Evidence for an Intuitive Physics Engine in the Human Brain

Sarah E. Schwettmann (schwett@mit.edu)
Department of Brain and Cognitive Sciences, MIT

Joshua Tenenbaum (jbt@mit.edu)
Department of Brain and Cognitive Sciences, MIT

Nancy Kanwisher (ngk@mit.edu)
Department of Brain and Cognitive Sciences, MIT
Cambridge MA 02139

Abstract

Humans demonstrate a remarkable ability to infer physical properties of objects and predict physical events in dynamic scenes. These abilities have been modeled as probabilistic simulations of a mental physics engine akin to 3D physics engines used in computer simulations and video games (Battaglia, Hamrick & Tenenbaum 2013; Sanborn, Mansinghka & Griffiths 2013), but it is unknown if and how such a physics engine is implemented in the brain. Does the brain represent quantities corresponding to the key latent variables of physical objects that contribute to their dynamics? To find out, we used multivariate pattern classification analyses of fMRI data from subjects viewing videos of dynamic objects. The masses of depicted objects could be decoded from parietal and frontal brain regions previously implicated in intuitive physics (Fischer et al., 2016). Crucially, this decoding was invariant to the scenario revealing the object's mass, as well as the the material, friction, and amount of motion of the object. These regions may support a generalized engine for intuitive physics where this invariant representation of mass serves as a key variable.

Keywords: **intuitive physics; physical reasoning; vision**

Introduction

Engaging with the world requires a model of its physical structure and dynamics – how objects rest on and support each other, how much force would be required to move them, and how they behave when they fall, roll, or collide. This intuitive understanding of physics develops early in childhood and in a consistent order; by 3 to 4 months of age, infants understand that the world is composed of bounded, continuous objects (Kestenbaum & Spelke, 1987; Spelke, Kestenbaum, Simons & Wein, 1995), by 5 months, they can differentiate liquids and solids using expectations about the behavior of nonsolid objects (Hespos, Ferry & Rips 2009; Hespos, Ferry, Anderson, Hollenbeck, Rips 2016), and by 11 months they can infer an object's weight based on its compression of a soft material (Hauf, Paulus, & Baillargeon, 2012). By adulthood, humans are capable of making sophisticated physical predictions in many different tasks requiring implicit physical reasoning or simulation (Battaglia, Hamrick & Tenenbaum, 2013). This body of behavioral evidence suggests that the brain contains detailed knowledge of the physical attributes of objects and the laws of physical interactions between them. We consider these concepts and the laws relating them to constitute an intuitive theory of physics that enables us to successfully interact with objects, make predictions about physical events, and perform novel physical tasks.

Recent computational efforts have explained human physical reasoning via an intuitive physics model that is quantitative, approximate, compositional, and probabilistic. Battaglia et al. propose a computational architecture shared by many physics engines, with two core parts: an object-based representation of a 3D scene (which encodes static variables such an object's size and mass), and a model of physical forces that govern the scene's dynamics. This type of simulation-based model can make robust inferences about configurations of many rigid objects subject to gravity and friction, with varying numbers, sizes, and masses (Battaglia, Hamrick & Tenenbaum, 2013). If such a physics model were implemented in the brain, we would expect underlying brain regions to represent relevant physical dimensions as concepts that generalize across scenarios. Here we applied pattern classification methods to fMRI data obtained from subjects viewing videos of dynamic objects, to test for invariant representations of mass in brain regions previously implicated in intuitive physical inference (dorsal premotor cortex/supplementary motor area, and bilateral parietal regions; Fischer et al, 2016).

Experiments

Participants were scanned with fMRI while performing physical inference, prediction, and orthogonal tasks on visually-presented stimuli. "Localizer" scans enabled us to identify key nodes of the candidate physics network in each subject individually, following Fischer et al. (2016). We then conducted three experiments to (i) test whether we can decode relevant physical variables from candidate physics fROIs with various degrees of invariance and (ii) test the automaticity of these representations.

