s_1, s_2, \dots, s_N , what about of a_1 and b_1 ? For example a_1 can be of the following: "B will pick answer s_1 ", "B will pick answer s_2 ", ..., "B will pick answer s_N ". These answers form another space, which we will name as Turing space T for the question Q . There is a nature one-to-one mapping from space S to space T :

$$\begin{aligned} s_1 &\leftrightarrow t_1 \text{ (A/B will pick answer } s_1\text{)} \\ s_2 &\leftrightarrow t_2 \text{ (A/B will pick answer } s_2\text{)} \\ &\dots \\ s_N &\leftrightarrow t_N \text{ (A/B will pick answer } s_N\text{)}, \end{aligned}$$

we will use $*$ to denote this mapping, so $t_i = s_i^*$ ³⁴.

In cases where there is no danger of confusion, we make no distinction between s_i and t_i and simply say that $a_1 = b_0$ when A correctly guessed B 's answer (e.g. $b_0 = s_i$, $a_1 = t_i$), or more accurately $a_1 = b_0^*$.

Higher order Turing space. We don't have to stop at here. We can ask A to guess not only B 's answer to the original question Q (b_0) but also B 's guess of A 's answer to the original question ($b_1 = \text{guess of } a_0$). This "second" guess, which will be denoted as a_2 , could be like "B will answer 'A will pick answer s_1 '", etc. Obviously, all possible a_2 form another space, Turing space T^2 . Similar to the mapping $*$ just introduced, there is a nature one-to-one mapping from Space T to T^2 , which we will also denote by $*$. Under such mapping, we can say $a_2 = b_1$ (when A guessed B 's guess correctly, i.e. $a_2 = b_1^*$), etc. Even higher order space can easily be defined but will not be used in this paper.

The distinction between space S and T might seem to be trivial. However there is something fundamental here. Perhaps, one simple example will make the subtle but important difference clear. In Space S , it is always that " $1+1 = 2$ " is true and that " $1+1 = 3$ " is false. However, in space T it is *possible* that *sometimes* "A thinks $1+1 = 3$ " is true while "A thinks $1+1 = 2$ " is actually false. That is for questions that has a GAP answer. For questions that do not have a GAP answer, we cannot assign a true/false value in space S . For example, to the question "do you support pro-life or pro-choice?", we cannot say pro-life is true and pro-choice is false or vice versa. However, there is no problem to assign a true/false value to "A think B is pro-life" once we know the answers A and B just given. Therefore, although one can not always scoring answers in S space, one can

always do so in T space. In TTT where possible we get a Subject score from answers like a_0 . We always get a Turing score from answers like a_1 .

Generally, a TTT session will not be just one question but rather consists of a series of more or less related questions. The product space $S_n = S_{Q_1} \times S_{Q_2} \times \dots \times S_{Q_n}$ (and $T_n = T_{Q_1} \times T_{Q_2} \times \dots \times T_{Q_n}$) is of particular interest. Different individual's answers $a_{0_{Q_1}}, a_{0_{Q_2}}, \dots, a_{0_{Q_n}}$ is simply a point in the space S_n . The distribution of such points from a population of human being is not random and Turing intelligence is basically the ability to see and use patterns in such space.

Notes

1. See Turing, A. M. (1950).
2. See Saygin, A. P., Cicekli, I. and Akman V. (2000).
3. See Harnad, S. (1989).
4. See Hayes, P and Ford, K. (1995).
5. See Hofstadter, D. R. (1981), for an interesting example where human imitated a computer.
6. See Zdenek, S. (2001).
7. See http://en.wikipedia.org/wiki/Rock%2C_Paper%2C_Scissors .
8. See P44. Kotov, A. (1971) for an intriguing case where several chess grandmasters failed to guess the next move in a Soviet Championship match.
9. In the most rigorous sense, what someone really thinks is only knowable to himself/herself and even that maybe questionable, at least in some situations; see Trivers R. (1990). Although it is important to be able to guess both what others really think and what they say they think, we can only score a game according what one says what he thinks but not what he really thinks, until we have a scientific instrument that can measure accurately what people really think. On the other hand, just as we cannot see quarks directly, or atoms at that, does not prevent us from building quantum mechanics, the most successful theory in physics, the fact that we cannot know directly what someone really thinks should not prevent us from getting a “theory of the mind” of the person and make predictions about, say, what he will say what he thinks even when that is different from what he really thinks. In fact, we are doing that all the time and sometimes quite successfully at that. The basic assumption of this paper is that what someone thinks is something real but not currently directly measurable and which may or may not agree with what he says he thinks. It is also worth to note that it is not all that simple that what one really thinks is real and deep and what one says one thinks is superficial and therefore un-important. Certainly, one would assume, a lady planning for the wedding would be more interested in whether her fiancé really loves her, than whether he says he loves her. On the other hand, a political party is probably much more interested in whether you vote their way than whether you really believe what it proclaims in its manifesto.
10. Heap, S.P. Hargreaves and Varoufakis, Y. (2004).
11. One has to wonder why people would still be interested in playing the game in such cases, but that’s besides the points.

