Scientific method and game theory as basis of knowledge and language

Krystian Zawistowski
krystian.zawistowski@zoho.com

April 21, 2022

Contents

1 Introduction

We use methods of science (parts of falsificationism) as a foundation of knowledge and language. We draw some parallels to human sensory experience, using recent progress in AI and demonstrate how do we know basic facts about space or ourselves or other people. Then we demonstrate how we can understand and make language with these methods giving examples from Tok Pisin language. Of basic hypotheses of understanding that emerge, one is related to focal point theory, a game-theoretical concept. Then we demonstrate the viability of this approach for clarification of philosophy. We demonstrate that our theory is a good answer to many linguistic conundrums given in "Philosophical Investigations" by Wittgenstein. We also demonstrate an application to other philosophical problems. It appears that scientific thinking is very important to our lives and also that it can shed much light on the riddles of philosophy. In appendix, we add a few remarks on the connection between falsificationism and induction.

2 Sensory data, thoughts and AI

2.1 Senses

One could think of our sensory experience as something like a movie (a series of pictures, sounds and other sensory data) in our head - we indeed have some sensory apparatus that talks to our brain through electric signals. One could also access sensory experiences of the past, which we often call a part of memory. The accuracy of both of those apparatuses is limited - and memory is way more limited than senses. We perceive present and past - sensory 'now' is way more vivid and intense and affects our senses, compared to memories.

All these sensory apparatus could be adequately replicated with the use of electronics - cameras, microphones, electronic memory, neural nets detecting objects etc - thus it is mostly understood by us as a possible result of some mathematical computation happening in our neural system

We retrieve and process sensory data in a way that is understood. But we think about this data and make decisions in a way that is not particularly clear. I would call the latter part 'abstraction' or 'hypothesis generation. This is overlapping but not identical to some concepts in philosophy: e.g. Kant's perception and conception division.

2.2 Example - Chollett corpus for abstraction and reasoning.

Francois Chollett, a computer scientist, proposed a method and dataset to measure the intelligence of AI systems [1].

He prepared a dataset for testing a general artificial intelligence system. Its openly available part consists of 400 challenges, simple games dealing with coloured squares on a grid, often no greater than 30x30 squares. Challenge has the structure of a few examples, showing what question and solution should look like and one test case to solve.

An algorithm to solve one given challenge could be like 'there's a red dot and green dot on the grid. draw blue path going from red dot (coming from it vertically) to green dot (coming to it horizontally)'. Or 'there are coloured blocks in line with some squares coloured, some grey. align squares together so that grey squares stick to each other.

If a human looks at these riddles, a solution comes to his mind sooner or later. It is not clear how to solve it programmatically in the general case. You can figure out, perhaps that most challenges involve squares and lines and points, and they can be represented as such, but solving most of them requires additional consideration.

Let us denote geometric objects as the dot (pixel), line with their coordinates and colour in parentheses, colours red, green, blue denoted as r, g, b. Dot $\text{dot}(r, x, y)$ is one red pixel at x, y. Line $\text{line}(b, x, y_1, x, y_2)$ is blue, horizontal line from (x, y_1) to (x, y_2).

Let us consider one riddle. Input and expected output to it

Figure 1: Example input and output for one of riddles

is given on Figure 2.2.

After seeing it and two similar examples one can summarize it as follows: 'There's a red dot and green dot on the grid. draw blue path going from red dot (coming from it vertically) to green dot (coming to it horizontally)'. Here's a solution - a symbolic algorithm to solve this riddle (assuming we give input and output on the left and right-hand sides respectively).

$$\begin{aligned}(\mathrm{dot}(g, x_1, y_1), \mathrm{dot}(r, x_2, y_2)) \rightarrow (&\mathrm{dot}(g, x_1, y_1), \mathrm{dot}(r, x_2, y_2),\\ &\mathrm{line}(b, x_1 + \mathrm{sgn}(x_2 - x_1), y_1, x_2, y_1),\\ &\mathrm{line}(b, x_2, y_1 - \mathrm{sgn}(y_2 - y_1), x_2, y_2 + \mathrm{sgn}(y_2 - y_1)))\end{aligned}$$

This is example of a challenge and a solution from Chollet's dataset.

The problem is to come up with a computer program that does it on an unknown set of problems similar to 400 problems that were made public.

There was a challenge being held for the best solution. The winner scored 20% accuracy on the test data set, writing (7k lines long) symbolic solver that worked adequately on known part of problems At the time of writing unofficial state of the art is 30%. At the moment thus it is not possible to replicate human performance on this with a computer program. Given a hypothesis for a solution it is not hard to test it programmatically against examples, but the generation of adequate hypotheses in the general case is the hard part.

2.3 Example - Object detection and CAPTCHA

The above example suggests that a distinction must be made, between the detection of objects and abstraction. Detection of objects in Chollet's corpus case would detect, dots, rectangles, lines, circles etc - a rather simple task programmatically. In a real-world case, it can detect cars, rabbits, or trees in visual input or words in the audio input (represented as spectra). As of 2020, these are largely solved (on par with human performance) problems, with the best solutions being neural nets. Neural nets are mathematical equations with millions of parameters. Inputs to them could be e.g. pixel representations of images or sound spectrograms and outputs are answers like 'which word or item is found in the input', given as index or vector with a discrete probability distribution. Parameters are optimized on real data.

Let us draw a line (border) between mechanical object detection of state-of-the-art neural networks and human object detection. One useful tool of it is CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart), a technique often used to prevent programs (crawlers etc) from access to services that are meant to be used only by humans (such as creating accounts, login). One example of such a system is reCAPTCHA by Google. It typically involves solving challenges such as “select pictures with traffic lights”. Sometimes ’weird’ traffic lights are given - ones that are horizontal or bigger than usual. Also, photos are often deliberately of low quality. A traffic light could just look like a black box with a red dot, nonetheless, humans know what it is. There’s road, cars and earthwork - and the box is just in the right place to be a traffic light. This is similar to Chollet-type riddles and also something artificial neural object detectors are unable to cope with.

3 Method of Science

3.1 Hypothetico-deductive method and falsificationism

Hypothetico deductive method, an algorithm of scientific research, is often attributed to Popper, an influential philosopher of science, but its origins can be traced to some earlier thinkers as well.

Here's a simple description.

- We come up with a hypothesis.
- We infer with an unlikely or relevant prediction that follows if the hypothesis is true and won't happen if the hypothesis is false.
- We compare prediction with empirical sensory evidence.

'Relevant' or 'unlikely' here means that prediction should contain knowledge we otherwise don't have - such as prediction of numerical value we don't know (e.g. the location or the time of occurrence of something)[1] We would say that hypothesis is 'corroborated' if it has confirmed unlikely predictions. That is 'corroborated' is to be confirmed by experience and there are degrees to this confirmation depending on abundance, relevance and variability of unlikely predictions. Abundance and variability are connected, as after one unlikely prediction is done,

[1] For detailed treatment in application to science one could see concept of logical probability in 'Logic of Scientific Discovery' paragraph 34).

another similar prediction is not unlikely, because we already expect this happens. To corroborate is to test by making unlikely predictions successfully. If the prediction is not verified, a refutation happens, but one could start the process over with a slightly different hypothesis.

An important distinction (to see if the process is scientific) is between progressive problem shifts and degenerative problem shifts. If our changes to hypothesis produce lots of new relevant predictions and correctly predicted results - it matters not that wrong predictions and refutations sometimes happen. We are producing new knowledge on average and our final hypothesis is correct and corroborated by new prediction. If we are proposing explanations for negative results without prediction of new results that's a degenerative or unscientific problem shift. This is a crucial part of 'sophisticated' falsificationism described by Lakatos[6].

4 From science to more general knowing

Philosophers are in disagreement over issues such as "do objects exist", "are we in simulation", "are there other minds", "what is the language", "what is the structure of our language". We will show that some underlying assumptions of our actions in the world are extremely well corroborated empirical hypotheses, with the use of H-D method. (At least, not at most).

4.1 Discovery of our existence in the world

How to know that we see our hand in front of us? We look at pictures of hands or impressions. They sometimes move, according to some pattern in our senses. We can move our hands or our legs, seeing their pictures being moved. This implies being able to predict very well, where they could be for example in the next few seconds. My hand will grab a cup if I choose so, or it will wave before my nose if I choose so, or it will remain at rest.

This is an example of relevant prediction. Given certain circumstances, something specific should happen ("my hand will move" is not as specific as "my hand will grab cup now"). We control circumstances, so we can call it "experiment". "What happens when my 'self' send certain signals - something changes, there's a pattern to it."[2] Similarly, movement of our eyes causes

[2] It is not clear how we "will", but the good thing is we can't fail to will (see Wittgenstein P.I. - 612), so it matters not.

movement of our sensory view - we can point our eyes in different directions, see our view being transposed over a larger picture (picture slides to the left if we move our eyes to the right) If we look to the left, inspect what we see in detail, and then look to the right, we can predict that if we look again to the left we would see the same thing we saw already. This way we corroborate that we are looking at parts of some bigger field of view.

