<<back to home

Notes On Karl Popper

Published: January 2020. Substack version (NB. I don’t endorse many of these views, but I found it interesting to write about them.)

> *The gods did not reveal, from the beginning,*
> *All things to us, but in the course of time*
> *Through seeking we may learn and know things better.*
> *But as for certain truth, no man has known it,*
> *Nor shall he know it,neither of the gods*
> *Nor yet of all the things of which I speak.*
> *For even if by chance he were to utter*
> *The final truth, he would himself not know it:*
> *For all is but a woven web of guesses.*
>
> *— Xenophanes*

This quotation from Xenophanes, who lived around 500 BC, contains the essence of Karl Popper’s philosophy: fallibilism; gradual improvement via a process of conjecture and criticism; scientific knowledge as an externalized, objective entity (a "woven web of guesses").

Popper is one of the most underrated philosophers, because he was right about several important things. These include democracy, the paradox of tolerance, scientific method, fallibilism, and open society/anti-Marxism [1]. In addition, Popper managed to solve several important philosophical problems (e.g. induction) in convincing ways, a feat which has eluded most philosophers. So I think most people should know more about his ideas.

Let’s go into them a bit more.

Fallibilism

The essence of fallibilism is: you can never *know* whether you have reached certain truth. Every assertion is provisional, and may be false.

The best theories we have are those which have so far withstood a barrage of criticism in the form of arguments, experimental tests, and so on, and have so far *not been found to be false*. (Not: “found to be true”.)

In some cases, we even *know* that our best theories are false -- for example, we know that at least one of general relativity or quantum theory is false in current form, since they contradict each other.

Why might fallibilism be true? There are two main arguments [2]:

First, the asymmetry between verification and falsification: if your theory predicts X, then observing not-X can, under certain conditions, “falsify” your theory [3]; whereas any number of observations of “X” never conclusively proves your theory correct. (The cliche example is the theory “all swans are white”: you cannot prove it for sure! A black swan may always be lurking...) [4]

Since we cannot rule out observations that haven’t been made yet, theories will always remain provisional and may contain hidden errors.

Second, the usual alternative to fallibilism is to start with some axioms that one knows to be true; or to designate a particular *source* of beliefs (such as the senses) to be foundational, such that any beliefs that come from that source may be considered true. But the idea of a "foundational" belief is problematic because of the infinite regress problem: what does that foundational belief rest on? Either it rests on something else, in which case it isn’t foundational, or it is “self-justifying”. And the idea of a particular type of belief being “self-justifying” is arbitrary & question-begging.

Popper’s solution to this problem is to dismiss the notion of “justification” altogether: you can judge which of two theories is better using criticism, *without* having to appeal of whether belief in either theory is justified. *Any* idea, from any source, may or may not be valid, but judging ideas from the source is *not* valid.

(People sometimes dismiss emotions and prefer reason as a reliable source. According to Popperian epistemology, this is a fallacy: sometimes your emotions are right and your reason is wrong, and vice versa. Ideas must be judged on their merits, not on where they came from! So listen to your emotions.)

So if fallibilism is true, then certain knowledge is impossible, which implies that science works by attempt to *disprove* theories (error-correction), as hard as possible.

Spending your time looking for “confirmatory evidence” is a waste of time. What you should do, and what good scientists do, is sketch out what your theory *forbids,* and then try your hardest to find evidence of that; try to *disconfirm* your theory. [5][6]

This doesn't mean all knowledge is hopeless, but it does mean that we can never get absolute certainty about anything, because a criticism may always come along and destroy it. And if we ever did reach the truth, we would never be able to tell, because there isn't any known way of conclusively proving that our current state of knowledge really is the final state.

More interestingly, it implies that science is an infinite process; we will never know whether have reached the end, and may never. In practice, the more knowledge we accumulate, the more ignorance we recognize around us; solving a problem with a new theory raises a host of new, more interesting problems.

