People always come up with *new* ideas to deal with the complex world. This seems natural, but too natural to answer *how* the ideas originate—sometimes proposals and hypotheses *just come out* when we generate plans for a problem and explain causes for an effect. Such sense of *natural* may come from inferring ***implicit knowledge*** given specific context or situation, *i.e.*, knowing how to solve a problem before starting to solve it. Implicit knowledge, a.k.a. *mindset*, does not directly lead to concrete solutions, but shapes the way to figure out the solutions, such as high-level methodologies and problem space representations. The use of implicit knowledge tend to vary in person due to diverse prior experiences, but sometimes tend to converge across individuals due to utility of the problem. In the crowd, the diverse mindsets provide multiple perspectives for a deeper understanding of a problem, but have to be aligned to a commonground for communicating with each other. Ideas thrive in the trade-off between inner bias and external utility; ideas emerge from the collision between diverse mindsets. The emergence, propagation, and synthesis of ideas drive human progress, from simple tool use to sophisticated scientific research.

I propose to study how people *form* and *use* implicit knowledge. The overall question can be analyzed hierarchically in Marr's paradigm (Marr, 1982) (see Fig. 1). **On the computational level,** I study from the view of individuals and the view of the crowd, given that people come with diverse mindsets. For individuals, I ask: When solving different problems, how does implicit knowledge control the interplay between personal bias and problem utility, such that solutions diverge on some problems, but converge on others? For the crowd, I ask: (1) How to leverage such diversity for synthesizing the diverse perspectives to a deeper understanding of the target problem? (2) How to reach a commonground through communications, such that ideas could propagate between diverse mindsets? **On the representational (algorithmic) level**, I build statistical models and programming language libraries, with updating operations defined upon, to model human priors in problem solving. **On the implementational level**, I build infrastructures for large-scaled behavioral studies to study how people externalize implicit knowledge; and to study how people internalize knowledge from the interface with other people and machines. Studies are launched from simple tasks that abstract the corresponding problems appropriately, toward domain-specific problems.

Understanding and modeling the use of implicit knowledge, *the origins of new ideas*, takes the first step toward my long-term goal—building an artificial intelligence (AI) agent that leverages the implicit knowledge from human to solve complex problems. The problems are connected to the ultimate goal—the computational and introspective study on scientific research.