A reflective surface such as a mirror or water or even a monitor displaying camera input - allows us to know more about ourselves. When we look at ourselves in it (not knowing what it is yet) we can quickly learn a few things. We stretch our hand - if we see back of it - we see the palm in the mirror, if we see the palm, we see the hand back in the mirror. Any movements we make are seen in the mirror, any scene behind us is seen in it too, thus we can predictively corroborate the hypothesis that this thing allows us to see anything in front of it. We can confirm that our eyes and face are about in the place where we would expect to see apparatus related to our sensory vision and that our whites move precisely with our sight. To 'confirm' here means 'first observe, then abstract, then corroborate by prediction'.

We can 'move' (altering our vision in a certain way) with the use of our legs and the rest of the body. Then we can corroborate that we navigate through space structured like 3D Euclidean geometry. For example, seeing the shapes of a cuboid room we can predict what we will see when we move through it. Seeing a cup we can watch it from different angles, then lift it, put it somewhere else and see it stay there - corroborating hypothesis that there are solid objects (as we understand it).

As we will see later, Kant argued that “we must have knowledge of space independent from experience, we couldn't aquire it from experience”. I think not, it can be established by experience extremely well - much better than Faraday law.

4.2 Bounded hypotheses

Some of these hypotheses will involve 'boundaries' sooner or later. It is not always true that we will see same thing if we look again in the given direction after a time. Some things move or change and a description of this will require further hypothetico deductive study. A similar approach applies in science - in XIXc people were resolving deviations from the Newtonian model of Solar System. For example, Le Verrier used perturbations of Uranus orbit, to predict that there's another planet - Neptune. Anomalous movements of Mercury weren't fixed that way - they were explained by Einstein's General Relativity, a competitive theory. Thus a counterexample was made against Newtonian gravity and dynamics. It doesn't mean that Newton's gravity was abandoned - on the contrary - it is used much more often than General Relativity.

Is then Newtonian theory true? Not as it was understood in the XIX century, perhaps. But we know that these Newtonian equations are extremely accurate as long as we deal with, for example, objects on Earth's surface with a mass small compared to the mass of the Earth. One could say that Newton's theory was falsified, but bounded Newton's theory reemerged (and was corroborated).

It is somewhat different in the philosophy of science, as the focus is made on the development of science

4.3 Background assumptions

A handful of problems emerges. Hypotheses that were established in the past are being used at a certain point in the future. One hypothesis may depend on another, or corroboration of one hypothesis may depend on the validity of many other hypotheses. What if the latter is wrong?

It turns out complex hypotheses depending on other hypotheses can in practice be tested as long as they have predictions and as long as we can test the assumptions too, though their ability of refutation suffers. This is related to Duhem-Quine problem in the philosophy of science - but I would say that this problem is not a serious obstacle.

4.3.1 Science example 1 - OPERA neutrinos

We will look at some scientific examples. It is hard to find something more convoluted these days than big particle physics experiments with a hundred or a thousand people working on it and the discovery of new physics relying on the validity of a large body of old physics. One of such experiments, OPERA produced in 2011 a highly anomalous result of faster than light neutrinos. If a controllable beam of superluminous particles was real it would be a theoretical problem. On paper it would allow us to send information backwards in time, for instance, thus we would prefer to cap the speed of neutrinos at speed of light.

Second, it was not confirmed by other measurements of neutrino speed.

What could physicists do in that case? Apply H-D method of course. OPERA people went on testing every bit of electronics and found some fault (failure in time synchronization electronics + loose optical link). Fixing machinery, electronics or software or cracking code is often done with hypothetico-deductive approach as well. When we see an issue in a program, then we can test isolated parts, predict that they should satisfy tests and narrow down on anomalies. After we fix the issue we expect that the whole system will work in a use case that was not working, which is another prediction.

If out of a sudden, some elementary physical law needed to make this work stop working, the procedure would be the same. Let us consider a thought experiment: consider that we do mass spectrometry of some bit of matter with accelerator producing ions, magnet deflecting them, electromagnetic measurement apparatus and certain set of predictions of the measurement - and also background assumption such that there's reality being modelled as 3D Euclidean space, there's Lorentz law governing forces acting on particles in the magnetic field and slow-enough particles move according to Newton dynamics. Out of a sudden a tiny, constant, space-time deflection forms up and makes our particles go in a different direction - and we get failed prediction, not knowing why it happen. It looks like a problem, but if we keep reiterating our H-D method we can sooner or later find out the issue. For instance, if we try doing the experiment at a different place it would work as expected. So apparatus is broken, we could hypothesise. Then we would keep testing our apparatus - e.g. electromagnet - by putting a cathode ray tube in the magnetic field. We would find a place where deflection is not quite right. Moreover - anomaly would persist when we turn off the electromagnet and all particle trajectories would be subject to it, regardless of the charge. Photons would be affected too, thus possibly we would be able to see it directly - a weird 'lens' floating air. And if Euclidean geometry suffered a global breakdown it will be seen in other experiments. Sooner or later we would find the issue, regardless of what it is, as long as we can keep iterating.

A side note: OPERA neutrino anomaly of 2011 was occasionally attacked as some kind of huge fault. I think it can be used to demonstrate that researchers (at least in experimental physics) most of the time act rationally, with the use of hypothetico-deductive method and getting apparently suboptimal results doesn't imply irrationality. Summing up:

1. If after refuted prediction or a surprise, researchers further apply H-D method, to narrow down what is the cause - this is often overlooked in methodological analyses especially done by non-physicists as no one writes this in textbooks or papers.

2. This investigation is often limited by time and budget and also accuracy and capability of equipment, at no fault of scientists or fault of the method.

4.3.2 Science example 2 - Tycho's Stellar Parallax

Thus it sooner or later happens that we can't immediately investigate further - for instance in the case of rare phenomena, that can't be controlled. Or when we don't have any means for further investigation. One of the examples of H-D method sending scientists in wrongish direction is Tycho Brahe's attempt to see solar parallax, a result he inferred from Copernicus's theory.

What is solar parallax? If we look at a distant celestial object from Earth and Earth goes around the Sun, then we should see this object in a slightly different direction on the opposing points of Earth's orbit (in June and in December, for instance). It is indeed true that we do, but the difference is small for distant objects (we know that parallax of nearest stars doesn't exceed 1 arcsecond) and Tycho Brahe couldn't see it. It is instructive to look at the inference Tycho Brahe did with this result. First, seeing that stars are about the same size, he conjectured that stars are about the size of a Sun (we know that it is mostly accurate - the Sun is a medium-sized main-sequence star). He then compared the size of a star to the size of the Sun and deemed that a) either star are like Sun or somewhat larger and should be close enough to see parallax b) stars are extremely big and extremely far compared to Sun. Considering that the latter was rather problematic and considering the obvious sensory experience of the stillness of Earth (the relationship between atmosphere and gravity wasn't yet understood), he concluded that Earth is still. What's the catch here? We are unable to see the real size of stars. What we see as their size is in fact the size of the diffraction picture. If our eyes were better we would see them as really tiny. Of course theory of light wasn't formulated yet either for Tycho to know that - so he picked up a simple, well-corroborated explanation for his data.

As we mentioned Solar parallax indeed exists, but it is so small that Tycho's instruments couldn't find it. It was discovered in XIXc. It is sometimes being looked at as an argument against the method, but I would not agree - people got it right when equipment got better. No one said that this is an immediate solution to all the world's problems - like not having good enough technology.

4.3.3 Summary

Based on these examples, I think it is acceptable to have background hypotheses when coming up with new hypotheses or testing them. If the result is against what we predicted, we can still just keep testing and narrow down on what failed. Limits to this approach are limits of our capability to investigate - such as equipment accuracy in Tycho's case.

We can presuppose that the physical real world exists, or that it is accurately described by Euclid's geometry (within our domain of interest) and use it in new hypotheses - for example, that massive bodies fall, that liquid spills on flat surfaces, that night and day come one after other. Empirical evidence for these hypotheses is very good.

All the previous observations don't require any application of language on behalf of the mind in question - hypotheses could be sets of thoughts in mind. When I call something 'Euclid geometry', it is to refer to concept readers often know, instead of deriving it from scratch. We will investigate the emergence of language next.

5 Other minds and language

The problem of other minds can be stated as "do other humans I see have minds or mental processes similar to one I have". I don't see their minds in the same way I see mine. I see their behaviour or I hear their talking - that's it. After our 'mind' figured out what he looks like (as we would put it), with the use of a mirror surface, he can move on to meeting creatures similar to what he has seen as himself.

After he sees some, he doesn't know what they are. But a reasonable working hypothesis would be 'they are like me' (we will refer to it as **TLM** hypothesis). They can't talk though, there's no language yet. I think that they can exchange some signs meaning anything like 'positive, pleasant, happy, agree' or 'negative, unhappy, sad, disagree' with certain mimics and behaviour - the former we will call **P-sign**, the latter **N-sign**. The mind can see, with the use of the mirror, that these correspond to certain states of thoughts they reinforce (if we smile willingly, we feel somewhat better), and are produced by these states, unwillingly. This can be thought of the special case of sensory IO (input-output).