12. See Pfungst, O. and Wozniak, R. H. (2000).
13. See Shieber, S. M. (2004).
14. This should not reflect badly on either Turing or Turing Test. Pioneers, by definition, see things far far away and as a side effect future looks too close to them too.
15. Turing did not specify in great detail how the human interrogator should try to discriminate between human and computer. Since currently computer are far away from passing the Turing Test, there is not a great deal of research done on this. From what little that has been reported, it can be speculated that the human interrogator might be looking for either things a human being is expected to know while the computer is not; or the other way around where the computer may just “knows too much”. See Shieber, S. M. (1994) for an example where a human player was misidentified as computer for knowing too much about Shakespeare. Basically, one can imagine that interrogator is putting some sort of “humanness” probability on each statement. When the combined probability becomes too small, the computer’s identity is exposed.
16. See Hofstadter, D. R. (1981).
17. This author’s personal opinion of IQ test is that one should just view it as any of the many sport games. As long as everyone knows the rule of the game, willingly joins the game and plays according to the rule of the game, then one should just accept the outcome of the game. Of course, we all know that the best teams don’t always win.
18. Though it is possible to measure such bias and compensate for them.
19. See Carroll, J. B. (1993).
20. See Sun Tzu (2003). Of course, “know others” in “the art of war” means knowing everything about the enemy but this includes knowing the enemy’s thinking and it is arguably the most important part.
21. See p220. Streufert, S and Swezey, R. (1986).
22. See p191. Kotov, A. (1971).
23. One can at least hope to put a significant fraction of space S on the internet. There are bits and pieces of Turing spaces coming to the internet but they are on a different curve and there are 6.6 billions of them to be worked on. Most importantly it is not clear that people really want and if they do how much they want to be put out there.
24. There is no question that space S can contain statements like “X is beautiful”. However generally one cannot assign a true/false value to such statements in space S. Without a true/false assignment, “X is beautiful” is logically no different from “X is

ugly” in space S. It is in Turing space that statements like “A thinks X is beautiful” get true/false value and become live, logically. Note, although one cannot assign true/false value to every statement in space S, it does not mean that they are all equally likely, that’s why in Turing Test, after lengthy interaction, it is possible for the interrogator to tell which is which. Most likely one can still assign some probability to at least some of the statements in space S, even though these statements cannot be simply labeled as true or false. However, it is much simpler, more elegant and satisfying to deal with an issue logically than probabilistically, where possible, of course.

25. See De Groot, A. (1966) and Chase, W. G. and Simon, H. A. (1971). World chess champion certainly has influence beyond chess. However this is mostly celebrity effect.

26. Nothing to against Chess or Go whatsoever. For some, Chess and Go is life.

27. We are only talking about learning mathematics that is ready known. Creating new mathematics could be something altogether different.

28. See Hayes, P. and Ford, K. (1995).

29. In case you don’t know, see http://en.wikipedia.org/wiki/Jerry_Seinfeld .

30. See p352. Karier, C. J. (1976) for a test item “which is prettier?” from an IQ test and the related controversy.

31. For the truly paranoid, A and B can be separated from the twelve after the cross-examination and before the question is put to both parties. It should also be clear that, unlike what happens in a court case, there is no need for the twelve to reach any consensus or whatsoever, instead it is their individual opinions that are valued here.