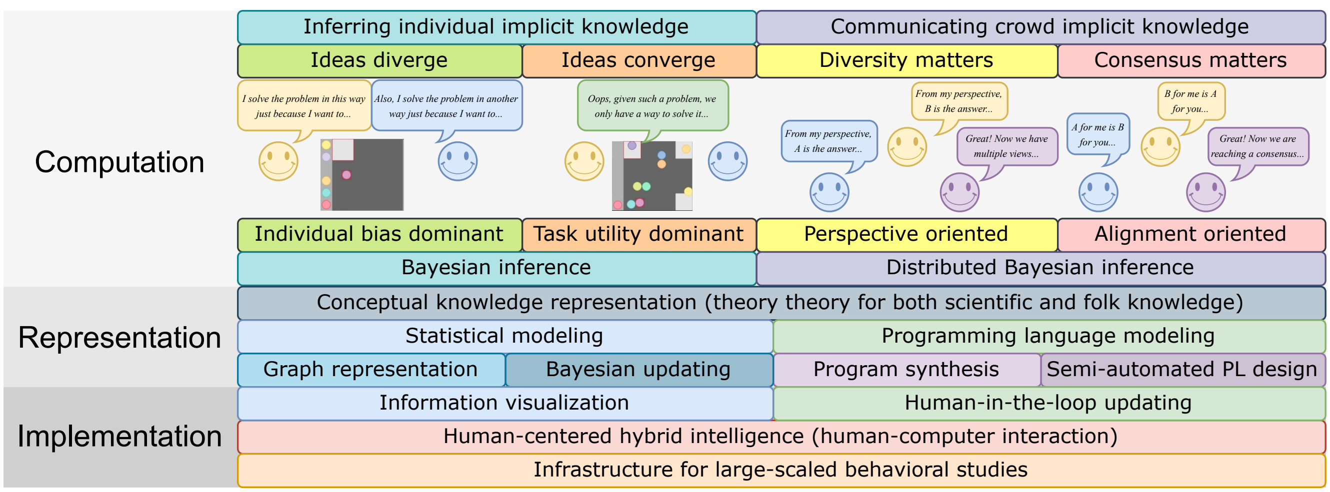


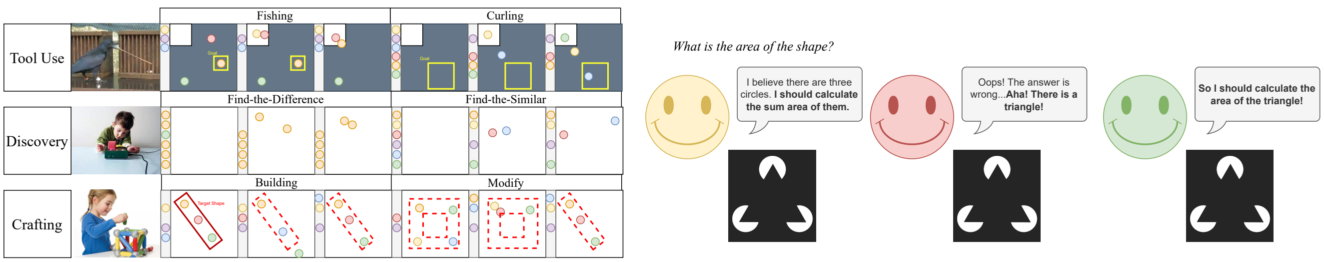
Fig 1: Investigating implicit knowledge on three levels

Computational level

To study computational level from the view of individuals, I frame two scenarios: (1) the interplay between implicit knowledge and task utility (Shi et al., 2022b); (2) the interplay between implicit knowledge and domain knowledge of problem (Shi et al., 2022c). These two scenarios has the potential to mimic planning and discovery in scientific research. From the view of the crowd, I frame two scenarios: (1) communicating between different mindsets over the same objective; (2) combining multiple perspectives over the same objective. These two scenarios has the potential to mimic idea propagation and synthesis in scientific research.

Individual-bias-dominant *vs.* task-utility-dominant Evidences from cognitive and developmental psychology support that people solve unseen problems with the help of high-level strategies according to prior knowledge that is agnostic to problem context. Specifically, I argue that people construct such strategies by assigning semantics to the elements in the problem to make connection with prior experience. Such assignment can be viewed as Bayesian inference given goals and constraints as prior, which implies the diversity and the convergence of strategies used by people. To understand and model the capability of people, I propose the **ProbSol Worlds (ProbSol)** environment (see Fig. 2(a))—the world is driven by the same dynamics but can be configured as different tasks, such as tool use, causal inference, and sketching. Equipped with magnetism-based dynamics, ProbSol alleviates the confounding variables of prior semantics brought by conventional physically-grounded problem solving tasks. Hence, the environment-agnostic prior of goals and constraints can be disentangled from problem solving. ProbSol has the potential for carrying out large-scale behavioral studies and benchmarking computational models to probe people's and machine's diverse understanding of goals and constraints in problem solving.

Insight-seeking *vs.* domain-knowledge-relying If scientific discovery is one of the main driving forces of human progress, insight is the fuel for the engine, which has long attracted behavior-level research to understand and model its underlying cognitive process. However, current tasks that abstract scientific discovery mostly focus on the emergence of insight, ignoring the special role played by domain knowledge. In this project, I view scientific discovery as an interplay between *thinking out of the box* that actively seeks insightful solutions and *thinking inside the box* that generalizes on conceptual domain knowledge to keep correct. Accordingly, I propose **Mindle** (see Fig. 2(b)), a semantic searching game that triggers scientific-discovery-like thinking spontaneously, as infrastructure for exploring scientific discovery on a large scale. On this basis, the meta-strategies for insights and the usage of concepts can be investigated reciprocally. In the pilot studies, several interesting observations inspire elaborated hypotheses on meta-strategies, context, and individual diversity for further investigations.



