Commonsense Physical Reasoning in Humans and Machines

Mécanismes de l'intuition mathématique chez les êtres humains et les machines Collège de France Ernest Davis January 13, 2023

What the talk is about

People (and animals) interact in many different ways with an immensely complex physical world.

Computer systems (artificial intelligence) for many applications will have to do likewise.

How?

- How do human beings (or animals) carry out commonsense physical reasoning?
- How can you build an artificial intelligence program with commonsense physical reasoning?

Outline of talk

- What do I mean by "commonsense" physical reasoning?
- Why is it important?
- Characteristics of commonsense physical reasoning
- A few notable experimental results in people
- Approaches
- Challenges and directions

What do I mean by "commonsense"?

- Learned by experience. Not formally taught in school.
- Mostly learned young (by 7 or so).
- Common. X can assume that Y knows it, and so he doesn't have to explain it to Y.

Why is commonsense physical reasoning important?

It is a component of many other intelligent tasks:

- Action
- Vision
- Planning
- Design
- Language
- Learning
- Science

Commonsense physical reasoning is part of: Acting, Planning, Vision



Design

From Jacques Carelman, *Objets Introuvables*



Commonsense physical reasoning is a part of: Language use and comprehension

E Google Translate

FI	RENCH - DETECT	ED ENG	SLISH	GERMAN	RU	ISSIAN	~		
Au théâtre, Henry qui est très grand, était assis devant Marie, > > sa vue sur la scène était donc bloquée.									
*	GERMAN	ENGLISH	FRENC	н	~				
	In the theatre Marie, so his	e, Henry, w view of th	/ho is ve ne stage	ry tall, v was ble	vas se ocked.	ated in fr	ont of	**	

Commonsense physical reasoning is a part of: Learning, Science

 "Subpopulations of a species can become isolated if the water level of a lake falls, dividing it into two lakes."



Chemical reaction



Passing steam over heated iron filings, the iron rusts, and you generate hydrogen.

 $2Fe + 3H_2O \rightarrow Fe_2O_3 + 3H_2$

Faraday, The Natural History of a Candle

Characteristics of commonsense physics

- Wide range of phenomena
- Dependent on complex spatial properties and relations
- Robust under partial knowledge
- Wide range of scales

Wide range of phenomena

- Different types of materials: Hard solids (utensils, potatoes). Soft solids (cheese, fresh parsley). Powders (flour, salt). Liquids (milk). Light. Fire.
- Different kind of processes.
 - Cutting, warming, chilling, mixing, dissolving, kneading, carrying, breaking (eggs) ...

Tools for cutting: knives, scissors, shears, wire cutters, files, lawn mowers, cheese graters ...

 Biophysics. What human bodies (your own and other people's) can do, in terms of motion, manipulation, perception, and biological processes.

Complex spatial properties and relations



Robust under partial knowledge

A person/AI knowledge of a physical situation may be incomplete because:

- Limits of perception.
- The situation is described in language.
- Some of the information is inferred.
- You need to reason generically.
- You are in the middle of designing it.
- Unknown external events.
- The physical processes are only partially understood.

Partial knowledge

You probably don't know the physical chemistry of frying an egg.

But you do know: If you pick up the pan, the egg stays inside.

As the egg cooks, it becomes easier to pick it up.

The more raw egg you start with, the more fried egg you end up with.



Reasoning with partial knowledge

If you chop an iPhone in half with an axe, it won't work any more.

Wide range of scales

An event that lasts 1/10 of a second may affect your life for 70+ years. Ratio = 10^9

Driving from Paris to Delhi (6000 km) involves shifting the steering wheel by centimeters.

Ratio = $6 \cdot 10^8$

In scientific reasoning, of course, these ratios are much larger.

A few notable experimental results in human physical reasoning

- Physical reasoning in infants
- Erroneous predictions
- Cutting pendulums, revisited
- Visualization

Pre-verbal infant understanding

An infant looks at a situation set up by the experimenter.

If what they see surprises them, then they will stare longer than if it is what they expect.

Violations of physical law surprise them.

Infants 2.5-5 months are surprised when an object is hidden in one place and reappears in another, when an object disappears from behind a screen, when object goes through an obstacle behind a screen, etc.

Lin, Stavans, and Baillargeon, "Infants' Physical Reasoning and the cognitive architecture that supports it."

Erroneous predictions

The diagram shows a thin curved metal tube. In the diagram you are looking down on the tube. In other words, the tube is lying flat. A metal ball is put into the end of the tube indicated by the arrow and is shot out of the other end of the tube at high speed. Draw the path the ball would follow.

Michael McCloskey, Naïve theories of motion



Erroneous predictions

The diagram shows a thin curved metal tube. In the diagram you are looking down on the tube. In other words, the tube is lying flat. A metal ball is put into the end of the tube indicated by the arrow and is shot out of the other end of the tube at high speed. Draw the path the ball would follow.



