2016

Neural representations of physics concepts

Robert A. Mason, *Carnegie Mellon University*
Marcel Adam Just, *Carnegie Mellon University*

Research Article

Psychological Science
2016, Vol. 27(6) 904–913

DOI: 10.1177/0956797616641941
pss.sagepub.com

Neural Representations of Physics Concepts

Robert A. Mason and Marcel Adam Just
Center for Cognitive Brain Imaging, Psychology Department, Carnegie Mellon University

Abstract

We used functional MRI (fMRI) to assess neural representations of physics concepts (momentum, energy, etc.) in juniors, seniors, and graduate students majoring in physics or engineering. Our goal was to identify the underlying neural dimensions of these representations. Using factor analysis to reduce the number of dimensions of activation, we obtained four physics-related factors that were mapped to sets of voxels. The four factors were interpretable as causal motion visualization, periodicity, algebraic form, and energy flow. The individual concepts were identifiable from their fMRI signatures with a mean rank accuracy of .75 using a machine-learning (multivoxel) classifier. Furthermore, there was commonality in participants' neural representation of physics; a classifier trained on data from all but one participant identified the concepts in the left-out participant (mean accuracy = .71 across all nine participant samples). The findings indicate that abstract scientific concepts acquired in an educational setting evoke activation patterns that are identifiable and common, indicating that science education builds abstract knowledge using inherent, repurposed brain systems.

Keywords

neural representations, scientific concepts, fMRI, physics semantics

Received 7/20/15; Revision accepted 3/7/16

Considerable advances have been made in developing brain-based theories of semantic knowledge, such as knowledge of concrete objects or emotions. Brain-imaging research has uncovered sets of brain systems that collectively contain the neural representations of such concepts, including information about the way the human body interacts with them (in the case of objects) or their intensity (in the case of emotions; Just, Cherkassky, Aryal, & Mitchell, 2010; Kassam, Markey, Cherkassky, Loewenstein, & Just, 2013). What has not yet been investigated with this approach is the neural representation of specialized abstract knowledge acquired through academic study, such as science learning. The current article addresses this issue in the area of physics knowledge. In the current study, we investigated the patterns of brain activity, as measured by functional MRI (fMRI), in students majoring in physics or engineering while they thought about physics concepts. It was possible to identify sets of brain locations and dimensions of knowledge representation that underlie the concepts and to test the resulting model in terms of its ability to classify physics-evoked activation patterns to which it had not been previously exposed.

Although physics is one of the fundamental sciences to which many students are exposed, there is sparse research into the brain basis of physics knowledge (Petitto & Dunbar, 2009). The current research begins to fill that void as well as to investigate what it means to have acquired knowledge of physics. In a superficial sense, physics terms are just new concepts whose definitions have to be learned. However, physics concepts are different in kind from concrete nouns, action verbs, and even simple abstract concepts. They require new formulations of knowledge that go beyond membership in a known category and include conceptions of nonvisible aspects of the physical world, mathematical knowledge, and the combination of complex features into a whole.

Corresponding Author:
Robert A. Mason, Department of Psychology, Baker Hall, Carnegie Mellon University, Pittsburgh, PA 15213
E-mail: rmason@andrew.cmu.edu

These complex concepts, acquired through formalized training, are qualitatively different from other abstract concepts that have been successfully identified from brain imaging data, such as emotions (Kassam et al., 2013), generic abstract concepts (Wang, Baucom, & Shinkareva, 2013), and numerical quantities (Damarla & Just, 2013). It is possible that this recent record of success may have been due in part to the basic or primitive nature of the concepts studied.

Two properties set physics concepts apart: They refer to abstract properties of the physical world, and they are acquired through schooling. Of course, everyone has some naive knowledge of physics (e.g., "the bigger they are, the harder they fall" may be a naive definition of momentum, lacking the velocity component), but sciences develop new concepts that are eventually communicated during formal schooling. In general, schooling has been shown to build not only new knowledge but also new brain capabilities; for example, instruction in reading brings forth a brain-based word recognition system that has no innate biological support but becomes capable of rapid word recognition (Cohen et al., 2002; McCandliss, Cohen, & Dehaene, 2003).

The referents of physics concepts are sometimes tangible, but at some level, they are always abstract. Abstractness suggests some separation from perceptual and motor representations and the daily activities and thoughts that human beings have engaged in for millennia. Physics concerns itself with matter and energy, which in some senses are very concrete, but understanding their nature in scientific terms often involves concepts that are typically not directly observable. Yet people with knowledge of physics develop systematic conceptions or representations of such entities as velocity and momentum and, as we show, they also develop systematic neural representations that can be assessed in several ways that are unavailable to behavioral analysis. The aim of the current study was to uncover the neural structure of such entities.

The neural representations of 30 physics concepts were assessed in physics students using fMRI. The goals were (a) to determine whether the concepts have consistent neural signatures identifiable by a classifier, (b) to characterize the underlying neural dimensions of representation that compose the signatures, and (c) to assess the commonality of the representations across participants.

