Questioning the Intuitive Physics Engine: An Occam's Razor View with Cognitive Generalization

Zhiyuan Zhang
Department of Automation
Tsinghua University
z-zy20@mails.tsinghua.edu.cn

Abstract

Humans possess the ability to make fast, approximate estimations of physics, from predicting the falling direction of a tower to inferring the mass ratio of two colliding balls given the trajectories. Yet the mechanism under the hood is not known. A popular explanation is the Intuitive Physics Engine (IPE) hypothesis, which posits that there is a simulation engine of physics in our brain that is responsible for the predictions. However, in this paper we argue that this assumption is not necessary, and should be replaced by a more general model building process that is key to human intelligence. We first argue that the model building process is compatible with the IPE hypothesis but more general and simpler, thus we should use Occam's razor to remove the IPE hypothesis. Then we demonstrate that the model building process is consistent with many other findings, including evidence from language acquisition. Finally, we briefly discuss the problems of addressing the intuitive physics using pure deep-learning based methods and propose future directions of bringing an explicit model building process into deep learning.

1 Introduction

Humans have a remarkable skill for understanding physical events, such as predicting where a ball will land when thrown. This is a challenging task for machines to accomplish. Despite our ability to understand these things almost effortlessly, the way our brain does it is still not fully understood.

One popular explanation for this human ability is the Intuitive Physics Engine (IPE) hypothesis[3]. This theory suggests that our brains contain a specialized "engine" that simulates physical interactions to make predictions. Researchers have conducted experiments to support this idea, including one where a computer simulation mimicked human judgments about physical events[1]. The simulation used probabilistic and approximate methods to make predictions, and the results were strikingly similar to human predictions, even when humans commonly made errors.

While the Intuitive Physics Engine hypothesis offers an intriguing explanation, this essay will argue that it might be an unnecessary assumption. Using Occam's Razor as a guiding principle, I propose that a more generalized learning mechanism can account for our intuitive understanding of physics as well as other domains (language, intuitive psychology). This alternative view suggests that our brains are exceptional at learning from data, forming abstractions, inducting rules and generalizing well across different scenarios.

2 Summary of the IPE hypothesis

The Intuitive Physics Engine (IPE) hypothesis claims that humans have a specialized cognitive mechanism, akin to a simulation engine, to understand and predict physical phenomena. According to this theory, the human brain runs quick, approximate simulations to forecast outcomes in a physical world. These simulations work probabilistically and approximately, thus leading to the mistakes people make in predicting physical results[1].

For this hypothesis, perhaps the most compelling evidence comes from a computational experiments that almost mimiced human behavior with an IPE[1]. This engine was designed to operate on approximate and probabilistic simulations, and researchers tested the simulation results and compared human predictions on 5 tasks, finding that the simulation's predictions were surprisingly closely aligned with human predictions, even when humans commonly made mistakes. Other approaches, however, cannot account for the mistakes well. This is considered the strongest evidence for the IPE hypothesis. However, taking inspirations from other domains, we argue that there exists a more general formulation that is compatible with the IPE hypothesis and can Occam-razor the IPE hypothesis.

3 An alternative explanation: cognitive generalization

While the Intuitive Physics Engine hypothesis has the simulation experiment as empirical support, it may not be the simplest or most comprehensive explanation for humans' intuitive understanding of physics. According to the principle of Occam's Razor, the simplest explanation that fits the data should be preferred. In this regard, a more generalized cognitive mechanism may be better.

The alternative theory claims that, the human brain excels at learning from data, forming abstractions, inducting rules, and then generalizing. Or we may term this process as "model building". This ability is shared across domains such as intuitive physics, intuitive psychology, and language. The process begins with data collection through observation (maybe also interaction with the environment). Then brain then fits this data by building forming abstractions and building models, inducting rules that generate these data. Finally the brain reaches a stage where the model is good enough that no more surprises are met, and learning in this domain stops.

3.1 Compatibility with the IPE hypothesis, using Occam's razor

This general learning mechanism is fully compatible with the IPE hypothesis. If we view learning as a model building process, then for learning world physics the model, or the rules behind are Newton's rules. So a good model that fits the data of daily observations, must somehow align with, or approximate Newton's rules. People's beliefs about Aristotelian model of physics[3] are the approximations of Newtonian physics. Then, the unconscious use of this mental model, may lead to results similar to a disturbed Newtonian physics simulation engine.