(A contrasting view would be: there are a finite number of important questions, with a finite set of knowable answers; ignorance decreases in constant proportion to our increase in knowledge; and there will eventually be an end to knowledge discovery. Popper opposes this.)

As Einstein wrote:

> *[in the struggle for new solutions] new and deeper problems have been created. Our knowledge is now wider and more profound than that of the physicist of the nineteenth century, but so our doubts and difficulties.*

Around the same time as Einstein was inventing special and general relativity, Lord Kelvin was saying that physics had no further surprises to yield; Einstein was proven right.

The history of science seems to bear this out: Newton's theories were thought to be true until Einstein's theories came along and explained the facts in a new and better way. Similarly, at some point Einstein's theories will be augmented and superseded by even better theories. The history of science is one of a long train of false theories that were replaced by better ones. Constant improvement is possible.

One may object to this by reaching for the most obvious form of certain knowledge we have, i.e. mathematical proof. Surely we know that 2+2 = 4, for example?

The counter-argument here is to ask: *why* that example; it’s not because it comes with a shiny “truth” tag on it. It’s because you can’t see any way that it could be false. In short, you’ve tried your hardest to criticize it, and you can’t disprove it.

We may be wrong at any moment! Even mathematical proofs are fallible; we could all believe that a certain proof was watertight, & then discover a critical mistake in it at some later point.

Conjecture and criticism

For Popper, the growth of knowledge is very much like biological evolution.

In evolution, we are presented with an *environment*. Various *organisms* attempt to fill the niche, and random *mutations* happen. Some of those mutations and organisms fit the environment; some don't, and die off. Over time, the process converges towards optimal organism/environment fit. Until the environment changes, and so on...

In knowledge (or science), we are presented with a *problem*. Various *theories* (and variants of these theories) attempt to solve the problem. Some of the theories are criticized and fail, so they die off; the good theories remain. Over time, the process converges towards *truth*, and the problem is gradually solved. Until the new theories give rise to new problems, and so on...

Since no theory is final, each theory presents us with a set of further problems. The best theories are very fertile: theye give rise to a host of new, interesting, and deeper problems, e.g. Newton's account of gravity then gives rise to the problem of how gravity can act at a distance; Maxwell's electromagnetic theory gives rise to the problem of whether Galilean relativity applies to the speed of light, which in turn gives rise to special relativity, and so on... [7]

A nice implication of this "evolutionary" view is that problem selection and problem generation are at least as important as problem solving in the activity of science; organisms in barren environments will take much longer to flourish compared to fertile environments. It's important for scientists to pick the right problems and then try and solve those. Michael Nielsen divides researchers into "solvers" and "generators" on this basis.

Einstein again:

> *[solving problems] may be merely a matter of mathematical or experimental skill. To raise new questions, new possibilities, to regard old problems from a new angle, requires creative imagination and works real advance in science*

Szent-Gyorgi:

> *A scientific researcher has to be attracted to these (blank) spots on the map of human*

> *knowledge, and if need be, be willing to give up his life for filling them in.*

Magee:

> *Popper considered it a waste of time for a thinker to address himself merely to a topic...there is often a feeling of so-what-ness hanging in the air, since no particular problem has been solved, or question answered. The whole procedure is arbitrary. So Popper suggests as a general principle that a thinker should address himself not to a topic but to a problem, which he chooses for its practical importance or its intrinsic interest, and which he tries to formulate as clearly and as consequentially as he can. His task is then manifest, namely to solve this problem...*

Objective knowledge

Economists sometimes distinguish between *human capital* (stored in neuronal connections) and *knowledge* (stored in books, explicit, codified). Writing then becomes the act of converting human capital to knowledge, which in turn creates more human capital. (This is the starting point for Paul Romer's Nobel-winning work on endogenous growth theory).

Just as the evolutionary phylogenetic tree is an objective fact, so too is scientific knowledge: all the books, theories and papers that constitute scientific knowledge exist independently of any human's mental state, even though they are originally created by humans.