Method

Participants

Nine right-handed adults (3 women, 6 men; age range = 19–25 years) from the Carnegie Mellon community participated. This sample size was in the 9-to-12 range used in previous machine-learning research from our lab (e.g., Mitchell et al., 2008). All participants gave signed informed consent approved by the Carnegie Mellon University institutional review board. All participants were undergraduates who had completed at least two years of college (6 juniors or seniors) or graduate students (3) and were in physics or engineering. These students had taken physics courses beyond an introductory level and had numerous prior encounters with the physics concepts used as stimuli. Our goal was to characterize the neural representations of these concepts at this level of education; we acknowledge that the representations of professional physicists may be different (Maloney, O'Kuma, Hieggelke, & Van Heuvelen, 2001; McDermott, 1998; Stylos, Evangelaki, & Kotsis, 2008). All 9 participants contributed usable data (within movement parameters).

Experimental paradigm

The stimuli were 30 physics terms from several physics topic areas: mechanics, electricity, thermodynamics, energy, light, and sound. The full list consisted of the terms *acceleration, centripetal force, diffraction, direct current, displacement, electric charge, electric current, electric field, energy, entropy, force, frequency, gravity, heat transfer, inertia, kinetic energy, light, magnetic field, mass, momentum, potential energy, radio waves, refraction, sound waves, temperature, thermal energy, torque, velocity, voltage*, and *wavelength*. Each concept was presented six times (in six different random orders). Each word was presented for 3 s, during which the participant thought about the concept. This was followed by a 4-s rest period, during which the participant fixated on an "X" displayed in the center of the screen. There were seven additional presentations of a fixation "X," 17 s each, distributed across the session to provide a baseline measure.

Task

The participants' task was to actively think about the properties that they associated with the presented concepts. Although the task required no overt response, it was far from passive; it required a controlled iteration through the properties of the concept and was rather demanding. To promote the participants' consideration of a consistent set of properties or features across the six presentations of a term, we asked them to generate two or three properties for each term before the scanning session; for example, the properties for the term *velocity* might be "vector quantity," "movement related," and "directional." Each participant was free to choose any properties for a given item, and there was no attempt to impose consistency across participants in the choice of properties. Some of these participant interviews are excerpted in the Discussion section.

fMRI procedures

Functional images were acquired on a 3.0-T scanner (Verio; Siemens, Erlangen, Germany) at the Scientific Imaging and Brain Research Center of Carnegie Mellon University. We used a gradient-echo echo-planar imaging pulse sequence with a repetition time of 1,000 ms, echo time of 25 ms, and a flip angle of 60°. Twenty 5-mm slices, aligned along the anterior commissure-posterior commissure line, were imaged with a 1-mm interslice gap and a 32-channel head coil. The acquisition matrix was 64 × 64 with 3.125- × 3.125- × 5.0-mm in-plane resolution. Images were corrected for slice acquisition timing, motion, and linear trend, and they were normalized to the Montreal Neurological Institute template without changing voxel size (3.125 × 3.125 × 6 mm). The gray-matter voxels were assigned to anatomical areas using masks from Automated Anatomical Labeling software (Tzourio-Mazoyer et al., 2002).

The percentage signal change relative to the fixation condition was computed at each gray-matter voxel for each stimulus presentation. The main input measure for the subsequent analyses consisted of the mean activation level over the four brain images acquired within a 4-s window, offset 4 s from the stimulus onset (to account for the delay in hemodynamic response). The percentage-signal-change data of the voxels in the mean image for each word were converted to z scores.

Selecting voxels with stable activation patterns

A voxel's *activation profile* refers to the vector of its 30 responses (activation levels) to the 30 words during that presentation. The first criterion for voxel selection was a stable activation profile (i.e., stable tuning curves) over the 30 words across the six presentations of the set of words. A voxel's stability was computed as the mean pairwise correlation between its 30-word activation profiles across all pairs of the presentations that served as training input for a given classification model. A stable voxel is thus one that responds similarly to the 30-word stimulus set each time the set is presented.

Factor analysis methods

The second criterion for voxel selection was the presence of an association with one of the factors emerging from a factor analysis of the activation data. To reduce the dimensionality of the neural activity associated with the 30 different stimulus items to a modest number of components, we applied a multilevel exploratory factor analysis procedure. We implemented a principal factor analytic algorithm, including varimax rotation, in MATLAB (Version 6.5; The MathWorks, Natick, MA), equivalent to SAS (Version 9.2; Cary, NC). The aim was not only to identify some of the main factors underlying the activation patterns but also to determine the brain locations of the stable voxels that were associated with each of the factors, and we expected that multiple locations would be associated with each factor. This factor analytic procedure is described in detail elsewhere (Just, Cherkassky, Buchweitz, Keller, & Mitchell, 2014; Just et al., 2010). We describe some of the parameters that are specific to the current study.