From TLM, and the hypothesis that they want to 'communicate like I do' (we will call it **CLI** hypothesis), we can derive inferences sort of "what would I think when I saw them pointing fingers at a rock" and thus predict 'what they will think when I point fingers at rock'.

The Source of the core idea behind it can be found in the game theory:

Proposition 5.0.1 *(Schelling point) - in absence of communication, people can often concert their intention or expectations with others if each knows that others are trying to do the same"*

The quote is from T. Schelling who investigated such problems and found people can solve questions like: "three players play a game, where they need to put letters A, B, C in order and reward is given when they write same order" (Most people will write 'ABC').

This can be applied to finger-pointing and uttering sounds. I point a finger at a rock and say "rock". It can be derived that given TLM and CLI hold- I must mean something about rock, not about a tree that is next to it or about the sky above us, as I make rock 'special' by pointing at it. They use fingers to point, too and their senses focus on moving objects more than still and focus on voice more than silence or ambient sounds. They know that I know that they know that.

One peculiar detail on that: the act of raising finger could be pointing, or could be doing something like Socrates on a painting[3]. He "points" in one direction, but his face points in the other. If we point with our face and eyes and finger - then we are really pointing. Eyes are most important - from mirror, we established we use them to look at things - thus our new colleagues can think "he is looking at something", "perhaps I should be to"[4]. A lucky thing is that our eyes allow it - they

[3]Painting "Death of Socrates" by J.S. David

[4]The problem, why finger-pointing is done in the direction of the finger and not for example another way around, can be found in "Philosophical Investigations" (I think)

have whites - we can thus see where one is looking. There are experiments in psychology that work as follows: if a person on a crowded street stands looking upwards, others will start doing that too [5].

Thus, the mind meets other 'apparent humans' he can exchange these signs and also he can point fingers at objects and utter sounds, hoping they will understand on behalf of the TLM hypothesis. Predictions can be inferred, for example, I will send them a P-sign - if TLM holds and CLI holds, they should reply with P-sign. This is similar to a handshake often used in digital communication - an 'agreement' that connected parties will talk in a certain code.

[5] observations taken from J. Peterson

6 Hypothetico deductive language creation

Ostensive definitions involve pointing at something and naming it with a word - they are well known to philosophy of language (Wittgenstein "Philosophical Investigation" mention them in the beginning, quoting Augustine as a source. With H-D method and Schelling point theory we would make a more powerful way of constructing language.

6.1 Tok Pisin example

We won't be deal with babies language, I consider it rather not good for our purpose. They are notably different in their thinking than adults (no one remembers almost anything from the time he was a baby), also they learn the language as it is without participating in its development. A better suited way is to start with languages that were made from scratch not long ago (a few hundred years or less). One such language is Tok Pisin, loosely based on English and used by Papuans and other Oceanic islanders. It came into existence as a result of contact of these people's ancestors with Englishmen.

We don't know how this worked but we can look at real examples and try to explain what it could be like. Tok Pisin is a member of a broader category of Pidgin languages, often made between two groups of people trying to communicate without a common tongue or means of translation. Russenorsk, an extinct Norwegian-Russian pidgin is another example.

Let us imagine four people who don't share a common tongue. However, two pairs of these people can talk to each other. They are well aware of P-signs and N-signs as they talked to people already a lot. They can also point fingers at things and utter sounds, let it be words of their mother tongues. Let us assume without loss of generality pair will be called English, and talk English, and the other will be called Islanders and they don't talk English (without loss of generality) and English would be teaching English words to Islanders (as that is what mostly we ended up with for Tok Pisin, for whatever reason).

6.1.1 Word 'me', 'you', 'he'

One Englishman can point a finger at himself and say "Me". What it is, thinks islander: "My name is 'Me"', "I am from people of "me"', "his word for 'chest' is 'me'', "his sort of pale people is called 'me'"

The second and fourth hypotheses can be corroborated as follows: Islander points finger at other Englishman and says 'me'. But he gets N-sign from the first Englishman - thus hypothesis refuted - these are not people of 'me'. So it's something in this guy, his name, his body part. Islander then points on first English, about the place on his chest he pointed himself, utters 'me'. N response again is received, not this guy, not his body.

Two English now see they can clarify. The second Englishman points at himself and utters 'me'. He then points at the first guy and says 'me' with N-sign (not me).

Now it is clear, 'me' is when individual points at herself.

Islander can initially corroborate, by calling himself 'me' and receiving a P-sign from Englishman.

Similarly word 'you' can be established - one English person points at everyone calling them 'you', but showing that he is not 'you' himself. Another Englishman follows. "He" can be established when facing one person but pointing to other.

6.1.2 Numerals

Numerals can be explained as follows: English shows one finger, says 'one', shows two fingers, says 'two' and so on, up to four. Clearly 'one', 'two', 'three' don't mean 'fingers' in general, it could mean words referring to 1,2,3 finger specifically, but that would be a highly redundant concept. Then, English picks a few stones and again, one stone is 'one', two stones are 'two' etc.

Islander can now hypothesise what 'one', 'two', 'three', 'four' is, and he can corroborate it by showing English two sticks, uttering 'two' and getting the P-sign from English.

Then English can show a finger and say 'finger', pick a stone and say 'stone' etc. Then, he can explain plurals: 'fingers', 'stone', 'four'.

6.1.3 Living entity

English can point at all people involved uttering 'fellow' for each person. Islanders can repeat the experiment to see that word doesn't change its meaning concerning the person who uses it. And he can establish with English that rocks or trees aren't 'fellow'. Is then fellow a human, a living thing, a male, an adult?

Islander points at the tortoise as 'fellow', he could get P-sign, as a tortoise could be 'fellow' in vernacular English. Then the meaning of 'fellow' would be something like a living entity to the Islander. If he gets N-sign it would be human or something more specific.

6.1.4 Word 'We'

Let us think about the English word 'we'. It can be explained by pointing to and hypothesising as above examples, but it probably won't be. 'We' itself doesn't mean much without a context of conversation 'we - four people here, 'we - two English here', 'we - the sailors on this ship', 'we - the people of England', there's not much reason to remember it.

Now Tok Pisin's expression for 'we' is 'mitupela' - it is derived from the phrase 'me, two fellow'. It is made from concepts we discussed and it is way more exact than 'we'. Moreover, an English speaker would easily understand what 'me PAUSE two fellow' means. Me, two people. Like me and another person? Or me and two people, thus three? He can clarify that by responding 'three fellow'. The response would be 'two fellow' - two people including myself.

6.1.5 Analogies and metaphors

Associations in Tok Pisin are done with the word 'bilong', which comes from the English word 'belong'. In our example, after teaching few nouns, like shirt, stick, ship Englishman can say 'shirt belong me', 'me and he belongs ship' - pointing at the ship, shirt, himself and his colleagues. And it can be clarified with further examples on both sides, as we did before.

Having this word a lot of expressions can be made with use of analogies and metaphors - and this is what happens. Prince Phillip once remarked he is referred to in this language as 'fellow belongs Ms Queen'. Knee is 'screw belongs leg', hairs is 'grass belongs leg' (This is not real Tok Pisin, but I write obvious English equivalents for clarity). Continuous tense was made with the addition of the word 'stop' - which could be another metaphor, result of hypothetico-deductive conversation or both. If our hypothetical Islander wanted to say that a certain person is sleeping in his simplified English it could be like 'He sleep - stop'. 'He wake up?' response could be. 'No, he sleep, stop'. Many other phrases of Tok Pisin can be thought of as being the result of hypothetico-deductive approach either.

Thus, a language can be constructed with the use of hypothetico-deductive method.

6.1.6 Abstract concepts as parts of hypothesis

Notice previous example with numerals. An abstract concept is being communicated as part of hypothesis - a quantifier like 'one', 'two', 'three' cannot be explained as a thing. But it is successfully communicated using things. Let us extend it - Englishman shows his hand with three fingers extended and counts downwards - 'three', 'two','one', 'zero'. Zero is fist then? Not really, he repeats with sticks on his hand. 'Three' is to three fingers or three sticks, what zero is to lack of fingers or lack sticks.

I think this hypothetico-deductive approach is really common when teaching children mathematics. What is 'division'? Let us have two heaps of sticks. Each stick from one heap gets an equal number of sticks from the other - this number is the result. There are three children and nine candies - how many candies each one gets? A question is a prediction that student understands.

6.2 Why most languages are not like Pidgins - evolution.

Probably most languages are not like that, there are much fewer metaphors or structured phrases, where two or more words together refer to something different. 'Big cat' is one example of a composite - not being a big cat on the big picture, or big cat as an overgrown house cat - but a tiger or a lion. 'Big cat' is one of few species, that look a bit like a really big cat. It is similar to Tok Pisin word for kangaroo (big fella wallaby). Word 'computer' in Finnish is something like 'knowledge machine' - this seems to be somewhat more common for new words.

