www.sciencemag.org/cgi/content/full/329/5990/399/DC1

Supporting Online Material for

Machine Science

James Evans* and Andrey Rzhetsky

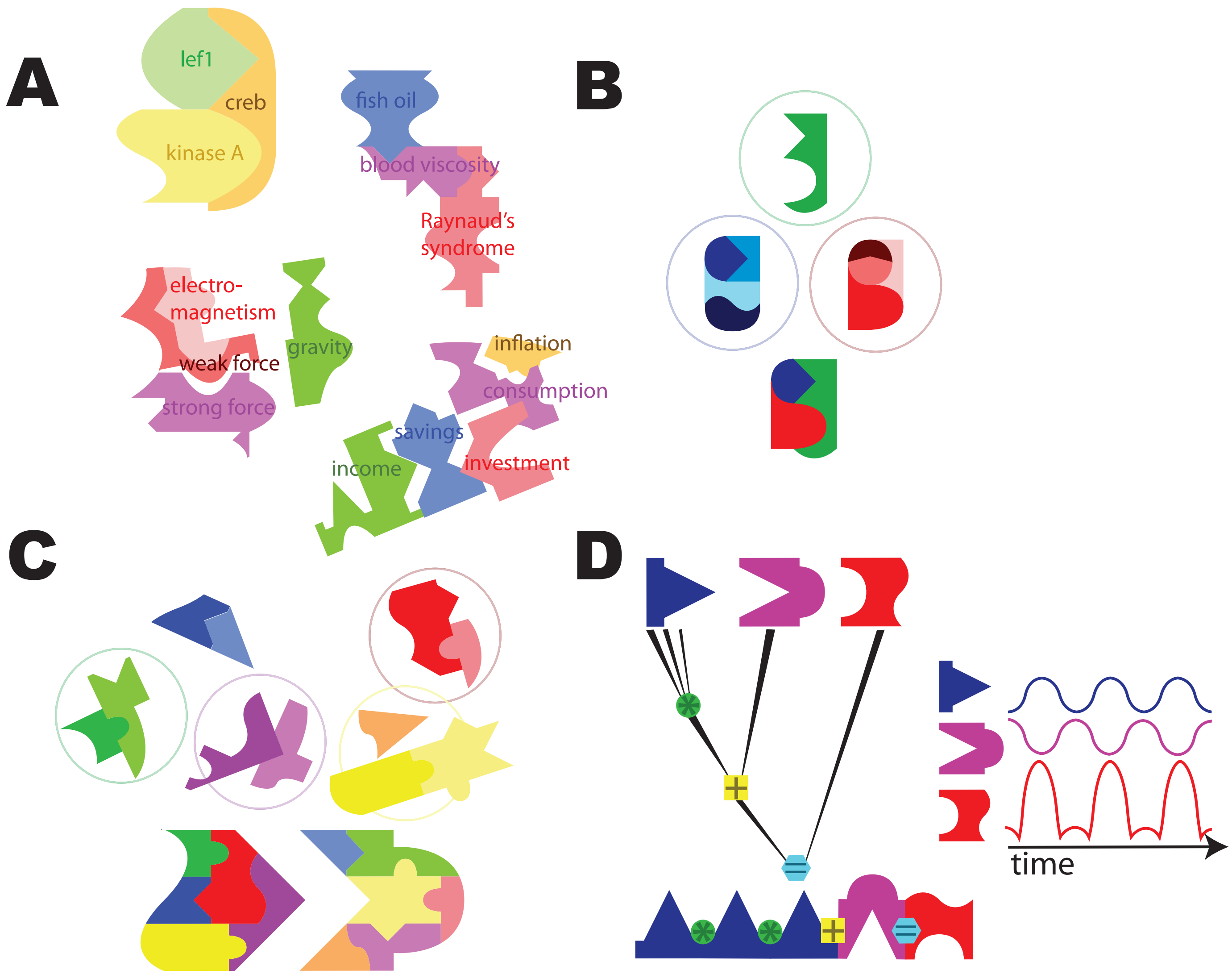
*To whom correspondence should be addressed. E-mail: jevans@uchicago.edu

Published 23 July 2010, *Science* **329**, 399 (2010)
DOI: 10.1126/science.1189416

This PDF file includes:

Fig. S1

Figure S1. Further representing human knowledge and hypotheses as jigsaw puzzles. (A) Concepts are rendered as puzzle pieces with edges that enable and constrain conceptual relations. Concepts assembled into a puzzle or conceptual network represent established theories. The first example from molecular biology illustrates two types of relations: T-cell protein **LEF1** *activates* (rendered as a sharp edge) **CREB**, a protein implicated in neural development, while **kinase A** *phosphorylates* (rendered as a rounded edge) it. Other examples are from medicine (fish oil as a treatment for Raynaud's syndrome), economics (macroeconomic variables), and physics (physical force theories). Some concepts are more complex than others (e.g., physical force theories may be more complex than a network of particles related by those forces), and some connections are precise and widely agreed upon (the unification of electromagnetic and weak forces in the electroweak force), while others shift and are widely contested (the relationship between macroeconomic variables, especially by microeconomists); (B) Three groups of concepts, from three distinctive scientific languages or ontologies, all attempting to describe the same underlying phenomenon, but with distinct, overlapping concepts. If the languages can be (partially) mapped to one another, then the analyst may computationally synthesize novel hypotheses that unite concepts across domains that have not previously be considered together; (D) Eleven concepts corresponding to five "input" and six "output" phenomena, spread across five overlapping communities, previously organized into cause-effect pairs. The concepts have been computationally reorganized into two higher-order, aggregate concepts and enabled the discovery of a higher-order relation between them; (E) The three concepts from Figure 1, but here a new composite *symbolic* relationship is inferred through patterns of their variation over time. This builds on a recent study in which Schmidt and Lipson use symbolic regression to derive equations that characterize physical systems based on patterns of covariation among system components (e.g, position coordinates, velocities, and accelerations of multiple points on a two-hinged pendulum). The space of possible equations is vast, but can be productively pruned in a manner analogous to coarse-graining, by requiring that equations predict connections between dynamics of system subcomponents [25,26]. Hypothetical models are then compared in terms of their parsimony and fit to data, and have been shown to reproduce many "laws" of physics (e.g., position coordinates lead to the discovery of manifold equations; velocities to energy laws; and accelerations to force identities).


A
lef1
creb
kinase A
fish oil
blood viscosity
Raynaud's
syndrome
electro-
magnetism
gravity
weak force
strong force
inflation
consumption
savings
income
investment
B
C
D
time