At the first level, 10 separate factor analyses were performed on the data for each participant, one analysis per region: left and right frontal lobes, left and right parietal lobes, left and right temporal lobes (minus fusiform gyrus), left and right occipital lobes, and left and right fusiform gyrus (as defined in the Automated Anatomical Labeling atlas; Tzourio-Mazoyer et al., 2002). The input data were the matrix of intercorrelations among the activation profiles of the 120 most stable voxels in each region.[1] The goal of each of these first-level factor analyses was to reduce the data from the activation profiles across concepts from many stable voxels in each region to a few factors that characterized the profiles of most of the stable voxels in each participant. This exploratory factor analysis indicated a reasonable estimate for the number of factors to be expected, using the Kaiser criterion (i.e., the minimal number of factors with an eigenvalue of 1).

A second-level, higher-order factor analysis was then run to identify factors that were common across regions and participants. (The search for commonality of factors across regions was motivated by the assumption that a factor would be composed of a large-scale cortical network with representation in multiple and disparate brain regions.) The input to the second-level analysis consisted of the 10 dominant first-level factors obtained from the 3 participants classified most accurately. Using the criterion that at least 5% of the variance had to be explained by a factor, we reduced this set of factors from 10 to 7. Using a stepwise addition of factors to the classification model until the accuracy stopped improving, we were able to remove 2 more factors, resulting in a final set of 5 factors. These factors were used in analyzing the data of the other 6 participants. (For the analysis of the data of the 3 participants classified most accurately, the factors were derived from a factor analysis performed on the other two participants classified most accurately.)

Each factor was then mapped to several clusters of voxels that had high factor loadings and similar activation profiles (i.e., tuning curves over the 30 terms). Clusters were defined as six or more contiguous voxels. The number of clusters per factor ranged from 2 to 6 (with a total of 22 clusters for the five factors). These clusters were converted into spherical volumes with a radius of 10 mm

centered at an estimated center of mass of each cluster (centroids are listed in Table S1 in the Supplemental Material available online) for the 20 non-word-form clusters (2 clusters centered in the left and right occipital cortex corresponded to a word-length factor and are not reported in the table); a voxel associated with more than one sphere was assigned to the factor that accounted for a higher percentage of the variance.

Machine-learning analyses

Gaussian naive Bayes classifiers with factor-based features were used to identify the 30 physics concepts (for an overview of Gaussian naive Bayes classifier cross-validation as applied to fMRI data, see Just et al., 2010). The classifiers were trained using stable voxels from only a subset of the data (the training set) and were then tested on the remaining data (the test set) using a cross-validation procedure. For the within-participants classification, the classifier was tested on the mean of the two left-out presentations. This procedure was reiterated for all 15 possible combinations (folds) of training on a set of four presentations and testing on the average of the two left-out presentations. The between-participants classification always left out the data of the to-be-classified participant and trained the classifier on the remaining participants' data. In the latter analysis, each participant's data (from the 22 spheres that emerged from the two-stage factor analysis) were averaged over the six presentations. Then the 120 voxels with the most similar activation profiles (assessed by correlation) across the 8 participants in the training set were selected as features for the classifier. This analysis assumed not only that the activation patterns across concepts were common across participants but also that the sphere locations of the key voxels instantiating that pattern were common.

Overview of the factor-labeling process

The interpretation of the factors that emerge in a factor analysis remains a subjective process, but in our case, the resulting interpretation was used to construct a classifier and then quantitatively evaluate its accuracy. The factor analysis provided two types of information that were helpful in interpreting the factors. One type of useful information was the rank ordering of the 30 concepts by their scores on the factor. A clear example of this type of information occurred when the rank order of the factor scores for the words in one of the factors matched the rank order based on the number of letters that the word or phrase contained, which indicated that the factor pertained to the encoding of the word form. In the ranked list, the terms at the two extremes of the ordering were particularly informative, sometimes immediately indicating what they have in common (e.g., a relation to thermal energy). The set of terms at the extremes might also have been correlated with the categorizations of the 30 items, or the extreme terms might have a shared superordinate concept. The second type of information from the factor analysis was the location of the voxel clusters that had high factor loadings for a particular factor. Although this evidence source for interpreting the factors included a reverse inference concerning the functions of various brain regions, these posited regions were subsequently used as part of a classifier model that was quantitatively evaluated. The classifier tested how accurately a concept could be identified from its neural signature, assuming that the signature included elements (voxels) from the regions posited to underlie the neural representation. The high classification accuracies indicated that the inductive assumptions were good ones.

Categorization of the physics concepts

After the scanning session and without prior notice, participants were asked to provide an open-ended categorization of the 30 terms into four to six categories and to label each category. The goal was to determine how the categorization of the 30 terms might be related to factor scores of these terms derived from the factor analysis. The responses were consolidated into six categories ("waves," "electricity," "thermodynamics," "mechanics," "energy," and "basic properties or other"); the precise labels varied across participants but were highly similar (e.g., "light & waves" for "waves"). Each category was represented as a vector of length 30; binary values indicated whether a word was modally placed in a category. On average, participants matched in their category assignments on 78% of the items (omitting 1 participant who produced qualitatively different categories). Very similar results (75% item agreement) were obtained in a norming sample of 5 participants who did not take part in the fMRI study (2 additional people with a physics background and 3 undergraduate psychology majors with little knowledge of physics).