Erroneous predictions

Suppose you have a swinging pendulum and you cut the string at some point in its motion. Draw the trajectory of the bob.



Follow-up to the pendulum experiment

If you ask subjects to place a bucket to catch the bob, then they do much better.

Smith, K., Battaglia, P., & Vul, E. (2018). Different physical intuitions exist between tasks, not domains.

Erroneous prediction: Anecdote

I regularly scuba dive, and on one weekend trip two friends and I decided to do a night dive. We needed our weight belts, which were on a boat in the harbor, so I offered to fetch them. I swam the few meters to the boat in my dry suit, clipped my own weight belt around my waist, held one weight belt in each hand, slipped overboard, and headed back to the ladder. To my surprise, I immediately sank to the bottom of the harbor.

-- Rebecca Lawson, Mirrors, mirrors on the wall. . . the ubiquitous multiple reflection error

Visualization

(b) Pulley problem

(a) Gear rotation problem

The diagram shows two interlocking

gears. Will the knob on the leftmost

gear mash with the groove on the

right gear if the gears are rotated

inward?



The diagram depicts a pulley system. When the free end of the rope is pulled, will the lower pulley turn clockwise? (c) Water pouring problem

The diagram shows two glasses of water. The glasses are the same height and filled to the same water level. If the glasses are tilted, will the water pour out of the two glasses at the same or different angles of tilt? If they are tilted at the same rate, which will pour first?

Hegarty, M. (2004). Mechanical reasoning by mental simulation.

Visualization vs. Qualitative Reasoning

Subjects were shown a video of the ball moving (blue lines not there). Which would the ball reach first: The red area or the green area? Smith, K. A., Dechter, E., Tenenbaum, J. B., & Vul, E. (2013). Physical predictions over time.



Answer time proportional to actual time

Answer immediate

Approaches to commonsense physical reasoning in cognitive psychology and Al

- Physics engine / simulation
- Symbolic reasoning
- Machine learning / pattern learning

 Reinforcement learning
 Deep learning
 - \circ Language models

Simulation / physics engine "Video game physics engine in the head"

Basic idea:

- You know the state of the world at some starting time.
- Use the laws of physics to predict how the world will change over a tiny time period (fraction of a second). Compute the next state.
- Iterate.

You generate a "mental movie" of what is going to occur.

Probabilistic version (Monte Carlo)

Either the starting state, or the laws of physics, or both are probabilistic.

You run multiple simulations, making random choices.

Do statistics over the outcomes

Extensively used in computer software. Proposed in cognitive theories.

Two versions of cognitive theory

Construct an actual "mental movie" that you can view in mind's eye.

Vs.

The simulation is not accessible to introspection. Its outcome is used subconsciously in carrying out tasks.

Strengths of the simulation theory

- Very general approach in principle for prediction
- Huge mathematical (back to 1700s) and computer (back to 1940s) theory. Exceptionally well understood.
- Lots of powerful software, both for scientific computing and for video games
- Some experimental evidence of various kinds and degrees of cogency.

Weaknesses of simulation theory

- Requires complete specification of starting state (or well-defined probabilistic distribution).
- Requires physical theory stated as a differential or difference equation.
- Fundamentally for prediction, only indirectly for other tasks.

Weaknesses of simulation theory (cntd).

Physics engines are hard to construct:

- Video games rarely have a physics for doors.
- In scientific computing, constructing a model requires immense expertise and labor.
- Building a useful simulator for a robot/human being is exceptionally hard: "The reality gap".

In scientific computing, expertise is also needed to

- Choose the appropriate model
- Formulate the problem
- Interpret the output.

Weaknesses of simulation theory (cntd).

Unnecessary precision demanded in input and produced in output.

The cyclist is carrying a half-filled, closed water bottle on a bumpy road.

Predict: The water stays in the bottle.



Symbolic reasoning: Qualitative reasoning

Characterize the state of a system in terms of one-dimensional parameters.

Characterize dynamics in terms of sign relations: "If X goes up, then Y goes down."

Often works well for systems with few degrees of freedom. E.g. swinging pendulum.

Some successful practical applications.

Doesn't work well for geometry

Knowledge-based methods: Logical analysis

 $\forall_{t:Time; o:Object}$ FeasiblePlace(o,Place(t,o)).

Every object always occupies a feasible place.

$$\forall_{p,q:Object; t:Time} p \neq q \Rightarrow$$

DR(Place(t,p), Place(t,q)).

Any two objects are spatially disjoint.