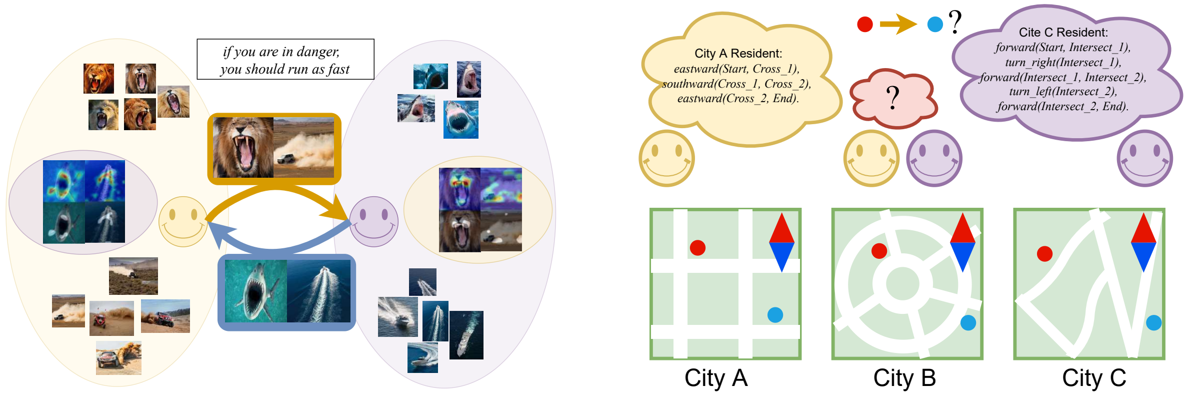
(a) The emergence of semantics

(b) The interplay between insight and knowledge

Fig 2: (a) Overview of the ProbSol Worlds. The simulated environment mimics physically-grounded problems in daily life. Circles denote objects, and areas with dark backgrounds indicate the inaccessible. These environments include three problem families: (1) tool use, such as fetching targets from inaccessible areas to manipulable areas with the help of other objects, reflecting the capability of meta-tool use and planning in both human and intelligent animals; (2) discovery, such as finding the only pair of objects that share the same in property, given a set of objects, reflecting the people's capability of causal inference in human; (3) crafting, such as arranging objects to resemble a given shape or transform a shape to another, reflecting the capability of concept abstracting and sketching in human. In these scenarios, magnetism plays totally different roles—helpful tool in tool use, clues and experimental materials in discovery, and harmful destroyer in crafting. **(b) Overview of insight in scientific discovery.** In a classic Gestalt problem, a problem solver first uses domain knowledge to analyze the problem, then seeks for insight once she gets trapped; after reconstructing the problem representation, she again uses domain knowledge to reach the solution. In this case, though domain knowledge constrains the thinking, it serves as the vehicle toward the target.

Communicating between diverse mindsets Abstract knowledge is one's explanation of perceptions from physical environments and interpretation of other minds from social environments—individuals gaining experience in different environments acquire divergent knowledge yet similar in essence. Hence, aligning knowledge (see Fig. 3(a)) or sharing environment (see Fig. 3(b)) become two ways toward reaching commongrounds. Modeling such communication needs to revise the pragmatics model—it is based on an ideal assumption that the speaker and listener share a prior model of the world—both of them map the object space to the semantic attribute space in the totally same way. But this prerequisite cannot always be satisfied in real world communications—different individuals have diverse world models due to distinctions between their backgrounds, understanding of concepts and pragmatics, while they may share an abstract semantic space as the prior to help bridge different worlds. Hence, I view this speaker-listener coordination problem as an extended version of cross-domain generalization. The key insight that distinct coordinate cross-domain generalization setting from the original one is that the two participants form a reciprocative learning process instead of a passive one. Thus in a view of system, the divergence between the two world models becomes an advance instead of an obstacle.