One very smart feature of Tok Pisin is that humans are good at consciously learning associations - there are memory techniques based on making up associations to whatever information we want to memorize. It could be the reason, why it spread and gained a lot of users compared to native languages. Babies are not good at reasoning - they are instead good at intuitive learning of language, so it is possible to teach everyone language early in their age, regardless of whether it is an easy language or not.

Proposition 6.2.1 *Babies learn language intuitively and quickly, adults learn language rationally and slowly.*

As for the evolution of languages, most of those used in Europe can be traced to languages used in the medieval and ancient eras and undergoing slow, gradual changes. These changes include optimizations. Tok Pisin does as well - for instance mitupela is pronounced shorter than English source. 'Screw belongs leg' is just 'screw', 'grass belong head' is just 'grass'. This could be done if frequent expression can be shortened without ambiguity - it would usually follow from the context what 'grass' refers to (unless one is a farmer in possession of a meadow).

Now, if Proposition 6.2.1 is true, one generation of people won't change their language a lot. Some people won't change their way of talking even when those around talk differently, this can be noticed if one compares the older generation to the younger. However - babies get a snapshot of the language of their parents and have 20-40 years to modify this language and pass it to their children. Thus over many generations language should evolve.

Let us look at two translations of the Gospel of John (362 years apart).

" "How do you know me?" Nathanael asked." - NIV.

"Nathanael saith unto him, Whence knowest thou me? "- KJV

The new translation is easier to clearly pronounce (in my opinion) with 'How' and 'do you'. The general composite phrase 'said unto him' was replaced with the specific verb 'asked'. Elsewhere words like 'saith' were replaced by 'said' (easier to pronounce) and all 'ye' and 'thou' become just you (the distinction is often redundant). One thus could think of it as optimization of communication. We optimize language for other uses, too - as long as we have the purpose to pick up words and use them, in place of other words. Some expressions are aesthetically pleasing, some are more polite or formal or less polite or formal, some are jokes. Pilots and the military have letter pronunciation codes: "Alpha“, "Bravo“,"Charlie“, "Delta“ etc - to spell text clearly. Those could be reasons for the evolution of language. While Tok Pisin suggests hypothetico-deductive structure, languages that are evolving for a much longer time would have lost most of it.

6.3 H-D conversations in our languages

The creation of our languages is rather hard to trace. On the other hand, I can demonstrate we are much accustomed to hypothetico-deductive thinking when using language - which is in agreement with the hypothesis that H-D influence was in our language at some point, like it is in Tok Pisin, but was obscured by subsequent changes.

6.4 Communication games

People use hypothetico-deductive method when talking in existing language as well. We could want to learn something, communicate a request, inquire about information, to double-check the information. A teller asks the client for his ID number - and reads it back to check if it is correct (as it is easy to confuse).

A doctor examines a patient - if he suspects lead poisoning - he could ask "Where you work", "do you have contact with lead" etc.

From my experience as a student or teacher, I think teaching concepts in mathematics or physics is often hypothetico-deductive, at least if there is a small number of students per teacher (a lecture is different to some extent). Often a concept is given in words, with any questions being answered and then a few examples of problems are solved, and then the student solves problems himself in subsequent classes (with the instructor's supervision). What is let's say a Langragian in physics? After the first lecture on classical mechanics, you can remember the definition, perhaps. After a few weeks of solving problems, you can apply this concept to a new problem.

6.5 Adversarial games

There is another kind of language games - where there's no intention to communicate, but one party wants to establish one thing and another party - another thing. TLM still holds, most often, CLI not really. One could call them **adversarial**. Debates, investigations, persuasion, salesmanship and frauds are some notable examples. Some of those involved in these fields use textbooks of so-called "Neurolinguistic programming" that is influencing other people to do what they want with the use of words.

Here's example of adversarial game:

> Tom knew his aunt wanted to catch him in a lie so

he said, "Ma'am, I was so hot today that I dumped water from the well on my head. See, my hair is still wet." Aunt Polly was annoyed that she had overlooked Tom's wet hair. "He skipped school again," she thought. She decided to use her secret weapon; before school each morning she sewed Tom's collar shut with thread. "Tom, show me your shirt collar," said Aunt Polly sternly. If he had taken off his shirt to go swimming, she would see the broken thread and know! Aunt Polly was really surprised because the collar was still sewn shut. "Tom, I'm sorry I suspected you of skipping school today. It appears that I was wrong," said Aunt Polly. Tom accepted her apology gracefully. He was just about to leave when Sid said, "Didn't you sew Tom's collar with white thread this morning? Look here." He pointed at Tom's shirt. "Tom's collar is sewn with black thread." "Why, yes, I did sew it with white thread today," said Aunt Polly thoughtfully, and then her face changed. It turned red, her glasses slipped to the tip of her nose, and she screamed, "Boy, that's it!" Tom turned to Sid and said menacingly. "I'll beat you good for that!" he growled. Not wanting to hear what his punishment would be, he ran out the door. "I'll have to punish him tomorrow," said Aunt Polly to herself. "I'll make him work. He hates work, but I've got to teach him a lesson."

(Mark Twain, Adventures of Tom Sawyer).

Aunt hypothesises that Tom skipped school and went swimming. If so - he should have broken the thread. But Tom predicted that already, both her suspicion and her strategy to find out. Thus 'collar was sewn shut'. But Sid predicted that if Tom indeed went swimming he would also prepared some countermeasures against getting caught and looked a bit more closely - and indeed there was a detail refuting Tom's version.

CLI hypothesis doesn't hold here, on the contrary - "They want to obscure truth“. Aunt Polly is immediately suspicious when seeing Tom's wet hair - pouring water from well on his head is hardly worth doing from Tom point of view, she thinks - thus it is an excuse.

One could compare it to mini-max algorithm used in games like chess. Computers play this game with variations of this algorithm, core idea is to hypothesise that opponent will make a move that would be most beneficial to him in the long run, knowing that I would pick moves that are most beneficial to myself. Computer thus would find a move such that e.g. in next 10 moves it would end up in best situation, on the assumption that both it and opponent choose best moves for themselves (look-ahead is limited by processing capability).

7 Late Wittgenstein theory of language

Ludwig Wittgenstein in his later publications ("Blue Book", "Philosophical Investigations") came up with an influential system of the philosophy of language.

In this system, words have no clear boundary of meaning, but the meanings of a word form a convoluted graph of relations of similarity. Meaning is determined by the way a given phrase is used. Philosophers should abandon looking for generalizations, instead just look at uses. Words are like tools he says: a hammer, a chisel, a saw - they have certain uses - similarly, words are defined by the way they are used. He gives an example of a game: game of chess, game of football, solitaire, hopscotch - all are examples of games. There's no universal abstraction of a game shared among these instances, but only some mutual similarity between instances.

One objection to this system (I know it from the preface to Polish edition by B. Wolniewicz, the translator) is as follows - we are told to just "look" at word's use, but there's no way how to look, as no theory is given and no theory could be given: (Preface to PI, paragraph 7)

"We are to establish that two words mean the same thing on behalf of the identity of their use. Very well, but how to recognize it? This critical problem is left undetermined in "Investigations"."

This would require a theory (Wolniewicz points out) - but Wittgenstein doesn't give any theory, just a bunch of loose comments with many backdoors to "sneak out".

I think I can give a theory that answers to a large part of remarks in PI: that language is being developed and understood with the use of hypothesis testing, as a means of communication. As we described: CLI, TLM and FPT are the most important hypotheses being used. - Words are related to sets of pictures ("rabbit", "brick") or sets of ideas ("four", "one meter", "walk") in mind. - Meanings of words are only in close-enough agreement among people so that communication is possible and efficient. - Universal definitions across a diverse group of speakers are impossible, similarly, very precise definitions of words are impossible. - But at the same time, there's a structure in between. A continuum of decreasing clarity, a continuum of increasing generality - as we go from small groups of people to large groups of people. - Language evolves - if few people establish new terms and uses by hypothesis-testing conversation and it spreads virally - it can be accommodated into language. On the other hand, unused terms may be forgotten.

7.1 "Pain" examples

Wittgenstein uses examples of pain, to demonstrate issues regarding mostly comparison of our mental states to knowledge about mental states of other people and related deceptive uses of language. Let us examine a few examples:

> "I grant you that you can't *know* when A has pain, you can only conjecture it", you don't see the difficulty which lies in the different uses of the words "conjecturing" and "knowing"

This is easy to clarify in our theory. We introduced TLM hypothesis as one of the fundaments of our understanding of language. This is one example where this is apparent. Of course, we have a direct sensory experience of pain (for instance) in our body and knowledge of reactions it produces in us. We don't have direct experience in other people, but well corroborated TLM hypothesis tells us that it should work the same in other people. If knowing that "I am in pain" is knowing the state of sensory input it is a somewhat better way of "knowing" than the latter. We have direct access to sensory input in the first case, we do inference from background hypothesis and sensory input in the second case. There's indeed a difference in the level of certainty, but I think it is a vast exaggeration to make a distinction between knowing and conjecture - TLM assumption is a very basic and essential thing. We commonly use phrases like "I see him smiling", "I know he's angry", "This person" - they all presuppose TLM.