Design

Each scanning session included two runs of a localizer task from Fischer et al. (2016) which used a univariate contrast

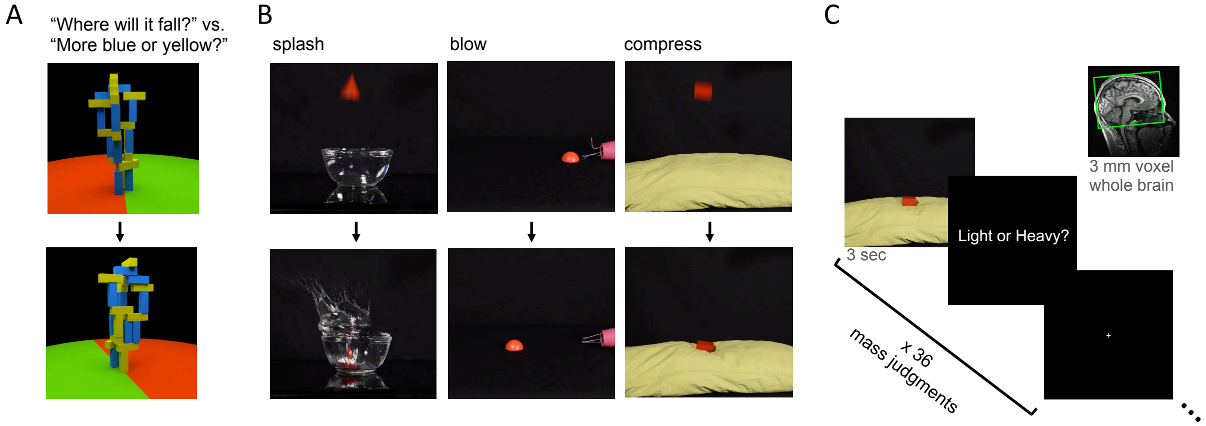
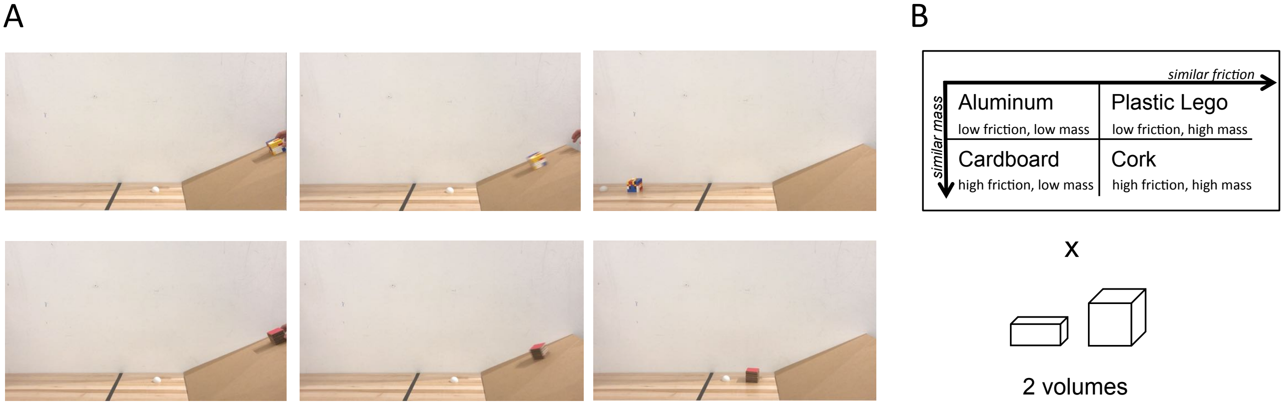


Figure 1: (A) Toppling tower task. Screenshots show an example tower from two different viewpoints during the 360° pan. (B) Example mass inference stimuli used in Experiments 1 and 2. Stills from splashing and compression movies depicting heavy objects; stills from the blowing scenario depict a light object. (C) Schematic of event-related scanning paradigm in Experiment 1.

of physical versus color judgments on toppling tower stimuli to isolate regions involved in physical reasoning. The stimuli were 6s movies created in Blender (Blender Online Community 2015) depicting towers of yellow, blue, and white blocks (Figure 1A) that were unstable and would tumble if gravity were to take effect.