Results

Overview

It was possible to identify which physics concept the participants were thinking about on the basis of the brain activation signature. The Gaussian naive Bayes classifiers were theory driven in that their features were voxels from regions derived from a factor analysis of the brain-activation data. This factor analysis was applied only to the conjoint data of the 3 participants with the most accurately identifiable concepts (using an atheoretical classifier). The factor analysis yielded five factors, each of

which could be traced back to a small set of voxel clusters, for a total of 22 clusters. These clusters were transformed into spheres from which voxels were drawn to serve as the classifier features. The four physics-related factors were interpretable as corresponding to causal motion visualization, periodicity, algebraic form (i.e., an equation or expression involving that term), and energy flow; a fifth factor was associated with word length.

Within-participants identification of physics concepts

The within-participants Gaussian naive Bayes classifier was trained on a subset of the data from a given participant and then tested on an independent subset of that same participant's data, drawing voxels as features from the factor-derived spheres. All of the within-participants classifications were based on spheres derived from the 3 best-classified participants (or from the 2 best, excluding a participant if he or she was 1 of that set of 3; see the Factor Analysis Methods section for a more detailed description). The 30 physics concepts were classified with a mean rank accuracy (hereafter, simply accuracy) of .74 (the critical value for above-chance performance at $p < .01$ was .55, obtained by permutation testing). All 9 participants were classified significantly above chance (range = .64–.89). Three participants had particularly high accuracies; their mean accuracy was above .84 (.89, .83, and .81). These were the participants from whose factor analyses the voxel locations were derived for classifying the 6 other participants. The mean accuracy for these 6 participants was .69 (range = .64–.74). (When selection of voxels was made randomly, classification accuracy was no different from chance, which indicates that the factor analysis captured physics-relevant regions of cortex.)

Between-participants identification of physics concepts

The between-participants Gaussian naive Bayes classifier examined the similarity of the concept representations across participants; the classifier was trained on data (voxel-activation levels) from all but one participant and then tested on the data of the left-out participant, using voxel locations in the factor-derived spheres and repeated using each participant as the test participant. The features were the 120 voxel locations that had a consistent profile (high pairwise correlations between participant means) across the 8 participants in the training set. The mean between-participant classification accuracy for the 9 participants was .71 (range = .54–.81). These reliable results (8 participants with accuracy greater than chance at $p < .01$; 1 participant was significantly different from chance at $p < .05$, critical value = .53) indicate that the factor-defined spheres captured a commonality across participants of the neural representations of these physics concepts.

Neural dimensions of physics-concepts representations

Four complementary approaches contributed to interpreting the factors: (a) the ordering of the 30 physics terms by their factor scores on a given factor, with particular attention to the terms at the extremes of the dimension; (b) the correlation between the factor scores and category groupings of the words by the participants (a vector of ones and zeros, where 1 = category membership and 0 = out of category); (c) applying the latent semantic analysis (LSA) nearest-neighbor function to the participants' property descriptions (Landauer & Dumais, 1997); and (d) the locations of the factor-defined spheres and knowledge of which experimental manipulations have previously produced activation in those locations.

Causal-motion-visualization factor. Concepts with high factor scores for this factor, such as *gravity* and *potential energy*, could be interpreted as having a role in the visualization of motion and the causality of forces. These concepts were associated with voxel cluster locations in the left occipital-temporal-parietal junction, left intraparietal sulcus and left middle frontal gyrus. The physics terms with extreme scores on this factor (e.g., *centripetal force, torque, displacement*, and *momentum*) tended to be sorted into the category of "mechanics." The factor-organized list of terms with high factor scores, correlated categories, and cluster locations are presented in Supplemental Table 1. A rendering of the four highly predictive sets of physics factor spheres is shown in Figure 1.

Periodicity factor. A second factor was interpretable as relating to periodicity. *Wavelength, radio waves, frequency, diffraction*, and *sound waves* had high factor scores on this factor. The scores on this factor were correlated with membership in the "waves" category ($r = .68$) and the "energy" category ($r = -.55$). Brain locations corresponding to this factor included bilateral superior parietal gyrus, left postcentral sulcus, left posterior superior frontal gyrus, and bilateral inferior temporal gyrus.

Algebraic-equation-representation factor. Concepts with high factor scores for this factor included *velocity, acceleration*, and *heat transfer*, all of which are particularly strongly associated with familiar equations. This set of words was correlated with the "mechanics" category