There are rare situations where this doesn't work: cases when someone may pretend to be in pain: a pupil who doesn't want to go to school and wants to convince his mother that he shouldn't - for instance. But in such a case we would differentiate that accordingly in our language: "I think he is really in pain" - "I think I know he is in pain".

> "If I point to the painful spot on my arm, in what sense can I be said to have known where the pain was before I pointed to the place? Before I pointed I could have said "The pain is in my left arm" "

As we argued we know (as well confirmed hypothesis) that we can move our body and we receive sensory input from it - temperature, touch etc. We can pinch our hand or leg, produce the feeling of pain - then corroborate by predictions that it produces predictable sensory input. Then if we feel pain in our tibia, we can relate it to the most similar feeling of pinch we can produce - this allows us to point to our shin. Our brain does it somehow more efficiently of course, I am deriving a plan to know it rationally.

> "It is conceivable that I feel pain in a tooth in another man's mouth; and the man who says that he cannot feel the other's toothache is not denying *this*." " Suppose I feel a pain which on the evidence of the pain alone, e.g., with closed eyes, I should call a pain in my left hand. Someone asks me to touch the painful spot with my right hand. I do so and looking round perceive that I am touching my neighbour's hand (meaning the hand connected to my neighbour's torso). " "Ask yourself; How do we know where to point to when we are asked to point to the painful spot?" 'I said that the man who contended that it was impossible to feel the other person's pain did nor thereby wish to deny that one person could feel pain in another person's body- In fact, he would have said: "I may have a toothache in another man's tooth, but not *his* toothache' 'Thus the propositions "A has a gold tooth" and "A has toothache" are not used analogously. They differ in their grammar where at first sight they might not

seem to differ.'

Wittgenstein in this paragraph argues that seeing that person A has a gold tooth is much different than seeing A has a toothache, and yet different from having a toothache in A mouth - and rightly so. Let me explain it - in the first case we see, through our sense of sight - a person (TLM used here) with a golden bit in his or her mouth. The second one (A has toothache) is just a special case of another person's pain we dealt with already. The third case (I have a toothache in A's mouth) is yet different. I think we can't comment on an intuitive perception of such a thing - but we can indeed know rationally that such pain exists. All it takes is to take a steel dental probe and hit A's tooth (possibly after a lot of trial and error to find where the pain comes from) - if it immediately intensifies our pain (per analogy to teeth in our mouth) we would know that we have pain in A's mouth. Other pain examples (feeling pain in furniture and such) can be clarified equivalently.

7.2 Ostensive definitions

The first few dozens of paragraphs of "Philosophical Investigations" deal with ostensive definitions - the general point is that they don't work - or not necessarily work.

(PI 2) introduces situation (language-game as he calls it - a situation of practical use of language) when two construction workers communicate as follows: Worker A yells "block¡‘, "slab¡‘, "pillar" or "beam", then worker B brings him a specified item. Then extensions are given in subsequent chapters -

worker A may yell "a-slab" or "d-slab" to obtain some specific type slab or "this-there" where he points to some item and a place to carry this item to. Then in (PI 10), he makes the point that the meaning of words must be defined by their use. This is probably an argument against theories, that would suppose that "slab" is attached to a picture of a "slab" in mind.

I think he is mistaken to derive support for it from this example, and it can be demonstrated by his own method of looking at such situations in ordinary language. I think they are not uncommon - construction workers may have good reasons to communicate by yelling a handful of words - for instance, if someone talks to you from a few floors above and next to a guy cutting steel with a grinder, you are hardly able to hear anything. So, a simplified language is made of yelling. In case a more elaborate conversation is needed, you need to come closer.

How do those come into being? One is explicit agreement: they may say: "Look, I will be doing this, and you will be carrying needed items - if I yell 'block' or 'slab' you bring me what I say" - . Regulations on a shooting range typically say something like: "In case of danger anyone should issue command 'STOP!'. If such command is issued no one is allowed to shoot". A soldier, when issued a command let it be "alert¡', knows exactly what he should do, according to rules he knows by memory.

One could think of it as the context of the conversation. "You will be carrying items, if I yell "brick" you bring me brick" and then "Brick¡', "Brick¡'. I think this is compatible with the picture-theory - word "brick" referring to a picture of brick in mind. One should carry bricks if it was specified so before.

One may argue against the above point I made as follows:

1. We can't assume that it is always (in these situations) agreed upon that one would use such and such "commands".

2. Even if we can, the same thing may be called "context of conversation" or "definition of the word".

Let me now argue that this explicit agreement is superfluous and that the usual definition of a word is applicable. I am going to use focal point theory (FPT). A bricklayer may yell to his assistant "Bricks¡', without prior agreement, having the intention to get some bricks from him and be almost certain that he will be understood. Bricklayer's job is to assemble a wall or other structure from bricks and the assistant's job is to help him, by carrying needed items. Both know that and know that the other party knows that as well. So sentence "Please bring me as many bricks as you can efficiently and safely carry" is reduced to "bricks" - other parts are superfluous. "Bricks¡' can be understood as "Please bring me as many bricks as you can carry", not as "Bricks have fallen on my leg", "What a wonderful pile of bricks" or " I don't like bricks a lot" or "Bring me only two bricks this time". Bricklayer knows what the assistant should expect to do and does not say what he should do. Assistant knowing that bricklayer knows that, does what he should do by default: "Carry bricks" and as many as he can (to do his job efficiently and safely). Same for "this-there" example - on behalf of FPT and above stated facts it means "Please bring this item (one I point at) to this position there (I point at)" (and not "This thing was there before", for instance). Same for examples of (PI 21) and (PI 27)

Similarly, if I come to shop, and say "a lighter" it will be understood as "please sell me lighter". If the shopkeeper says "one dollar" I will understand it as: "lighter cost one dollar, please pay me one dollar". A ticket inspector would say "ticket" and mean "Please show me your ticket". These are all obvious applications of focal point theory. But "lighter" here remains lighter, "dollar" remains dollar, "ticket" remains ticket - meaning of the word is not affected - and picture definitions are adequate.

This is a direct answer to (PI 19): Wittgenstein says that "Slab¡' is indeed "Bring me a slab¡' shortened, but why not another way around? "Bring me a slab¡' being an extension of "Slab¡'. "Slab" could be related to the picture of the object in mind, "Bring" refers to the default action, "me" to a default object. "Slab¡' is "bring me a slab" shortened with the use of FPT.

7.3 Colour, numerals, length

(PI 29) objects that there's no ostensive definition for colours, numerals, length etc. As we already demonstrated numerals can be defined in our Tok Pisin example through hypothetico-deductive discussion. Colours are not different - we can demonstrate the concept of blue to another person by demonstrating a couple of blue items - blue glass, sky, sea, ink and also a few non-blue items. After the other person tries to give us a few more examples and receives feedback from us - showing us items in different shades of blue and getting feedback - we can establish a close-enough agreement.

It is important to comment here, why 'close-enough agreement' works well in real life. For example, if all our examples were glossy blue, our colleague could be not certain whether 'blue' includes only glossy items, or whether it also applies to something mat or rugged. But he would call mat blue as well if only there's no better term for mat blue (or he could call it "mat blue" or clarify when he needs it).

The notion of length can be taught by a more complex set of analogies. First, we establish numerals. Then we grab a stick - we call it "one meter" ostensively. We use it to mark off one meter on the ground in various horizontal directions and also on a wall - and call all those examples "one meter" too. Then we proceed to mark off 2,3,5, 10 meters. Then we repeat this drill with a different template (let it be 20 cm-long stick or 25 cm long foot or 1-meter long stride) - that's how the hypothetico-deductive process is started. Our "student" can corroborate his understanding by performing measurements under our supervision. Then we proceed to say that "town is 2000 meters away" and so on.

7.4 A mistake in demanding what cannot be given

I stress that those language expressions correspond only to close-enough agreement of ideas or pictures in people's heads and this agreement is not exact. The exact agreement can't be achieved, as our knowledge of other people's internal mental states is very limited.

Expecting otherwise entangles us in a myriad of unsolvable problems and a lengthy discussion of reading in (PI 157-165) is a good example of it. Wittgenstein starts by pointing out the example of a pupil who "reads" a word and other people wonder whether he really did or just made a lucky guess. If he reads more words then we can see he indeed reads.

L.W. here points at "contradiction": on first-word teacher says that pupil doesn't read, on reading a couple more of words teacher is convinced that pupil indeed reads. But how about this first word L.W. asks: was it read? This question is nonsensical, he answers himself, unless we assume that we start reading.