3.2 It's all a learning process

By forming it as a learning (then generalization) process, we may find better consistency with other evidence. First, infants learn the physics while developing. Second, humans are better at predicting physical outcomes that people are more familiar with (e.g. human prediction is better "when an object exiting a curved tube is replaced by water exiting a curved hose" [3]), which corresponds to the imperfect generalization of the mental model: the model surely performs better when data is similar to the samples in the training set. Third, false physical intuitions can be corrected, if exposed to enough training samples (e.g. when I was in senior high school, I watched videos of Newton's cradles and practices lots of homework about conservation of momentum, then I found that I had perfect intuitions to know which ball is heavier given the trajectories of two balls colliding).

3.3 Evidence from language learning

Evidence supporting this general learning mechanism can be also found in how humans acquire language. Just like in physics, language learning starts with data collection—listening to and processing sentences from others. Over time, the brain abstracts the rules of grammar and syntax, allowing individuals to produce novel sentences that they've never encountered before, yet are grammatically correct. For language generation, we do not assume that in our brain there is a "grammar engine" that goes through the set of grammars and check or search the proper arrangements of words. Also, young children sometimes over-regularize words, like "feets" and "goed", indicating that children are indeed discovering rules from the sparse data they get (compared to current LLMs), and applying them.

Again, This view is in line with the 'learning as model building' perspective, which emphasizes the brain's ability to construct mental models based on data and then generalize well. It provides a unified framework for understanding various human cognitive abilities, simplifying our understanding of the brain's functions, and proposing new directions to work with.

4 Problems with deep-learning based intuitive physics

While deep learning algorithms have made significant advancements in various fields, they lag behind when it comes to intuitive physics understanding and generalization. Current models, such as Convolutional Neural Networks (CNNs) used for physics predictions in segmentation tasks[2], lack an explicit understanding of the data they process. For instance, these networks can only recognize patterns but have no inherent understanding of an "instance" or "gravity", which results in poor generalization abilities.

In tasks that require predicting the stability of tower-like structures, deep learning algorithms fail to account for changes in object appearance. Any variation in the objects used in the task, or increase of complexity leads to unpredictable outputs, demonstrating the lack of a comprehensive "mental model" within the algorithm. This highlights a fundamental difference between deep learning models and the human cognitive system. Whereas deep learning models require massive amounts of data and still struggle to generalize, humans can intuitively form models and rules from sparse data, and generalize well across diverse scenarios.

We believe that to achieve human level physics understanding, the model must build proper "mental model" of the world. An end-to-end training fashion without much inductive bias could result to an arbitrary mapping that satisfies the training set, yet the generalization coming from proper models are completely not guaranteed. Without a proper model building process, the representations may be incorrect, lacking abstraction, cannot used with other tasks, and generalize poorly. We may start with bringing inductive bias that are good for explicit model building (e.g. "abstraction" is good) into the strong expressive power of DNNs.

5 Conclusion

This paper has explored the limitations of the popular Intuitive Physics Engine (IPE) hypothesis in explaining human intuitive physics understanding. We give an alternative explanation, emphasizing a general learning and model building process as the core cognitive mechanism. This alternative approach is not only simpler but also consistent with human cognitive abilities across multiple domains, as demonstrated by evidence from both intuitive physics and language acquisition.

We argued that the general model building process is fully compatible with the IPE hypothesis, making the latter an unnecessary assumption based on Occam's Razor. In doing so, we presented a unified framework that ties together various cognitive processes, simplifying our understanding of the brain's complex functions.

Furthermore, we discussed the shortcomings of existing deep-learning methods in capturing human-like understanding of intuitive physics. These algorithms, despite their successes in other areas, lack the capacity for generalization and proper "mental model" formulation, thereby failing to mirror human cognition effectively, and we may devise ways to bring a more explicit model building process into DNN training/inference procedure.

In sum, this paper encourages a shift from specialized, domain-specific explanations like the IPE hypothesis, to a more generalized understanding of cognitive mechanisms.

References

[1] Peter W. Battaglia, Jessica B. Hamrick, and Joshua B. Tenenbaum. Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, page 18327–18332, Nov 2013. doi: 10.1073/pnas.1306572110. URL `http://dx.doi.org/10.1073/pnas.1306572110`. 1, 2

[2] Oliver Groth, Fabian B. Fuchs, Ingmar Posner, and Andrea Vedaldi. *ShapeStacks: Learning Vision-Based Physical Intuition for Generalised Object Stacking*, page 724–739. Jan

2018. doi: 10.1007/978-3-030-01246-5_43. URL http://dx.doi.org/10.1007/978-3-030-01246-5_43. 3

[3] James R. Kubricht, Keith J. Holyoak, and Hongjing Lu. Intuitive physics: Current research and controversies. *Trends in Cognitive Sciences*, page 749–759, Oct 2017. doi: 10.1016/j.tics.2017.06.002. URL http://dx.doi.org/10.1016/j.tics.2017.06.002. 1, 2