Popper anticipated this insight in his "three worlds" argument. The three worlds are:
- World 1: Physical objects
- World 2: Mental states
- World 3: Science, art, philosophy, objective knowledge... things that we created, but which exist independently of humans

Consider two possible scenarios: (1) all technology, civilizational artifacts, and human memories of these things are destroyed, but our books, libraries, scientific papers, and engineering manuals remain; (2) same as #1, but our books, libraries, science papers and so on are all destroyed too.

In case #1, we would be able to recover civilization painfully but within some reasonable timeframe, whereas in scenario #2 we would have to rediscover everything from the beginning, which would take us millenia. This shows that the world of scientific knowledge exists in those books/libraries/papers, as an entity that exists independently of human mental states — even though they are all the creations of human creativity.

Given the existence of world 3, Popper suggests that the best way to make a contribution to scientific knowledge is to understand the current set of theories and problems that constitute the current state of scientific knowledge, and then aim to extend it by considering a particular problem or theory and aiming to solve that problem. In doing so, one may create new theories and subject them to critique. One may put forward a scientific theory without actually believing the theory; in Popper's parlance, we can "let the theories die in our stead". (Contrasted to evolution, where organisms literally die if they are maladapted to their environments.)

Implications for other areas

Popper’s philosophy has a large number of implications for other areas of life. (The physicist David Deutsch has done a lot of work drawing this out & the below is indebted to reading and discussions with him.)

(1) Politics

In political philosophy, Popper is famous for (1) his theory of democracy (2) the paradox of tolerance. Both more or less follow from the above epistemology.

The theory of democracy goes like this: typically, political philosophers ask the question "who should rule?" But mistakes are inevitable, and nobody actually knows the best way to govern. The most important criterion for a society overall is that it makes progress and creates knowledge at the fastest rate possible on how to make its citizens affluent. In order to do this, citizens must be able to make mistakes and correct them. Thus, the essence of democracy isn't "who should rule", but *the ability to remove a bad ruler*, which is the same as being able to correct a mistake.

(For those of you who follow UK Politics, this is Dominic Cummings’s worldview as set out in his blog. Current political systems aren’t error-correcting fast enough to adapt to a rapidly changing world, which means they have to be reshaped to be able to do so. Technology helps get a better view of what’s actually going on — such as the effects of various policies — which in turns makes correcting errors easier; hence his emphasis on tech.)

The process of governance should follow the same process as that of all knowledge creation: a government commits to trying a set of policies to solve certain problems. After a sane period of time, say 5 years, they are assessed on whether or not they solved those problems, whether citizens believe they will continue to solve problems well, and so on. If citizens think they made a mistake electing that government, they should be able to kick them out (without violence) and enact a new government. The most important thing in knowledge creation, including political knowledge, is *error-correction*. This is Popper's criterion.

One interesting implication of it is that First Past the Post electoral systems are superior to Proportional Representation systems. In brief, this is because PR typically leads to coalition governments where it is not clear who is to blame for a given policy, and as a result the country as a whole doesn't learn much and doesn't make progress. FPTP leads to a clearly accountable government with the ability to enact its policies and an easy means of removing them if they fail. (David Deutsch explains this more in this YouTube video.)

The "paradox of tolerance": because of fallibilism, nobody actually knows anything, so everybody must be free to make guesses, criticize others' guesses, and engage in reasoning and free speech, so that we can improve our theories over time. However, anything that impedes this process should not be tolerated, because impeding this process means impeding the entire process of learning itself, which means that we will never make progress. Thus, we should tolerate people and ways of life but we cannot tolerate those that are intolerant, i.e. that seek to shut off criticism, free speech, and the means of making progress. Liberal democracies thus need to guard things like freedom of speech and critical thought jealously, otherwise they may end up failing. In short, running a tolerant polity requires a certain degree of intolerance towards anything that would threaten the polity's existence.