This apparent problem is very neatly solved by the method of hypothesis testing. The teacher wants to find out whether the pupil reads. "No, you don't read" can't be a conclusive statement of fact - the teacher has no access to student's internal mental states. The teacher hypothesises "you don't read". Pupil reads a bit more and the teacher is convinced otherwise. One could estimate the probability of guessing the content of the text by chance as exponentially convergent to zero with its length - which is more exact way of saying "this is impossible for all practical reasons". Teacher, in a contrived example, could say something like 'you don't read' indicating a conviction about such internal mental states - but this can be explained by FPT. Teacher and student both know that teacher can't state facts inaccessible to him and know that other party knows that. 'You don't read' is thus 'I think you don't read'.

In general, a comprehensive method of reading is taught by hypothesis testing - a teacher demonstrates how to read letters, syllables and words - and then expects the pupil to memorize, imitate him and then generalize to new words. If the pupil is mistaken, he is being corrected by the teacher. Later they work on other topics, such as understanding of the text being read. A hypothesis could be: if the pupil understands texts he should be able to pick matching paraphrase. Such exercises are common in textbooks for teaching foreign languages. And more importantly - they are an important part of exams. How do we determine proficiency in language? The student solves a bunch of exercises, writes an essay, participates in conversation - that is the content of exams.

It is hypothesis testing all over the place in the application to what can be externally observed - that's how people find out what "reading" is in our method. And this definition is shared closely-enough to allow communication. How about internal mental states regarding reading? No way to know and no reason to assume that they would be the same among different people - that's the universal answer to the rest of Wittgenstein's argument.

L.W.'s definition of "language games" is not very clear. We can think of a global definition of words: i.e. whether there's one meaning of a word for most people using it. Even if we limit ourselves to one language - pretty obviously it can't be so - there's regional variation in the language, there are lingos of various societal or professional groups etc. Thus meanings of words "in general" will end up being overlapping, convoluted graphs but that's hardly a surprise and no contrived examples are needed to see that.

Somewhere below language games starts indeed - a group of people engaging in the use of language form clear meanings of words. But is there a clear distinction of meaning 'in language game', and 'across language games'? Like: first is given use case of a word, second is convoluted graph without clear boundaries? I think not - I claim that meanings inside of game are only in close-enough agreement for communication to happen. Outside of given game they could be slightly less defined, but this is a continuum, not a discrete boundary.

Think of "Brick“, "Slab“ example - I demonstrated that picture theory is completely adequate here - "brick“ could be still understood as a picture of a "brick“ in mind (in the language game that seeks to demonstrate its inadequacy).

I think the problem here is in desiring two things 1) grand universal definitions simultaneously valid for at least most English speakers 2) definitions precisely connected to mental states shared by most people. It is not hard to see the absurdity of both by producing few counterexamples, but that doesn't mean that there's no objective structure somewhere in between (as L.W. seeks to establish).

On the contrary, if we assume that word meaning emerges as a hypothesis testing phenomenon and its meaning is an agreement of ideas close-enough to allow communication, we can explain his counterexamples.

7.5 Did N. exists?

The point in (PI 79) is as follows: a bunch of people argue about a historical person - let it be Socrates - known from written sources. Some believe the narrative that Socrates lived and performed certain things. Some deny parts of the narrative. Some of them even claim that Socrates never existed, Socrates wasn't known by that name when he lived or biography of Socrates in fact refers to a few different men.

Then the argument is made that e.g. "Socrates didn't exist" means a different thing, depending on what the meaning of Socrates is. "Socrates didn't exist", because his real name was Cleanthes, "Socrates didn't exist" because he's a myth etc.

One way to deal with it [3] is to assume that "Socrates" is a category that includes all options. Socrates didn't exist - all kinds of Socrates didn't exist, his story is mostly false and doesn't refer to any real people.

Surprisingly - application FPT/CLI produces the same result. If I say "Socrates didn't exist" I would mean that 'all kinds of Socrates didn't exist' by default (in absence of context) - I deliberately reference certain concepts to make this distinction - that's FPT at work. My audience reasonably expects that if there were two Socrates(es), I would tell them explicitly, on behalf of CLI hypothesis with FPT. There's no way to infer it from "Socrates didn't exist" without context, so it would be inefficient to say this that way - as they could know. I would deceive them with this phrase and we would lose time.

7.6 Chair example, pictures, computer games, simulations

L. W. in (PI 80) comes up with the following situation: We see a chair. "There is a chair" we think. But as we move closer, it disappears. Then, it reappears in a few seconds or reappears when we move a bit. Then he asks whether the reader had rules prepared to tell if this is still a chair. I think we have such rules indeed. What is violated in his example are the assumptions about perception of space and navigation through space. I would expect thus that chair is still a chair. We would refer to it as chair.

This works that way in various simulated realities we use in the XXI century - computer games or software used for design. A chair seen in such software is still called a "chair". Even less than that - a chair on a picture is called "chair". L.W. should know the latter well, part of (PI II) is dedicated to seeing things in pictures - a rabbit in a picture or a duck in the same, ambiguous picture. A chair in a space that sometimes disappears is much closer to a real object than a picture, so it is called a "chair".

Why do we call picture-rabbit a rabbit? First of all, "rabbit" is the best phrase in our vocabulary - it looks (mostly) like a rabbit. Why don't we call it "picture-rabbit" - the reason is again FPT and TLM. If we talk to other people in a room, with picture-rabbit on the wall, we know what kind of "rabbit" we refer to. Other people know that as well.

Imagine that we are talking by phone to a friend while playing a computer game. If we say "Hold on, bandits are here", a misunderstanding may happen. Our friend, unaware of the situation might imagine real bandits. If we play online while talking to people playing with us this phrase could be understood. I should talk efficiently and in a way that another person understands me. How to do that? TLM, CLI, and FPT are here to help.

Bear in mind that our hypothetico-deductive concepts of seeing objects, sensory experience of objects and locomotion through space are independent among each other. If we play a game with 3D graphics and a virtual reality set, we can apply the same spatial or geometric concepts as in reality. We can even sometimes forget that it is not "real" (it is an exercise left to the reader to think "real" means).

7.7 Failure to understand

(PI 185) describes pupil who is told to write down elements of the arithmetic sequence $a_k = 1000 + 2 * k$ (1000,1002,1004...) but gets it wrong (1000, 1004, 1008) - and can't be convinced that it is wrong.

Indeed it happens that some people won't understand advanced, academic mathematics. Capability and desire to do so varies through the population.

But in the given case, it is at least clear how the teacher should proceed with our method. He can do something like - "Look, you first wrote 1000, which is good, but if we subtract it from the next element, we get 4. This is in contradiction to the definition I wrote, as $a_{k+1} - a_k = 2$. how to fix that?".

The pupil can insist it is in agreement, or derivation is wrong, but in that case, we can continue towards more basic concepts. We would end up seeing that he denies something basic about math. The definition of addition in natural numbers, for instance. Why could it be that? Perhaps a deficiency similar to colour-blindness - but nothing to do with our method. It is known that our cognitive faculties sometimes fail.

8 Application to problems

We will look at two language problem that can be easily solved with our theory.

8.1 'Nothing itself nothings', Heidegger vs Carnap

Next we will look at major philosophical dispute in Germany in 30s of XX century, between local school of metaphysics (Hegel, Fichte, Heidegger) and Vienna Circle and we will disagree with both groups. Both were criticized many times and large part of our point is probably not very original, nonetheless I think that lens of our method allows to see what's wrong here with a remarkable clarity.

"Nothing itself nothings" is a quote from Heidegger. Rudolf Carnap criticized Heidegger's approach in "Elimination of metaphysical thoughts by a logical analysis of language".' Few sentences he quoted from Heidegger could be, he claims, fallacious use of the word 'nothing' as a noun, while it is a quantifier often used in negations ("Nothing can fix this issue" i.e "This issue can't be fixed (by any-thing) "). Then he claims it could also be that Heidegger's 'nothing" is something else than the ordinary use of this word - an "emotional constitution" for instance. But this is not the case, he claims, as Heidegger starts with "What is to be investigated is being only and-nothing else; being alone and further-nothing" clearly pointing to its ordinary use "nothing else", "and further nothing". Then quotes from Heidegger are given, on his dismissal of logic and science. Thus, Carnap concludes, Heidegger is, probably, talking meaningless sequences of words (Section 5)

Can Carnap establish that? It is rather difficult issue, especially for anyone not being native German speaker. There are different translations of this essay. 'being' could be 'what-is', 'Nothing nothings itself' - 'Nothing annihilates of itself' - depending on the translation. Nothing nihilates of itself. "Nothing“ is a thing that rids of this thing itself? This still looks to be nonsense as 'nothing' still must be a noun. On the other hand, the thing " that rids of itself“ involves some action and relation like meronymy or ownership being at the same time reflexive (between thing and itself). I am by no means certain there's no meaning here - but I don't know how to look for it. The riddles are not easily solved, the riddles multiply themselves when we attempt a solution. Carnap also points out that Heidegger scornfully dismisses logic and rationality - in that case, we indeed lack the tools to proceed with a solution. We don't have any clarification or response from Heidegger either.