In Experiment 1, 6 subjects viewed 3s movies of real objects interacting in various physical scenarios: splashing into a container of water, being blown across a flat surface by a hairdryer, and falling onto the soft surface of a pillow (Figure 1B). Three rigid 3D shapes of equal volume were used (a rectangular prism, a cone, and a half-sphere), and movies were filmed for two different colors and two different masses (45g, 90g) of each shape (36 total movies). Visual cues from the scene could be used to infer the mass of each object. After each movie, subjects responded to a text prompt ("Light or Heavy?") with a button press indicating their inferred mass.

Experiment 2 asked whether it was possible to decode mass from multivoxel activity in candidate physics fROIs during a color judgment task orthogonal to the physics task. Six new subjects viewed the same stimuli used in Experiment 1, and after each video were prompted with 1s of text to respond whether the object was "Light or Heavy?" or "Red or Orange?" Physics and color tasks were completed in blocks of 6 videos each.



Aluminum low friction, low mass	Plastic Lego low friction, high mass
Cardboard high friction, low mass	Cork high friction, high mass

Figure 2: (A) Screenshots of example videos from Experiment 3. Top: lego cube, bottom: cardboard cube. (B) Illustration of invariance dimensions used for mass decoding.

In Experiment 3, we asked whether mass could be decoded from candidate physics brain regions during a physical prediction task that requires mass knowledge but never explicitly interrogates it. We created 48 real-world movies. Each 6s video shows an object (made of aluminum, cardboard, lego, or cork) sliding down a ramp and colliding with a puck (half-ping-pong ball), whose initial location is consistent between videos (Figure 2A). Subjects answered, as immediately as they could, whether they predict the sliding object will launch the puck across a black line, which can lie in 3 different locations. The mass of the object and its coefficient of friction determine how far it will launch the puck. Each of the four different materials was used to make two objects, a 2.5" cube and a 2.5"x 2.5"x1.25" object with half of the volume of the cube and the same surface area in contact with the ramp (Figure 2B).

Importantly, these stimuli were designed in a way that orthogonalizes mass, friction, and motion in the videos (Figure 2B), allowing us to test whether it is possible to decode a generalized representation of mass invariant to friction and motion. Materials were chosen with densities such that same-volume objects made out of aluminum and cardboard have the same mass (30g, 15g), and same-volume objects made from lego and cork have the same mass (90g, 45g), while pairs along the other invariance dimension (aluminum and legos, cardboard and cork) share similar coefficients of friction with the ramp.

Multivariate Decoding Analyses

To test the representational content of multivoxel activity from candidate physics regions, decoding analyses (Naselaris et al., 2011; Haxby, Connolly & Guntupalli, 2014) were run on multivoxel activity pooled across these fROIs. An SVM was used for classification, restricted to linearly decodable signal under the assumption that a linear kernel implements a plausible readout mechanism for downstream neurons (Shamir & Sompolinsky 2006; DiCarlo & Cox 2007). In each of 3 experiments we tested the invariance of physical representations by testing the classifier on data from conditions that differed from those in the data used for training along a key dimension. For example, to decode mass in Experiment 1, an SVM was trained on beta values classified as corresponding to either "heavy" or "light" conditions, collapsing across shape and color. We used two of the three scenario types (splash, blow, compress; see Figure 2B) to train the classifier and tested on the third, left-out scenario, forcing the classifier to generalize across physical scenarios and iterating over left-out conditions to obtain a mean classification accuracy for each subject.