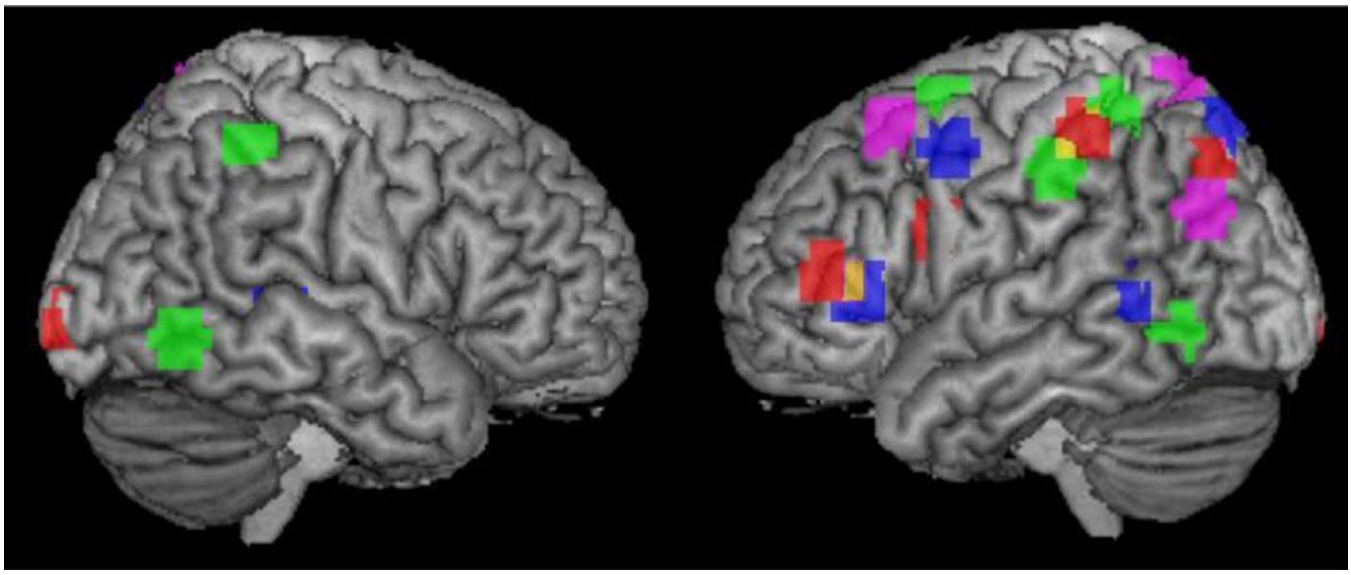

Fig. 1. Rendering of the four physics factors. Shown are right and left lateral views of the brain. The four sets of factor-related voxel clusters that emerged from the factor analysis of the activation evoked by the physics terms are colored as follows: pink indicates causal-motion visualization, green indicates periodicity, red indicates algebraic or equation representation, blue indicates energy flow, and yellow indicates overlap of adjacent clusters. The word-length factor is omitted here.

($r = -.43$) in the participants' postexperimental categorization task. The voxel clusters in this set were located in the precuneus, the left parietal lobe, the left inferior frontal gyrus, and the left occipital gyrus. Parietal regions typically activate in calculation tasks (Dehaene, Molko, Cohen, & Wilson, 2004; Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999). The participants' reports of the properties that they associated with each of the concepts provide converging evidence for the interpretation of this factor as involving an algebraic formulation. For example, for the concept of *velocity*, several participants listed the algebraic formulation "$\Delta x/\Delta t$" as a property. For the concept of *heat transfer*, several participants listed the Boltzmann constant (k). For *electric current*, several participants listed "Ohm's law" as a property.

Energy-flow factor. A fourth factor can be associated with energy flow. *Electric field, light, direct current, sound waves*, and *heat transfer* were among the concepts with high scores on this factor. The factor scores were correlated with membership in the electricity ($r = .38$) and thermodynamics ($r = -.37$) categories of physics terms. This factor was associated with voxel clusters in left intraparietal sulcus, left precentral sulcus, left posterior middle temporal gyrus, and inferior frontal gyrus. The participants' responses regarding the properties they associated with each of the concepts contributed to the interpretation of this factor as involving an energy flow. For example, for the concept of *direct current*, several participants listed "flow" or "flow of electrons" as a property. For *heat transfer*, participants listed "radiation" as a property. For *electric field*, participants listed properties such as "radial from a point charge." For this factor, the Suggested Upper Merged Ontology (Niles & Pease, 2001) was consulted to group the words with high factor scores into a common ontology. The higher level of abstraction that subsumes these concepts is the sense of energy flow in which energy radiates outward, or a "radiating of energy."

Word-length factor. A fifth factor corresponded to the alphabetic length of the term used to denote each concept (the factor scores of the concepts on this factor were highly correlated with word length, $r = .88$). The location of this nonphysics factor was almost exclusively in the occipital lobe, with small extensions into the inferior temporal-occipital junction and the intraparietal sulcus near the occipital lobe. This factor simply reflects the neural encoding of the visual word form of the concept.

The factor labels were consistent with a nearest-neighbor LSA of the properties of the concepts. The input to the LSA of each factor was a vector of the properties reported by the participants as being associated with the concepts that had high factor scores for that factor, and the output of the LSA was a list of semantic associates from the LSA space that were most related to the input. For example, for the algebraic-equation-representation factor, the word *proportional* was most related to the input properties, with a .71 LSA similarity score in the latent semantic analysis (on a scale from 0 to 1). Several other algebra-, measurement-, or equation-related words were among the top 15 LSA semantic associates (*arithmetic, joules, sec, constant, inversely*, and *magnitude*)

For the energy-flow factor, *electromagnetic* was the most related semantic associate, at .82; other energy-flow-related semantic associates included *hertz, sine, frequency, lambda, wave, amplitude, alternating*, and *radiate*. For the periodicity factor, the similar semantic associates included *lambda, vibrational, hertz, wavelength, frequency, sec, motion, amplitude*, and *crests*. For the causal-motion-visualization factor, the related semantic associates included *motion, acceleration, proportional, velocity*, and *exerting*. This complementary analysis of the participant-reported properties indicates that the proposed labeling of the factor analysis captured the participants' intuitions.