32. See Neisser, U. (1976), Klemp, G. O. and McClelland, D. C. (1986) and Lave, J, Murtaugh, M and de la Rocha, O. (1976). Many IQ tests and aptitude tests are thought to test things that are simply too “academic”. One of the reasons is probably that real world problems often do not have GAP answers and therefore are not easily given in standard format.

33. See Owen, D. (1999).

34. Given a space S, all elements in S can be mapped to space T using the map $*$ just introduced, form a subspace S^* of T. It could happen that $T = S^*$. But space T can be extended so it is larger than S^* . For example, for a question that does have a GAP answer, it is reasonable for a player to answer, for example, “he/she will NOT answer s1”, to express the prediction that the other player will give a wrong answer, assuming the player believes the correct answer is s1. This is, of course, when he can not be sure which wrong answer the other player might pick. $T = (S^*, \neg S^*)$.

Reference

- Carroll, J. B. (1993) *Human cognitive abilities: a survey of factor-analytic studies*. New York: Cambridge University Press.
- Chase, W. G. and Simon, H. A. (1971) *The mind's eye in chess*. In Chase, W. G. (Eds). *Visual information processing*. New York: Academic Press.
- De Groot, A. (1966) *Perception and memory versus thought: some old ideas and recent findings*. In Kleinmuntz B. (Eds). *Problem solving: research, method, and theory*. New York: John Wiley & Sons.
- Harnad, S. (1989) *Minds, machines and Searle*. *Journal of experimental and theoretical artificial intelligence*. 1(1), 5-25.
- Hayes, P. and Ford, K. (1995) *Turing test considered harmful*. *Proceedings of the fourteen international joint conference on artificial intelligence (IJCAI95-1)*. Montreal, Quebec, Canada. 972-997.
- Heap, S.P. Hargreaves and Varoufakis, Y. (2004) *Game theory: a critical introduction*. London:Routledge.
- Hofstadter, D. R. (1981) *The Turing test: a coffeehouse conversation*. In Hofstadter, D. R. and Dennett, D. C. (Eds). *The mind's I*. New York: Basic Books.
- Karier, C. J. (1976) *Testing for order and control in the corporate liberal state*. In Block, N. J. and Dworkin, G. (Eds). *The IQ controversy critical readings*. New York: Random House.
- Klemp, G. O. and McClelland, D. C. (1986) *What characterizes intelligent functioning among senior managers?* In Sternberg, R. and Wagner, R. (Eds). *Practical Intelligence: nature and origins of competence in the everyday world*. Cambridge: Cambridge University Press.
- Kotov, A. (1971) *Think like a grandmaster*. Dallas: Chess Digest Inc.
- Lave, J., Murtaugh, M. and de la Rocha, O. (1976) *The dialectic of arithmetic in grocery shopping*. In Lave, J. and Rogoff, B. (Eds). *Everyday cognition: its development in social context*. Cambridge: Harvard University Press.
- Neisser, U. (1976) *General, academic, and artificial intelligence*. In Resnick (Eds). *The nature of Intelligence*. Hillsdale,NJ: Lawrence Erlbaum Associates
- Owen, D. (1999) *None of the above*. Lanham: Rowman & Littlefield.

Pfungst, O. and Wozniak, R. H. (2000) *Clever Hans (The Horse of Mr. von Osten)*. London and New York: Thoemmes Continuum.

Saygin, A. P., Cicekli, I. and Akman, V. (2000) Turing test: 50 years later. *Minds and machines* 10(4), 463-518.

Shieber, S. M. (1994) Lessons from a restricted Turing test. *Communications of the ACM* 37(6), 70-78.

Shieber, S. M. (2004) *The Turing test: verbal behavior as the hallmark of intelligence*. Cambridge: MIT Press.

Streufer, S. and Swezey, R. (1986) *Complexity, managers and organizations*. Orlando: Academic Press.

Sun Tzu (2003) *The art of war*. Boston: Shambhala.

Trivers, R. (1990) The elements of a scientific theory of self-deception. *Annals of the New York Academy of Sciences*. 907, 114-131.

Turing, A. M. (1950) Computing machinery and intelligence. *Mind* 59(236), 433-460.

Zdenek, S. (2001) Passing Loebner's Turing test: a case of conflicting discourse functions. *Minds and machines* 11(1), 53-76.