We can't assume logic to work - that's the first problem - and we silently presupposed logic everywhere so far. When we read the above excerpts Heidegger it is apparent that it could be an adversarial language game - like one with Tom Sawyer and his aunt - in that case, we probably can't rely on the CLI hypothesis (and no way to show that we can).

Based on other Heidegger's works we can think that TLM is violated either. He [2] appeals to some sort of mystical revelation, his special inner insight, H. Phillise says:

> Heidegger relies on a epistemic model derived

> from theology, and assumes that he is the recipient of some kind of revelation... (...) Heidegger belonged to the elect, to those favored by Being, who were destined to hear Being's voice

Heidegger's mental states regarding these matters can't be related to others' mental states - thus TLM probably doesn't hold.

I would argue that Carnap critique goes too far. I would rather state that there's probably no way to know what Heidegger means, as he overthrows the most basic structures needed to understand language in our theory.

Then in section 6. Carnap extends this argument to the whole field of metaphysics - it is based on a couple of observations regarding use of language:

- Fallacious use of "to be". "To be" is often a prefix to a predicate - "I am hungry", while metaphysicists use it as a verb in "I am". - Type confusion: "Caesar is a prime number", "Twenty two was Roman general".

The story is as follows: Methaphysicist say something like that and Carnap says that it must be meaningless because it goes against such and such logical structure of language established by "modern logic".

I answer that: First, it implies no contradiction to assume that two metaphysicists that talk have associated some concept with word "being", "nothingness" or other. Cherrypicking few sentences that make no sense in ordinary use of language proves nothing - language is being extendend all the time according to people's needs, as we demonstrated. Second it is not hard to see that Hegel, for instance, was clearly interested in consciousness, self-awareness - this 'video' we see in our heads. Another story is what came out of this research (and if anything did), but it is wrong to say he talks.

Third, why one should assume that there's indeed logical structure to language? If by induction, then it would be rather bad induction (see 9.1 for clarification) as one could find counterexamples to any such idea. Could there be a real type confusion if word-meaning forms convoluted graph of family resemblance? I think not. One could say that "Ceasar is a prime number" refers to nothing in particular, but "Give me some Ceasar" at canteen will get us dressing named after restaurateur Caesar Cardini ("Philosophical Investigations" are rather effective attack on these ideas in general).

8.2 Putnam's Twin Earths

Putnam Twin-Earth problem is as follows: there are two identical planets with copies of the same people having identical sensory experiences of everything. One planet is our Earth. The other is Twin Earth - with the only difference being that our water (hydrogen dioxide) is replaced with compound XYZ with exactly the same physical properties - both populations refer to their compound as "water". Or when it freezes they refer to it as "ice", when it boils they refer to it as "steam". The date of the experiment is set in Earth's medieval era, where neither population knows any chemistry.

The question is "Whether Earthling says 'water' and his twin from Twin Earth says 'water', do they mean the same thing".

Putnam thinks not, Earthlings refer to H20, Twinearthlings to XYZ.

From our point of view, 'water' is just a hypothesis for something wet, liquid transparent, sometimes drinkable - thus they refer to the same abstraction equally applicable to H20 and XYZ, similarly as 'three' applies to 'three sticks' and 'three stones.

I think this is reflected in our language. There is a drink called 'coffee'. It can be prepared in different ways - but one could think that it should be made from roasted coffee beans (Coffea arabica or Coffea robusta). But drink called "coffee" is also made from rye, barley and chicory root - it has similar colour and similar taste. It would be called 'cereal coffee - for distinction, but there's still world 'coffee', referring, probably, to the similar experience of drinking.

How about a disk? The disk is sport equipment for throwing. An optical device for holding data is a compact disc - compact refers to the size, probably. It is abbreviated as CD. But same with same size device only slightly better is not 'compact' anymore - it is DVD - digital video disk. Of course, you can write both with music or video if you want. Of course, CD is digital storage too. So if I have a video on CD as it is often understood, I can tell my friend it is a DVD. I would not do that, because he would understand it wrong, it makes no sense. If I want to communicate with him I would clarify 'It's a video on CD, and you need such and such equipment to play it'.

8.3 Kant's transcendental aesthetics.

Kant's highly influential system explains how we supposedly know concepts like space or geometry a priori. Let us argue with it a bit to show, how different our position is (and why it is better).

Peikoff summarizes Kant's position on space as follows: 1. "Everything that comes to us through five senses, comes to us necessarily in form of space. By space, we mean spatial relationships." 2. "Space is not the content of experience; it is a way in which experience is organized". 3. "You couldn't get the idea of space from experience".

I hinted already that space as we see it, is not, in my opinion, the basic form of our experience. We see a flat picture, or a series of pictures and our brain is trained to detect spatial objects in it. Sometimes it will indeed find such objects, sometimes it won't find any and sometimes it finds objects where there are none.

It is necessary so, our visual apparatus is made of a controllable lens projecting the light on a flat matrix of light-sensitive cells - the primary format of data our vision produce is two dimensional. We can perceive distance indirectly and to a very limited extent as a) we have two eyes that may see a bit different picture when looking at a near object b) we are aware whether we are focusing sight on a close object (we feel some strain, movement). It won't work to tell whether we are looking at a small screen from a distance of 20 meters or a big screen from a distance of 200 meters (focus would be about the same as a focus to infinity and both eyes would see identical image).

Our brain reads data as two dimensional and 'infers' spatial objects - sometimes correctly, sometimes not. Consider the fake dome in the Church of St. Ignatius in Rome. It is a painting on the ceiling that appears like a dome if looked at from one certain point. If we move a bit, the illusion disappears. Our senses don't tell us that there's no real dome.

I introduced already space and geometry as a predictively corroborated hypothesis we establish by moving through space, looking at things and making predictions about how items look like, from the other side, or after we go around it a few times.

Now let us add that this notion of space cannot be the truth we know a priori, as it not necessarily always works. The same goes for a universal time at a distance. Physics of the XX century discovered that space and time is being affected by a) movement of observer b) masses, energy and momenta - a) is the result of Special Relativity and b) is the result of General Relativity. Two classical tests of GR: Mercury's apsidal precession and gravitational lensing during solar eclipse demonstrate space-time being not flat in presence of large amounts of matter. Clocks on GPS satellites go faster than they should in the classical sense, compared to clocks on Earth.

Thus I respond as follows: The idea of space can indeed be derived from experience as a corroborated hypothesis. This idea doesn't need to always describe reality, thus it is no more than a hypothesis. Experience doesn't need to come to us in form of space.

I think that two strong points may be taken from Kant's description. First of all, I think local time is indeed a form of our experience. Second, there's indeed some spatial structure to experience - either a flat structure of a picture we see with our eyes or a spatial structure emerging from the perception of depth. The inference that points at some given positions are at some distance and sections among these points are at certain angles to each other must be corroborated hypothesis, nothing more.

9 Appendix

9.1 Digression on unity of scientific method, induction and replication crisis

I would like to make a point that above popperist methods and modern, rigorously formulated induction appear to be significantly different in theory, but in practice are very overlapping, connected and mutually complementing.

This could appear a surprise in light of vulgarized or intuitive notions of induction, such as: "it happened many times, so it must happen again" and I want to stress that rigorous induction is nothing like that.

Certain ideas put forward by Popper can be traced to the thinking of earlier inductivists. Bertrand Russell points out that the business of science is to find laws that always work, without exceptions and that inductive judgements can be falsified. Then he proceeds to discuss the mistaken induction of a chicken that is fed every morning many times and expects it to be so in future. That's how our instincts work, he claims, but they could be rather badly wrong.

Bertrand Russell in "Problems of Philosophy" (1912), "On Induction"

> "The belief that the sun will rise tomorrow might be falsified if the earth came suddenly into contact with a large body which destroyed its rotation; but the laws of motion and the law of gravitation would not be infringed by such an event. The business of

> science is to find uniformities, such as the laws of motion and the law of gravitation, to which, so far as our experience extends, there are no exceptions." " The man who has fed the chicken every day throughout its life at last wrings its neck instead, showing that more refined views as to the uniformity of nature would have been useful to the chicken. But in spite of the misleadingness of such expectations, they nevertheless exist. The mere fact that something has happened a certain number of times causes animals and men to expect that it will happen again. Thus our instincts certainly cause us to believe that the sun will rise to-morrow, but we may be in no better a position than the chicken which unexpectedly has its neck wrung. "

I made the distinction between "something always happened" and "something happened many times" of intuitive induction. Our brains are often tempted to mistake "many" for "always". Moreover, it is hard to make "always" at least a good approximation. Induction theoreticians such as Solomonoff and von Mises and scientists like Feynman were aware of such issues.

For intuive, informal description of problems we can start with Feynman's "Cargo Cult Science". Title points to "Cargo Cult People" - people of Pacific Islands, who seeing military airplanes bringing supplies, coined idea that those must be divine messengers with gifts, and proper rituals done by personnel of the airfield should summon them.