Results

In Experiment 1, situation-invariant mass decoding in the candidate physics system had group mean accuracy 0.64, which was significantly above chance ($p<0.05$), and was found numerically in 6 out of 6 subjects. Critically, this representation of object mass does not depend on whether the object is splashing into water, being blown by a hair dryer, or being dropped onto a pillow: to obtain a decoding accuracy greater than 50%, the classifier must generalize a representation learned from two scenarios to a third scenario left out of training. Mean classification accuracies as well as classification accuracies for each left out scenario were greater than 50% in all subjects. Further, mass representations are not confounded with shape or color, as colors and shapes were represented in equal proportions for both masses in the training and testing data. While this result mirrors the situation-invariant representations expected in a physics engine, an alternative hypothesis is that we may be decoding a prepared response to the explicit mass task ("Light or Heavy?" which is constant across scenarios). To test this hypothesis, as well as the automaticity of the mass representation, in Experiment 2 we used a design that interleaves blocks of the physics task and a color task on the same stimuli. This design enabled us to ask whether a situation-invariant mass representation can also be decoded from multivoxel activity during blocks where subjects perform the orthogonal color task where mass was not relevant.

In 6 new subjects, we replicated the findings of Experiment 1: mean decoding accuracy 0.63 was significantly above chance ($p<0.05$), and present numerically in each subject individually during the mass task. More importantly, mass decoding was also significantly above chance (mean = 0.61, $p<0.05$), and present numerically in each subject individually, during the color task. This result shows that mass is represented even when the task does not require it, and further that the decoding of mass we observe cannot be explained as an abstract response code. Further evidence against the idea that the mass representations reflect response codes comes from the fact that color decoding from the same voxel activity during the color task was at chance in all subjects. Thus the candidate physics engine does not represent all task-relevant dimensions and may be more specific to physical variables.

In Experiment 3 we asked whether this representation can be decoded during a physical prediction task that requires an understanding of mass but does not explicitly suggest subjects attend to it, and tested the invariance of this representation to friction and motion. Experiment 3 replicated once again our finding that mass can be decoded from candidate physics regions (mean accuracy of 0.60 was significant, $p<0.05$, and numerically present in each of 13 out of 14 participants individually). Further, this experiment demonstrates an important new invariance of these mass representations beyond those already found in Experiments 1 and 2: the mass decoding in Experiment 3 required generalization across the friction and material of the object shown (lego to cork for heavy, and cardboard to aluminum for light).

Discussion

This work tested for a physics engine in the human brain by asking whether brain regions previously implicated in intuitive physical reasoning (Fischer et al., 2016) contain information about the physical properties of objects, specifically mass. Indeed, we showed using fMRI decoding methods that the candidate brain regions for physical inference contain information about mass in 25/26 subjects tested. Importantly, this mass information is present even when mass is irrelevant to the task (Experiment 2) or when mass is relevant but the participant is not asked to report it explicitly (Experiment 3). Further, mass information is invariant to the dynamic scenario in which the mass of the object is revealed (Experiments 1 and 2), as well as to the coefficient of friction of the object, the material it is made of, and the overall amount of motion in the scene (Experiment 3). Taken together these results show that the brain regions previously implicated in intuitive physical reasoning represent mass in an invariant manner that would be expected for an intuitive physics engine.

Importantly, the brain regions implicated in intuitive physical reasoning resemble those previously implicated in action planning, highlighting the tight link between these two functions. Indeed, recent work in robotics (Todorov 2017) combined optimization with a physics model to design a highly efficient goal-directed action planning system. And fMRI studies in which humans perform actions (Johansson & Flanagan 2009; Loh et al. 2010; Chouinard, Leonard & Paus 2005; van Neunen, et al. 2012) reveal representations of object mass in brain regions where we report mass decoding in inference tasks, further strengthening the link between physical inference and action. A model-based account of physics in the brain could support both physical inference and action planning in the same underlying brain regions, which may serve as the seat of a neural physics engine. The suggestion of overlapping activation could be strengthened in future work testing action planning and physical inference in the same subjects.