Neural representations associated with the extracted physics factors were present to a large degree in all of the participants. To demonstrate the high level of consistency in the nature of each factor across participants, we assessed the similarity between the factor scores emanating from the factor analysis of the pooled data of the 3 participants classified most accurately for the 5 highest-ranked and 5 lowest-ranked concepts and the factor scores of these 10 concepts in individual participants.

The mean correlation was .57. This conservative analysis excluded the 3 participants from whose data the factors were extracted (when computed for all participants, the mean correlation was higher, $r = .65$). These correlations also excluded the word-length factor (when the word-length factor was included, the mean correlation was higher, $r = .6$). The mean correlations between individual participants and the group of the 3 best participants for individual factors ranged from a high of .86 for 1 participant to a low of .35 for another participant, both for the periodicity factor. Thus the outcome of the factor analysis was very consistent across participants.

Discussion

Overview

The findings provide a new neurally based view of how physics concepts are represented in terms of brain organization. The activation patterns indicate a set of neural factors that underlie the representation of physics concepts and enable accurate classification of these concepts. The concepts are abstract and learned only through formal education, and yet interpretable factors associated with sets of regions that code various fundamental facets of these physics concepts do emerge. A classification model based on this postulated representation was successful in identifying the concepts from their neural signatures, which indicates that the factor analysis accurately extracted the underlying dimensions that organize the activation. The resulting classification accuracy using those factors stands on its own, independently of our interpretations of the factors.

The neural commonality of participants' representations of physics concepts reinforces our interpretation of the findings. We suggest that there is a common path from the basic capabilities of the human brain to the abstract physics concepts, developed only in the past few centuries and currently taught through formal schooling.

Repurposing basic brain capabilities to represent abstract physics concepts

Learning culturally developed knowledge may rely on repurposing neural structures that were originally evolved for other or general purposes (Dehaene & Cohen, 2007). Each of the factors that emerged in our analysis can be viewed from this perspective. The voxel clusters associated with each of the factors include executive regions (frontal), spatial regions (parietal), and several LH language areas implicating linguistic processing. Although the interpretations are speculative, the good performance of the theory-based classifier lends additional credence.

Causal-motion-visualization factor. The concepts associated with this factor entail motion that can be visualized and causality of the motion that can be conceptualized. Physics terms from mechanics (*centripetal force, displacement*) had high scores on this factor. Mechanics is a branch of physics concerned with motion and forces on objects. Students probably attempt to explain phenomena of nature in terms of "*X* causes *Y*," so it is not surprising that causality of forces emerges as one of the factors underlying the representation of scientific concepts in the minds of students. For example, *torque* is the force that causes an object to rotate around an axis. *Gravity* is a force that causes two bodies to be attracted.

Several of the regions associated with this factor (left intraparietal sulcus, left middle frontal gyrus) have been shown to play a role in attributing causality when viewing objects that collided (Fugelsang, Roser, Corballis, Gazzaniga, & Dunbar, 2005; Han, Mao, Qin, Friederici, & Ge, 2011). The parahippocampus was activated when participants had to link a causal theory to observed data (Fugelsang & Dunbar, 2005). Yet another region, the occipital-temporal-parietal junction, was activated during the visualization of movement of objects and actions in space (Jahn, Wendt, Lotze, Papenmeier, & Huff, 2012), suggesting that this factor may be related to causes of motion in particular. Of course, causality is involved in many other types of physics concepts (e.g., *heat transfer, force*), which are apparently less likely to evoke visualization of a causally understood event. In general, understanding the systematicity of the physical world entails imputing causal relations between concepts, and we propose that a factor corresponding to a visualizable and

causally understood motion constitutes one of the dimensions of representation of physics concepts.

Periodicity factor. Many of the concepts with high scores on this factor were associated with periodicity (*wavelength, frequency*). Periodicity is likely to be psychologically salient because of human sensitivity to periodic events in nature, such as biorhythms, lunar cycles, and ocean waves. The neural processing of periodicity has been studied in the context of music perception. Music and dance are typically associated with repeatable rhythms. Clusters in the dorsal premotor cortex associated with this factor activated when people tapped their fingers to rhythms (J. L. Chen, Zatorre, & Penhune, 2006). Interfacing sensory cues (the auditory stimulus rhythm) with temporally organized movement (the rate of tapping) builds on neural structures sensitive to periodicity that emerge in this factor.