(2) Parenting & Education

Although Popper didn't explore this himself, I think his philosophy entails taking movements such as Taking Children Seriously, Unschooling, and the homeschooling movement more seriously, at least at the margin.

The premise of these movements is that most parenting/education philosophies take it for granted that you have to coerce children into doing various things for their own good. But fallibilism implies that nobody knows for sure what good is, and good criticism must be taken seriously as an objective idea, regardless of the source. So in the case where the parent and child disagree, they must work it out using reason, which means *not* forcing children to do a certain thing. Coercion cuts off the reasoning/learning process by saying "person X is right because they are person X", which violates the rule that the source of an idea doesn’t matter, only the idea itself does.

Popperian epistemology also implies an account of education as being about each person creating knowledge for themselves, rather than “receiving” knowledge passively (which he called ‘the bucket theory of the mind', i.e. the idea that you can just pour ideas into the mind the way you pour water into a bucket). Reading instruction, or hearing them, is merely the beginning of the process; you then have to guess what the meaning of what you're reading/hearing is, and synthesize the essential "thing" behind the words, which is a highly active process.

For example, suppose we're both sitting in a lecture by Popper, and I ask you to imitate Popper's way of speaking; as any AI researcher knows, this instruction does not speak for itself. Should I copy Popper's Austrian accent? Should I be standing? Should I face the back of the room, as Popper is doing? Should I be copying Popper's formal diction, or just the content of his thoughts? Etc. As this set of questions makes obvious, understanding pretty much anything involves an active process of interpreting the meaning behind such statements, and this understanding may be revised as further thinking occurs. This is what Popper means when he says:

> *In fact, I contend that there is no such thing as instruction from without the structure, or the passive reception of a flow of information that impresses itself on our sense organs.*

(3) What to work on

Nobody can predict where important knowledge is going to come from -- we can guess, but there will always be an element of surprise. This is notoriously true in scientific and technology fields, which are full of entertaining examples of random nobodies/underdogs coming out of nowhere and reinventing fields based on seemingly ridiculous ideas. (See this excellent Quora answer, on how startup ideas often seem dumb, for an entertaining example.)

It is often said that people should work on the most impactful thing. But by the above, one doesn't know a priori what the most impactful thing is, because one doesn't know the state of future knowledge. Lots of impactful innovations come from working on areas that seem remarkably unpromising but are interesting/fun to the researcher.

For example, most of the 20th revolution in biology was driven by a bunch of nerds playing around with fruit flies, and the entire field of genetics comes out of an obscure Swiss monk called Gregor Mendel who was cross-breeding various strains of peas. Szent-Gyorgyi won the Nobel Prize for discovering Vitamin C (ascorbic acid) by investigating why bananas turn brown but some other fruits don't. Newton discovered much of contemporary mechanics, optics, and cosmology, but he also spent time on alchemy and Biblical numerology. As Paul Graham has pointed out, he couldn't see in advance which of these would bear fruit; things that are "huge if true" are worth working on, but one cannot guarantee that they will lead to important knowledge!

What you are interested in — what is "fun" — is an important signal of which part of the knowledge tree you should be trying to grow. As Popper says, seek fun problems and fall in love with them.

(4) Creativity

Another implication: there is no known way of reliably generating ideas that are only good; "idea generation" software has an error rate that is irreducible. (If there were such a thing as always generating only correct ideas, then one would be infallible, which we have shown is not possible). Nobody knows how ideas are generated; but if you want to generate good ones, you are best served by generating a lot of them (create mutations) and then filtering hard (selection) and pursuing the good/promising ones.