Combining multiple perspectives given diverse mindsets Understanding a complex concept through multiple perspectives can be formed as a distributed Bayesian inference problem (see Tab. 1 for details). First, a researcher cannot know everything about a piece of high-dimensional knowledge due to limited cognitive resources. This is in line with *the curse of dimensionality* in density estimation—let $P(x)$ be the marginal density that knowledge $x \in \mathcal{X}$ ($\mathcal{X}$ is the high-dimensional space of all modalities of knowledge), estimating $P(x)$ directly is intractable. Second, researchers have diverse mindsets due to expertise, experience, background, and interest. We can view the mindsets as prior z in latent space $\mathcal{Z}$ with all potential perspectives, where $P(z)$ is the distribution of the perspectives. Third, the human-centered knowledge can be viewed as the joint density $P(x, z)$, which can be generated through $P(z)P(x|z)$ according to Bayes theorem, where $P(x|z)$ the interpretation of knowledge x conditioned on perspective z. Finally, if the diversity of the researchers is enough and we are able to merge the perspectives, we obtain a deeper understanding of the piece of knowledge. Formally, merging perspectives is the marginalization over $P(x, z)$, *i.e.*, $P(x) = \sum_z P(x, z)$. In contrast to research synthesis, we explicitly model the process of



(a) Knowledge aligns experiences

(b) Environments shape language

Fig 3: **(a) Knowledge aligns experiences.** One growing up in the land refers to *if you are in danger then you run as fast* by showing two pictures, a tiger showing off sharp teeth and a jeep rushing dirt off sands; and her marine counterpart may use a shark opening mouth and a boat rushing in waves. Neuron activation maps visualize simulations of each other's minds. **(b) Environments shape languages.** In city A, people refer to directions with east or north, while they become left or forward in city C. But what language they would use at city B?

generating, interpreting, and merging. To note, this formulation views knowledge in a static way—if we append a temporal dimension to view science as a dynamic process, we may be able to study the evolution of science by integrating the propagation of perspectives into the graph-dynamics based science of science research as a constraint.

Table 1: The analogy between perspective synthesis and distributed Bayesian inference

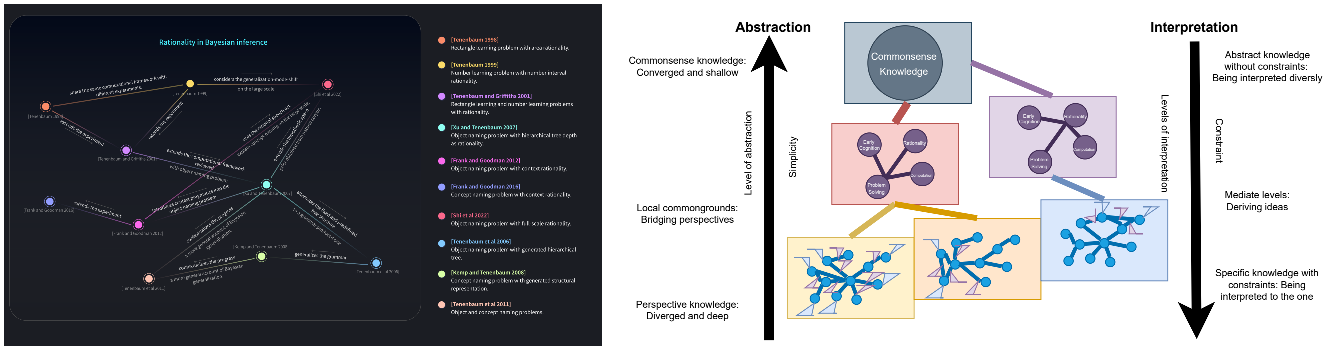
	challenge	prior	generation	merge
Bayesian inference	intractable $P(x)$	$P(z) \in \mathcal{Z}$	$P(x,z) = P(z)P(x\|z)$	$P(x) = \sum_z P(x,z)$
perspective synthesis	limited bandwidth	diverse prior	diverse interpretations	in-depth understanding

Representational level

To study the operations that mimic the use of implicit knowledge, I first define the representation of explicit conceptual knowledge—the use of implicit knowledge should be defined as operations over conceptual knowledge. I introduce two representation forms: (1) graph representation; and (2) programming language library representation. Further, I discuss the relation between perspective and commonground, as well as the role they are playing in implicit knowledge.