> "In the South Seas there is a Cargo Cult of peo-

> ple. During the war they saw airplanes land with lots of good materials, and they want the same thing to happen now. So they've arranged to make things like runways, to put fires along the sides of the runways, to make a wooden hut for a man to sit in, with two wooden pieces on his head like headphones and bars of bamboo sticking out like antennas—he's the controller—and they wait for the airplanes to land. They're doing everything right. The form is perfect. It looks exactly the way it looked before. But it doesn't work. No airplanes land. So I call these things Cargo Cult Science, because they follow all the apparent precepts and forms of scientific investigation, but they're missing something essential, because the planes don't land.

Feynmann point is that modern scientists sometimes are about the same as Cargo Cult people, fooling themselves or others. Main problem is bringing up "evidence" for hypothesis making no effort to look for counterexamples, report possible problems, explore alternatives etc.

His solution is given as a set of informal rules.

> It's a kind of scientific integrity, a principle of scientific thought that corresponds to a kind of utter honesty—a kind of leaning over backwards. For example, if you're doing an experiment, you should report everything that you think might make it invalid—not only what you think is right about it: other causes that could possibly explain your results;

> and things you thought of that you've eliminated by some other experiment, and how they worked—to make sure the other fellow can tell they have been eliminated.

We will show that similar principles follow from rigorous formulations of induction, like those by von Mises or Solomonoff. More precisely following naive induction:

- I see many examples of A implying B. - Thus, always A follows B.

in practice is supplemented by the following rules: 1. Complexity of hypothesis must be low compared to data it seeks to describe - simple hypotheses are much preferred. 2. Other than that, all hypotheses are equally probable a priori. 3. Negative evidence is vastly more important than positive evidence. 4. One needs to objectively investigate many different hypotheses.

If we look at Solomonoff induction, 1. and 2. are given explicitly as premises (Kolmogorov complexity and principle of equivalent explanations). As for 4, Solomonoff induction uses the theoretical apparatus of Turing machine and enumerates all possible input programs to Turing machine to make sure that all hypotheses are tested - but this is not possible to do in practice. This should not be considered a limitation of the method. We can't generate all hypotheses in general and this affects any system of inference that relies on hypotheses.

Instead, we can get close to truth in the long run if we would be testing more and more hypotheses - that is essentially going from theoretical induction (Solomonoff's) to practical induction - similar to Bayesian statistics. However we need to test all possible hypotheses we can - if we skip some classes, that would likely be a mistake - thus conscious effort to look for alternatives, particularly ones we don't like is badly needed.

Point 3. follows from the mathematical form of likelihood. For instance, Bernoulli likelihood of knowing a certain hypothesis that something happens with probability p is $p^P(1-p)^N$, where N are negative cases (in disagreement with hypothesis) and P are positive ones (in agreement). If $p = 1$ (something certainly happens) and there are negative cases then the hypothesis is refuted. If $p = 0.999$ - that is something almost certainly happens - then one negative example counts as much as 1000 positive ones. Thus ignoring, or not actively looking for negative cases makes a tremendous difference.

Let us consider an example - if we lived 300 years ago and we observed swans in Eurasia, seeing thousands of swans - we would come to the conclusion that all swans are certainly white. If we then go to Australia and see a dozen of black swans our hypothesis is refuted. We could then argue, that swans are white with 99% probability - but that is not the best hypothesis. A better one (according to likelihood and posterior) would be: Swans in Eurasia are white, but swans in Australia are black.

Richard von Mises induction relies on two axioms (it is very close to Reichenbach's induction):

- If the relative frequency of a sequence of random events converges to a limit and the limit is independent of the place in the sequence we start counting, then this limit is called probability. - If we do fair bets on the outcomes of these random events it is impossible for us to make any consistent profit, regardless of the strategy we would employ.

The first point is an axiom of convergence, second is an axiom of randomness.

Consider a fair, balanced coin - it lands on heads with 50% probability - and these odds are never changed. If we make following bet: we win one dollar for heads and we lose one dollar for tails - it is impossible to consistently make money from this bet. That is a good example of von Mises induction. [6]

Axioms can be rather easily violated. If we have an asymmetric coin made of soft metal, that deforms as we toss it - it could happen that odds are different after 1000 tosses than they were in the beginning. In that case axiom of convergence is violated. We can't expect that currently observed odds would be the same in the long run. If we see that coin is being tossed with heads or tails upwards it is known that it affects the outcome a bit - in that case, we can indeed propose a strategy that outperforms a fair bet - which is a violation of the randomness axiom. Thus, bear in mind that the application of these axioms in real-world is non-trivial, except in the case of physical theories (statistical physics, quantum mechanics) and random games.

Similarly to Solomonoff case if we propose a hypothesis that something happens with certainty ($p = 1$), seeing a small amount of negative examples in a series changes a lot - the axiom of convergence is violated (this follows easily from Cauchy's definition of limit) - and it is not clear what to do about it - thus as point 3 states, negative examples change the outcome a lot. Point

[6] It could be that coin is not evenly balanced, e.g. it falls on tails 75% of the time. A fair bet would be 1 dollar lost for tails, 3 dollars paid for heads.

2 and point 4 follows from the following observations: - If we apply the above induction to two competing hypotheses and see both of them confirmed - there's no obvious reason to prefer one or the other. - If we see a strategy to outperform fair bets then it at the same time violates the axiom of randomness and is a candidate for a new hypothesis.

Point 1 - regarding the complexity, immediately follows if our hypotheses become models forecasting certain probabilities - as typically done in statistics and machine learning. Often it is dealt with a set of heuristics (bias/variance tradeoff.)

We can see a profound familiarity of both methods with the popperist method - there's indeed a lot in common in these systems:

1. Hypotheses that claim certainty are quickly refuted when negative evidence shows up (this can only happen when we do prediction) and, in Solomonoff/Bayesian case accurate hypothesis can replace inaccurate ones on behalf of a small amount of negative evidence. This is an implicit, practical hypothetetico-deductive method.

2. How to know that von Mises axioms are satisfied in the real world? We can't know precisely if the axiom of convergence is satisfied, but we can corroborate it through hypothetico-deductive method. The axiom of randomness is satisfied if we can't improve upon the existing hypothesis. Thus we need to look for alternative hypotheses as a rule needed to provide H-D justification for our axioms.

3. Complexity tradeoff is related to progressive and degenerative problem shifts in falsificationism - if theory grows more complex without an increase in predictive power, that's an ex-

ample of a non-scientific shift.

4. Induction is justified by its predictive power in the long run: "it is guaranteed to eventually approximate the limiting frequency if such a limit exists. Therefore, the rule of induction is justified as an instrument of posting because it is a method by which we know that if it is possible to make statements about the future we shall find them by means of this method" Reichenbach stated.

A major strength of inductive methods is that they can deal with probability, while hypothetico-deductive method alone can only deal with discrete rules. This comes at cost of big complexity and weaker rational grounds - thus it is generally considered that Popper's theory is more effective and provides stronger justification if it is applicable.

In this book, we concerned ourselves almost exclusively with hypothetico-deductive method, but one could think of application of Bayesian induction in cases, where one needs to consider uncertainties.

What we consider example of inductive error is so-called "Replication Crisis" in science[7][8]. There's growing concern that a large part of research findings in some fields (sometimes even most of them) are false, as they cannot be reproduced. Paper by Dacrema et al, from the field of machine learning, demonstrates how fake, apparently-excellent results can be produced by fallacious research that seeks no alternative explanations of these results by double checking all the assumptions needed for producing results (see Rule 2 and 4).

Ioannidis demonstrates replication failures in the medical research - claiming that they could be caused by faulty assumptions on the probability of a study finding a true result. The fault here is obviously not about irreproducible trials happening sometimes, but happening in such big numbers and producing confusion and loss of time and money that could be avoided. If scientists "leaned over backwards" and paid attention to negative evidence and axioms of probability were tested on regular basis, this would be nowhere near as serious.

References

[1] F. Chollett, "On Measure of The Intelligence", 2019, https://arxiv.org/abs/1911.01547

[2] Herman Philipse, Heidegger's Philosophy of Being, cited in David Auerbach, "Authority by Obscurity".

[3] "Tragedia werbalnej metafizyki", L. Chwistek, 1932

[4] Ludwig Wittgenstein "Philosophical Investigations" (in Polish), PWN , 2012

[5] Karl Popper, "Logic of Scientific Discovery" (in Polish), Aletheia, 2002

[6] Imre Lakatos, "Criticism and Methodology of Scientific Research Programmes"

[7] John Ioannidis, "Why most published research finding are false"

[8] "Are We Really Making Much Progress? A Worrying Analysis of Recent Neural Recommendation Approaches", Maurizio Ferrari Dacrema, Paolo Cremonesi, Dietmar Jannach

[9] "Immanuel Kant: Transcendental Aesthetic—Identifying the Forms of Perception", Leonard Peikoff, Lecture in History of Philosophy