∀ _{o:Object} Continuous(HPlace(o))
 An object moves continuously

Problem statement for water bottle

Given:

- RigidObject(Bottle). RigidObject(Cap). Stuff(Water).
- Holds(Ta,ClosedContainer(Place(Bottle) U Place(Cap), Water))
- Throughout(Ta,Tb,Sealed(Bottle,Cap))

Infer:

Holds(Tb,ClosedContainer(Place(Bottle) U Place(Cap), Water))

Strengths of symbolic approach

- Incomplete information and physical theories
- Not inherently geared toward prediction
- Can incorporate non-physical information
- Deals with range of scales.

Weaknesses of symbolic reasoning

- Extremely hard to formulate adequate theories.
- Extreme degree of fussy logical precision required.
- Hard to be confident (impossible to be sure) that no absurd inferences can be derived.
- No generally adequate inference engines exist
- Very few practical applications so far.
- Questionable how plausible as a cognitive theory.

Current methods of machine learning

- Reinforcement learning
- Deep learning neural network learning
- Large language models

Some form of learning theory must be right. Much of commonsense physical knowledge is learned by babies/children. Whatever is innate was learned by evolution.

Reinforcement learning

A creature / a real robot / a simulated robot learns by trial and error what actions work to achieve specified goals in a given situation.

As a cognitive theory, this was behaviorism (B.F. Skinner).

Strength: Practical successes in some forms of AI, both robotics and game playing. AlphaGo etc. use forms of RL.

Weakness: Limited to learning relation of action to goals. Does not learn domain theory. Very limited as cognitive theory.

Deep learning

- A very general model of learning to carry out intelligent tasks.
- Trained from large to enormous data corpora.
- Remotely similar to neuronal activity in the brain.
- To some extent, extracts abstract features
- Enormously successful in AI. A major component of almost every practical application in AI in the last decade.
 - Completely dominant in vision and natural language
 - Combined with reinforcement learning in robotics and games playing

Limitations of Deep Learning

- Hard to incorporate domain knowledge.
- Focuses on unimportant feature; e.g. texture rather than shape in computer vision.
- Fails to generalize. A self-driving car trained in one city may not work in another.
- Opaque. Results cannot be explained.
- Cannot be debugged. All you can do is change the training set or the architecture.

As applied to physical reasoning

- Important successes in highly specific scientific problems. E.g. AlphaFold as applied to protein folding.
- Limited success as applied to robotics.
- Little success as applied to commonsense physical reasoning. Little reason to think that it finds the "right" generalization or abstract categories.

Large Language Models – GPT3, ChatGPT, etc.

- Deep learning models specifically for language. Trained from huge text corpora on task of predicting the next word.
- Output is often breathtakingly impressive
- Output is sometimes hilariously off-base.
- Subject to hallucinations (invents facts, because they sound good). "Stochastic parrots".
- Commonsense reasoning, particularly spatial and physical, is a particular weak point. No inherent model of space, and looking at text is not a good way to learn about space.

ChatGPT



The trophy didn't fit in the suitcase because it was too small. What was too small?

ChatGPT

The trophy didn't fit in the suitcase because it was too small. What was too small?



ĐΔ

In the statement "The trophy didn't fit in the suitcase because it was too small," the trophy is the object that is too small to fit in the suitcase. The suitcase is a larger container or enclosure, and the trophy is the object that is being placed inside the suitcase. However, without more information, it is not clear why the trophy was too small to fit in the suitcase or what size or dimensions the trophy and suitcase are.



I'm frying an egg, but when I flip the egg I use too much force. What happens?



If you use too much force when flipping an egg, it is likely that the egg will break. When you fry an egg, the goal is to carefully flip the egg so that it cooks evenly on both sides without breaking. Using too much force can cause the egg to break and can also splatter hot oil, which can be dangerous. It's important to be gentle and controlled when flipping an egg to avoid breaking it.



What part of the egg is breaking?



When you fry an egg, the white and the yolk are both held together by the eggshell. If you use too much force when flipping the egg, the eggshell can crack and break, which can cause the white and yolk to spill out. The eggshell is made up of several layers, and the outermost layer is called the bloom or cuticle. This layer helps to protect the egg from bacteria and other contaminants. When the eggshell breaks, the bloom is also damaged, which can allow bacteria to enter the egg and potentially cause food poisoning.

Overall

Multiple cognitive processes

- Learned manipulation strategies
- Visualization
- Direct perception
- Corpus-based learning
- Language based method
- Analogy
- Abstract reasoning

Overall

Multiple cognitive processes

- Learned manipulation strategies
- Visualization
- Direct perception
- Corpus-based learning
- Language-based methods
- Analogy
- Knowledge-based/abstract reasoning

- We do not have adequate theories of any of these.
- We have no idea how to integrate them.
- Multimodal learning may advance things somewhat but probably won't give a complete solution.

One possible point of leverage

Diagrams that have only a vague qualitative resemblance to the actual situation can be very useful aids to understanding.





Feynman, Lectures on Physics

"Red shift and blue shift" Wikipedia

Thank you!