Graph representation Conceptual knowledge systematically combines declarative and procedural knowledge, consisting of both facts about the concepts and active processes about how concepts interact with each other. Though there are many perspectives on concept representation, I take the theory theory as a prerequisite because it is the most accepted theory both in terms of scientific knowledge representation and folk psychology (Gopnik and Wellman, 1994), thus we can formulate the domain knowledge under the form of theory theory. One implementation of theory theory is that concepts are maintained in a fully-connected network, where each concept is related to all other concepts in the set—that is, each concept may be shaped by all other concepts in the world. The graph is highly flexible—many calculi on fully-connected graphs can be applied to formalize the operations over concepts—such as the general pattern theory (*i.e.*, viewing all the nodes as words) (Grenander, 2012), or the representativeness of attribute (*i.e.*, viewing all other concepts as attributes of the concept) (Shi et al., 2022d). The graph matches generative modeling well, thus can be updated through sampling methods (Shi and Wu, 2022). If we view each node a paper projected to a specific perspective, the graph can also represent a frame in scientific research (see Fig. 4(a)).

Programming language library representation On the prerequisite of naturalism, the end-user of conceptual knowledge is human, while conceptual knowledge aims to describe the world. That is, people work with knowledge from diverse specific perspectives, but the world described by knowledge is the one. There seems to be a contradiction between the human-centered requirement and the content-centered nature of knowledge. Has any domain balance the description of the world and satisfying human need? The answer is yes. As a basic tool for computation, programming languages communicates between human users and the world—(1) a programming language is used to describe or model an object, a scenario, or a system, *e.g.*, 3D masks of human faces, biological experiment protocols, and electronic systems; (2) a programming language is designed to satisfy users' need and bias, such as the requirements for object-oriented programming, for functions, and for stream and procedures (Abelson and Sussman, 1996). The former comes as low-level instruction sets that only consider content modeling, while the latter comes as high-level programming languages that only consider user requirement, assuming that all programming languages can be translated to instructions. And the compiler bridges the two ends. Hence in knowledge representation, we also need such compilers that translate human-centered perspectives toward a mediator that can be merged with or be trans-



(a) (b)

Fig 4: (a) Showcase of the rationality perspective (visualized by Shi and Zhu (2022), a tool for generating frames of perspectives in scientific research); (b) Illustration of perspective and commonground as two ends of human scientific knowledge

lated to other perspectives. The compilers should capture how the perspective is framed from the content. And the mediator is a commonground that reaches the consensus between different perspectives. I have made initial trials on building a framework that (1) generates content-centered instruction proposals via statistical features from the natural corpus; and (2) supports human users to indicate their preference on instructions to form a human-centered program library. The framework works in the domain of biological experiment protocols (Shi et al., 2022a), and has the potential to be applied to arbitrary domains.