This factor's cluster centroids also included somatosensory and bilateral parietal regions that have been linked to motor imagery and simulation (H. Chen, Yang, Liao, Gong, & Shen, 2009). Thus the periodicity factor may involve a neural representation of temporally regular events. For example, experienced dancers activated a similar left postcentral-parietal region (Montreal Neurological Institute coordinates: $x = -57$, $y = -27$, $z = 36$) when watching rehearsed movements (Cross, Hamilton, & Grafton, 2006). When dancers watched any dance movements, the left intraparietal sulcus activated. These physics students may internally simulate motions and sensations related to periodic physics concepts. Thinking about periodic physics concepts could evoke an embodied sense of experiencing the corresponding real-world phenomena.

Algebraic-equation-representation factor. The physics concepts that had high factor scores and algebraic equation associations included *velocity* and *acceleration*. Not surprisingly, the factor scores were correlated with the participant-generated members of the "mechanics" category (presumed to have strong associations with equations). Although many of the physics terms appear in a well-known equation, some terms have a stronger relationship to their algebraic expression than others. For example, an equation involving *velocity* comes to mind readily whereas an equation involving *diffraction* is less familiar. Other concepts with weaker equation associations (*magnetic field*) have lower factor scores.

This factor was associated with brain locations that were activated in algebraic or arithmetic processing (Gruber, 2001) including the precuneus, left intraparietal sulcus, left inferior frontal gyrus, and occipital lobe. Thinking of physics terms that are strongly associated with an equation need not entail calculation per se; it may simply trigger the retrieval of the equation. This retrieval may result in activation of regions involved in calculation even if the calculation is not actively occurring. Thinking of equation-based physics concepts engaged parietal regions: a language-based fact-retrieval region (extending superiorly to the intraparietal sulcus) seen in calculation tasks and an approximate calculation region (postcentral cluster extending posteriorly; Dehaene et al., 2004, 1999). In addition, the precuneus, left parietal, and left inferior frontal gyrus have been shown to activate in conjunction with executive processing and integration of visuospatial and linguistic information in calculation (Benn, Zheng, Wilkinson, Siegal, & Varley, 2012).

Energy-flow factor. Several physics concepts involve the idea of an energy or force that flows or radiates outward. Concepts with high factor scores on this factor included members of the "electricity" category (*electric field, direct current, voltage*), the "thermodynamics" category (*entropy*), and various other categories (*torque, waves*). Sensing the radiating warmth of the sun or fire is a universal experience that may be part of the basis of the brain processes that activated in association with this factor.

Energy flow may also entail visualization of a physical object that is related to the radiating of energy (e.g., *radio waves* could bring to mind a source, such as a musical instrument or an electronic communication device). Thinking about the semantic associations between abstract concepts and visualized concrete objects can modulate activity in the middle temporal area (Binder, Desai, Graves, & Conant, 2009). The voxel clusters in the classic language areas (left inferior frontal gyrus, superior temporal gyrus) associated with this factor activated during the decoding of abstract concepts (Wang et al., 2013). Thus, thinking of energy flow concepts may evoke activation in regions that are involved in sensing energy, visualizing concrete objects, and semantically linking the concrete objects to energy flow.

Summary

There is a commonality among people in how they neurally represent physics concepts. The commonality consists of the sets of brain regions involved, the brain-activation signatures of specific concepts, and the organizing factors that underlie the activation patterns for the 30 concepts. These schooling-acquired concepts engender a brain-based systematicity (i.e., a repeatable activation pattern) and commonality (across people).

This research lays the foundation for a neural description of physics comprehension that goes beyond brain locations. The localization of factors enabled the formulation of speculative descriptions of knowledge types and

informational codes. Although human brains are not expressly intended for representing physics knowledge, they are expressly intended for representing knowledge of the physical world. The findings suggest that physics schooling develops concepts grounded in the brain systems that represent the physical world.

Action Editor

Philippe G. Schyns served as action editor for this article.

Author Contributions

R. A. Mason and M. A. Just contributed equally to all aspects of this project.

Acknowledgments

We thank Tim Keller, Vladimir Cherkassky, Andrew Bauer, Chelsea McGrath, Xiaoxiao Lei, and Robert Vargas.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This work was supported by Office of Naval Research Grant N00014-131-0250.

Supplemental Material

Additional supporting information can be found at http://pss.sagepub.com/content/by/supplemental-data

Note

1. The rationale for performing a separate factor analysis in each region rather than one factor analysis for the entire cortex was to prevent any of the regions from dominating the set of input stable voxels, which the occipital regions would have otherwise. The choice of the particular number of voxels per region (120) was motivated by similar analyses in other data sets in which 120 was the smallest number of voxels that maximized classification accuracy.