This has practical implications. From Arthur Jensen's paper on creativity:

> *I once asked another Nobel Prize winner, William Shockley, whose creativity resulted in about a hundred patented inventions in electronics, what he considered the main factors involved in his success. He said there were two: (1) he had an ability to generate, with respect to any given problem , a good many hypotheses, with little initial constraint by previous knowledge as to their plausibility or feasibility; and (2) he worked much harder than most people would at trying to figure out how a zany idea might be shaped into something technically feasible. Some of the ideas that eventually proved most fruitful, he said, were even a physical impossibility in their initial conception. For that very reason, most knowledgeable people would have dismissed such unrealistic ideas immediately, before searching their imaginations for transformations that might make them feasible.*

Jensen again:

> *The individuals in whom this mental manipulation process turns out to be truly creative most often are those who are relatively rich in each of three sources of variance in creativity: (1) ideational fluency, or the capacity to tap a flow of relevant ideas, themes, or images, and to play with them,*

> *also known as “brainstorming”; (2) what Eysenck (1995) has termed the individuals’ relevance horizon; that is, the range or variety of elements, ideas, and associations that seem relevant to the problem (creativity involves a wide relevance horizon); and (3) suspension of critical judgment. Creative persons are intellectually high risk takers. They are not afraid of zany ideas and can hold the inhibitions of self-criticism temporarily in abeyance. Both Darwin and Freud mentioned their gullibility and receptiveness to highly speculative ideas and believed that these traits were probably characteristic of creative thinkers in general. Darwin occasionally performed what he called “fool’s experiments,” trying out improbable ideas that most people would have instantly dismissed as foolish. Francis Crick once told me that Linus Pauling’s scientific ideas turned out to be wrong about 80 percent of the time, but the other 20 percent finally proved to be so important that it would be a mistake to ignore any of his hunches.*

Further Reading

Bryan Magee’s “Popper” is a nice short overview of Popper’s work.

“Popper Selections” is a great primary source: excerpts from various of Karl Popper’s books, most of which are collections of essays.

The best Popper books to start with are “Conjectures and Refutations” and “Objective Knowledge”.

Many more implications of Popper's philosophy are sketched out in David Deutsch's “Beginning of Infinity”.

Thanks to David Deutsch and Karl Wilzen for commenting on drafts of this.

----

[1] Any thinker who was strongly against Marxism, for reasons which remain valid today, at a time when Marxism was in intellectual fashion and indeed seemed like a good bet, is impressive in my view.

[2] I don’t attempt to prove this thesis rigorously; this essay is just a sketch of Popper’s viewpoints.

[3] “Falsify” is a slight simplification. It makes the theory worse; theories can’t be decisively falsified, but there is such a thing as a “really good” theory and a “really bad” theory, and rationality means we pick the best ones.

[4] There is an elegant analogue of this asymmetry in Buddhism. Buddhism observes that there is an

asymmetry between suffering and pleasure; suffering is more bad than pleasure is good. A small amount of suffering can ruin any amount of pleasure, but pleasure does not fix suffering. Moreover, suffering and death are inevitable (for now), and no amount of pleasure changes this fact. One can cope with an infinity of pleasures and never have perfect satiation -- one can never reach the perfect upper bound -- but there is a lower bound we eventually reach, i.e. death. Human life can be thought of as a downward-biased random walk in which suffering and bad feeling are inevitable; the only rational response is to stop playing the game and reach perfect equanimity, etc.

[5] Chapter 22 of “Harry Potter and the Methods of Rationality” by Eliezer Yudkowsky is a memorable dramatization of this way of doing science.

[6] This, in short, is why astrology isn’t a science: it seeks confirmatory evidence, but never sets itself up to fail. It explains everything, and therefore nothing! Good theories take epistemological risks.

[7] This account contrasts radically with the induction-based conception of science, which says that one starts with data and then finds patterns in that data. The problem with this account is that science doesn’t *have* to start with data. Moreover, all 'observing' involves a human making decisions -- implicitly or explicitly -- about what features of the observed sense data are relevant, and interesting, and require explanation. In that sense, "all observation is theory-laden". Popper used to demonstrate this in lectures by asking his students to just “observe!”, prompting the question “what should we observe”?