Perspective *vs.* commonground: rethinking the value of diversity If a personal perspective is a conditional view of the world, a commonground should be the consensus reached by a group of persons without conditioning. In the extreme situations, if the group of consensus is large enough to include all people in the world, the commonground hold by the group becomes the commonsense. Here we hypothesize that perspective and commonsense should be the two ends of the entire knowledge space, between which are different levels of commongrounds, different in levels according to group sizes (see Fig. 4(b)). Knowledge close to perspective is in-depth, narrow, and diverse across population; knowledge close to commonsense is shallow, general, and convergent over population. Going from perspective (lower levels of commonground) to commonsense (high levels of commonground) is reducing knowledge to their more basic and abstract forms, washing off condition-specific considerations and human bias; going from higher to lower levels of commonground is interpreting knowledge according to their bias and other constraints. This may echo the finding that semantic representations of abstract words are more diverse than that of concrete words—compared with concrete words, the abstract ones have 'a longer way' to be grounded, which means more choices for the interpretation. Similarly, given the great diversity of the science community, we must admit that personal bias may lead to scientific flaws, but some of them become the starting points of great breakthroughs in the history of science, since they are from numerous thinking based on both expertise and insight, such as the discovery of Kekule structure. Kekule's perspective on organic chemistry comes from strong personal bias. Hence, we should rethink the value of diversity—instead of trying to eliminate all bias in science, we can try embracing the bias that can be subsequently framed into perspectives which lead to meaningful research directions. After all, although commongrounds bridge the obstacles between sciences for communication and learning, it is perspective that extends the boundary of science, from the deep and narrow.

Personal Background

I am inspired by great thinkers and their ideas, including C. S. Peirce's Abduction, K. Popper's Hypothetical-Deduction, T. Kuhn's Science Evolution, F. Znaniecki's statement on Logology (the study on science), D. Marr's Levels of Analysis, U. Grenander's General Pattern Theory, J. Pearl's Causal Intervention, and M. Tomasello's Origins of Communications.

Skills I have developed strong capabilities to derive and implement computational models in Logic Programming, Neural Programming, Optimization-based Programming, and Probabilistic Programming. I am also a full-stack software developer—I write both environments for algorithmic evaluations and platforms for behavioral experiments.

Methodologies My work begins at intuitive yet under-researched intelligent phenomena. A first-step-work on introspective study should be: (1) formalize the problem with plausible evaluation metrics; (2) engineer computational models and run behavioral experiments in parallel, where human results inspire computational improvement and model limitations guide subsequent behavioral studies; (3) analyze complexity and convergence of the model to understand *why* it works; (4) define the problem as an evaluable task for AI community. During doctoral studies, I plan to take the first steps in my three research thrusts and leave deeper investigations for future work. I value working together with psychologists, mathematicians, and AI researchers, ranging from behavioral-level interpretations to computational theories with AI applications that have a broad impact on downstream fields.

References

Abelson, H. and Sussman, G. J. (1996). *Structure and interpretation of computer programs*. The MIT Press. 3

Gopnik, A. and Wellman, H. M. (1994). The theory theory. In *The Society for Research in Child Development Meeting*. 3

Grenander, U. (2012). *A calculus of ideas: a mathematical study of human thought*. World Scientific. 3

Marr, D. (1982). *Vision*. W. H. Freeman and Company. 1

Shi, Y.-Z., Bi, Z., and Zhu, Y. (2022a). The AutoBio instruction set manual. AutoBio. 4

Shi, Y.-Z., Hou, H., Ho, M. K., Yang, Y., and Zhu, Y. (2022b). Semantics emerge from solving problems given abstract prior. *ProbSol Concept Paper*. 1

Shi, Y.-Z. and Wu, Y. N. (2022). Generative Modeling Explained. In *Statistical Machine Learning Tutorials*. Department of Statistics, UCLA, Summer 2022 edition. 3

Shi, Y.-Z., Xu, M., Han, W., and Zhu, Y. (2022c). To think inside the box, or to think out of the box? Scientific discovery via the reciprocation of insights and concepts. *arXiv preprint arXiv:2211.16192*. 1

Shi, Y.-Z., Xu, M., Hopcroft, J. E., He, K., Tenenbaum, J. B., Zhu, S.-C., Wu, Y. N., Han, W., and Zhu, Y. (2022d). On the complexity of bayesian generalization. *arXiv preprint arXiv:2211.11033*. 3

Shi, Y.-Z. and Zhu, Y. (2022). PersLEARN: a platform for visualizing and synthesizing diverse perspectives. `https://perslearn.com/`. Accessed: 2022-09-30. 4