References

Benn, Y., Zheng, Y., Wilkinson, I. D., Siegal, M., & Varley, R. (2012). Language in calculation: A core mechanism? *Neuropsychologia*, *50*, 1–10. doi:10.1016/j.neuropsychologia.2011.09.045

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*, 2767–2796. doi:10.1093/cercor/bhp055

Chen, H., Yang, Q., Liao, W., Gong, Q., & Shen, S. (2009). Evaluation of the effective connectivity of supplementary motor areas during motor imagery using Granger causality mapping. *NeuroImage*, *47*, 1844–1853. doi:10.1016/j.neuroimage.2009.06.026

Chen, J. L., Zatorre, R. J., & Penhune, V. B. (2006). Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. *NeuroImage*, *32*, 1771–1781. doi:10.1016/j.neuroimage.2006.04.207

Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain*, *125*, 1054–1069. doi:10.1093/brain/awf094

Cross, E. S., Hamilton, A. F. de C., & Grafton, S. T. (2006). Building a motor simulation de novo: Observation of dance by dancers. *NeuroImage*, *31*, 1257–1267. doi:10.1016/j.neuroimage.2006.01.033

Damarla, S. R., & Just, M. A. (2013). Decoding the representation of numerical values from brain activation patterns. *Human Brain Mapping*, *34*, 2624–2634. doi:10.1002/hbm.22087

Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, *56*, 384–398. doi:10.1016/j.neuron.2007.10.004

Dehaene, S., Molko, N., Cohen, L., & Wilson, A. J. (2004). Arithmetic and the brain. *Current Opinion in Neurobiology*, *14*, 218–224. doi:10.1016/j.conb.2004.03.008

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, *284*, 970–974.

Fugelsang, J. A., & Dunbar, K. N. (2005). Brain-based mechanisms underlying complex causal thinking. *Neuropsychologia*, *43*, 1204–1213. doi:10.1016/j.neuropsychologia.2004.10.012

Fugelsang, J. A., Roser, M. E., Corballis, P. M., Gazzaniga, M. S., & Dunbar, K. N. (2005). Brain mechanisms underlying perceptual causality. *Cognitive Brain Research*, *24*, 41–47. doi:10.1016/j.cogbrainres.2004.12.001

Gruber, O. (2001). Effects of domain-specific interference on brain activation associated with verbal working memory task performance. *Cerebral Cortex*, *11*, 1047–1055.

Han, S., Mao, L., Qin, J., Friederici, A. D., & Ge, J. (2011). Functional roles and cultural modulations of the medial prefrontal and parietal activity associated with causal attribution. *Neuropsychologia*, *49*, 83–91. doi:10.1016/j.neuropsychologia.2010.11.003

Jahn, G., Wendt, J., Lotze, M., Papenmeier, F., & Huff, M. (2012). Brain activation during spatial updating and attentive tracking of moving targets. *Brain and Cognition*, *78*, 105–113. doi:10.1016/j.bandc.2011.12.001

Just, M. A., Cherkassky, V. L., Aryal, S., & Mitchell, T. M. (2010). A neurosemantic theory of concrete noun representation based on the underlying brain codes. *PLoS ONE*, *5*(1), Article e8622. doi:10.1371/journal.pone.0008622

Just, M. A., Cherkassky, V. L., Buchweitz, A., Keller, T. A., & Mitchell, T. M. (2014). Identifying autism from neural representations of social interactions: Neurocognitive markers of autism. *PLoS ONE*, *9*(12), Article e113879. doi:10.1371/journal.pone.0113879

Kassam, K. S., Markey, A. R., Cherkassky, V. L., Loewenstein, G., & Just, M. A. (2013). Identifying emotions on the basis

of neural activation. *PLoS ONE*, *8*(6), Article e66032. doi:10.1371/journal.pone.0066032

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240. doi:10.1037/0033-295X.104.2.211

Maloney, D. P., O'Kuma, T. L., Hieggelke, C. J., & Van Heuvelen, A. (2001). Surveying students' conceptual knowledge of electricity and magnetism. *American Journal of Physics*, *69* (Suppl. S1), S12–S23. doi:10.1119/1.1371296

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, *7*, 293–299. doi:10.1016/S1364-6613(03)00134-7

McDermott, L. C. (1998). Students' conceptions and problem solving in mechanics. In A. Tiberghien, E. L. Jossem, & J. Barojas (Eds.), *Connecting research in physics education with teacher education* (Vol. 1, pp. 42–48). Retrieved from http://www.iupap-icpe.org/publications/teach1/ConnectingResInPhysEducWithTeacherEduc_Vol_1.pdf

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K.-M., Malave, V. L., Mason, R. A., & Just, M. A. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, *320*, 1191–1195. doi:10.1126/science.1152876

Niles, I., & Pease, A. (2001). Towards a standard upper ontology. In *Proceedings of the international conference on Formal Ontology in Information Systems—FOIS '01* (Vol. 2001, pp. 2–9). doi:10.1145/505168.505170

Petitto, L.-A., & Dunbar, K. N. (2009). Educational neuroscience: New discoveries from bilingual brains, scientific brains, and the educated mind. *Mind, Brain, and Education*, *3*, 185–197. doi:10.1111/j.1751-228X.2009.01069.x

Stylos, G., Evangelaki, G. A., & Kotsis, K. T. (2008). Misconceptions on classical mechanics by freshman university students: A case study in a Physics Department in Greece. *Themes in Science and Technology Education*, *1*, 157–177.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., . . . Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*, 273–289. doi:10.1006/nimg.2001.0978

Wang, J., Baucom, L. B., & Shinkareva, S. V. (2013). Decoding abstract and concrete concept representations based on single-trial fMRI data. *Human Brain Mapping*, *34*, 1133–1147. doi:10.1002/hbm.21498