Acknowledgements

This work was supported by NIH Grants DP1HD091947 (to N.K.) and National Science Foundation Science and Technology Center for Brains, Minds, and Machines Grant CCF-1231216.

References

Battaglia, P. W., Hamrick, J. B., Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences, 110*(45), 18327–18332.

Blender Online Community. (2015). Blender - a 3d modelling and rendering package [Computer software manual]. Blender Institute, Amsterdam. Retrieved from http://www.blender.org

Chouinard, P.A., Leonard, G., & Paus, T. (2005). Role of the primary motor and dorsal premotor cortices in the anticipation of forces during object lifting. *Journal of Neuroscience 25*, 2277–2284.

DiCarlo, J.J., & D.D. Cox. (2007). Untangling invariant object recognition. *Trends in Cognitive Science 11*(8), 333-41.

Fischer, J., Mikhael, J.G., Tenenbaum, J.B., & Kanwisher, N. (2016). Functional neuroanatomy of intuitive physical inference. *Proceedings of the National Academy of Sciences*, *113*(34), E5072–E5081.

Hauf, P., Paulus, M. & Baillargeon, R. (2012). Infants Use Compression Information to Infer Objects' Weights. *Child Development, 83*, 1978–1995.

Haxby, J.V., Connolly, A.C., & Guntupalli, J.S. (2014). Decoding neural representational spaces using multivariate pattern analysis. *Annual review of Neuroscience*, *37*, 435-456.

Hespos S.J., Ferry A.L., & Rips L.J. (2009). Five-month-old infants have different expectations for solids and liquids. *Psychological Science, 20*(5), 603–611.

Hespos S.J., Ferry A.L., Anderson E.M., Hollenbeck E.N., & Rips L.J. (2016) Five- month-old infants have general knowledge of how nonsolid substances behave and interact. *Psychological Science, 27*(2), 244–256.

Johansson, R.S., Flanagan, J.R. (2009). Coding and use of tactile signals from the fingertips in object manipulation tasks. *Nature Reviews Neuroscience*, *10*, 345–359.

Kestenbaum R., Termine N., & Spelke E.S. (1987). Perception of objects and object boundaries by 3-month-old infants. *British Journal of Developmental Psychology*, *5*(4), 367–383.

Loh, M.N., Kirsch, L., Rothwell, J.C., Lemon, R.N., & Davare, M. (2010). Information about the weight of grasped objects from vision and internal models interacts within the primary motor cortex. *Journal of Neuroscience, 30*, 6984–6990.

Naselaris, T., K.N. Kay, S. Nishimoto, & Gallant, J.L. (2011). Encoding and decoding in fMRI. *Neuroimage, 56*(2), 400-10.

Sanborn, A.N., Mansinghka, V.K., & Griffiths, T.L. (2013). Reconciling intuitive physics and newtonian mechanics for colliding objects. *Psychological review*, *120*(2), 411.

Shamir, M., & Sompolinsky, H. (2006). Implications of neuronal diversity on population coding. *Neural Computation, 18*(8), 1951-86.

Spelke E.S., Kestenbaum R., Simons D.J., & Wein D. (1995). Spatiotemporal continuity, smoothness of motion and object identity in infancy. *British Journal of Developmental Psychology, 13*(2), 113–142.

Todorov, E. (2018). *Goal Directed Dynamics.* Paper presented at International Conference on Robotics and Automation, Brisbane, Australia.

van Neunen, B.F., Kuhtz-Buschbeck, J., Schulz, C., Bloem, B.R., & Siebner, H.R. (2012). Weight-specific anticipatory coding of grip force in human dorsal premotor cortex. *Journal of Neuroscience, 32*, 5272–528.

Wu, J., Yildirim, I., Lim, J. J., Freeman, B., & Tenenbaum, J. (2015). Galileo: Perceiving physical object properties by integrating a physics engine with deep learning. In *Advances in neural information processing systems* (pp. 127– 135).