# Statement of Objectives

Yu-Zhe Shi

November 1, 2021

I study Abduction, an extended cognitive process of Induction, introduced by C.S.Peirce. Inspired by the philosopher, I hypothesis there exists another kind of core knowledge other than physical and psychological ones—human explicitly develop the way they explain the world and discover the unknown, and that is the generalized meaning of Abduction. As of present, most investigations on Abduction remain metaphysical debates or some programmatic algorithms professing the term, remaining the essence sort of under-researched. Hence, my long-term objective is to develop a unified perspective of Abduction as a significant component of human intelligence in three-fold: computational basis, cognitive underpinnings, and applications in AI.

Since sophomore year, I've been working closely with Dr.Song-Chun Zhu, Dr.Yixin Zhu, Dr.Stephen Muggleton, and Dr.Wang-Zhou Dai to pursuit the goal. I enjoy formulizing the world, with strong capabilities to derive and implement computational models in Logic Programming, Neural Programming, Optimization-based Programming, and Probabilistic Programming. I'm also a full-stack software developer—I write both backend environments for algorithmic evaluations and frontend platforms for behavioral experiments. I think hard to design impeccable behavioral experiments with integrity and completeness. These academic skills accelerate me iterate over new ideas.

In the preliminary work on Abduction, I constructed an agent discovering unknown unseen objects and operations when solving visual deciphering games and Minecraft puzzles—the agent succeeded given prior knowledge about dynamics of the world, solved most problems by maximizing the consistency between the composed theories and observations, iteratively in an expectation-maximization loop, or by MCMC sampling. However, paralleling agent's solutions with their human counterparts, I found the gap between them is much greater than expectation—humans solve the puzzles with much more flexibility, leveraging much weaker domain knowledge. Most interestingly, they describe the same way to hack problems with entirely different appearances, as well as exploit different heuristics to consider problems with seemingly similar appearances. How do humans discover such a variety of ways to solve problems, even without knowing the problems in advance?

This lead me to rethink the essence of Abduction—are there any prior other than prior knowledge? When failing to plug a USB-head into an interface, you tend to turn it upside down and make another trial, which often leads to a success—you are being epistemic in these scenarios, and knowing you should be epistemic when solving problems with similar structures. In another case, one is trying to add a LEGO block onto another—she keeps fine-tuning the posture of the block without changing the target and starting again from scratch—she is being persistent, and obviously is aware of being persistent in the tasks sharing the same problem structure with building LEGO blocks. A problem structure is a content-free hierarchical hypothesis space of solutions. Being epistemic and being persistent, in the example above, are two contrasting task-agnostic strategies to solve problems, which are beyond simply maximizing utility. Intuitively, humans compose theories from a qualitative language that uses such higher-level strategies as primitives, and generalize to problems with novel objects, unknown environments, but somehow similar structures. Once discovered structure of a family of problems, one solves them by stochastic trials under guidance of the speculative strategy.

H.A.Simon and D.Klahr have described problem solving as scientific discovery. I take a step further, by arguing that problem solving is not only about concept learning, theory acquisition, but also ways to solve them. I formulate general problem solving with three indicators—Flexibility, Rationality, and Acceptancibility. Flexibility describes how likely one is generating new hypotheses once the older one is inconsistent with observation. Computationally speaking, it rearranges human biases on different hierarchicals among the hypothesis space. Rationality directly relates to the top-level goal, indicating how strictly one follows the most rational reward of a given goal. Acceptancibility handles bottom-level learning, determining whether to revise a learned concept or to invent a new concept for novel objects in the problem. Under this formulation, humans explore a rich space spanned by the top-down goal planner, the bottom-up theory learner, and the bias controller.

Besides specific work, I also search for other intelligent phenomena as candidate evidences for my Abduction hypothesis: 1) human languages are shaped by environments physically and socially, and they know how to adapt the general pragmatics in different communication scenarios; 2) humans develop a variety of methdologies for scientific

discovery, varying in different domains; 3) humans explain the art, especially music, through synesthesia, and they are aware of how to appreciate music by different composers. These phenomena cover human cognitive processes from objective ones to subjective ones, sharing the same essence. I posit, as a complement to Bayesian Induction, Abduction highlights the core knowledge of how to assign pragmatic prior to select an explanation from hypotheses, and this is sort of in line with Laura Schulz's hypothesis on exploring how to explore. Most speculatory, let us consider a child building a house with simple LEGO blocks—she thinks much more than the house itself, but constructs an imagined world of her family living in it, with all plots of their happy life. She may excitedly ask her parents to visit the tiny world and guess what it is—while parents try to abduce, the child builds contexts in a Inverse-Abduction way, and this is similar to scientists designing puzzles, or musicians composing music. Obviously, complementing and verifying these ideas is a big question and is impossible to be solved throughout doctoral career. My objective for PhD study is to take a preliminary step towards computational formulations of these phenomena.

At BCS, my research interest matches that of Dr.Josh Tenenbaum, Dr.Laura Schulz, Dr.Rebecca Saxe, and Dr.Nancy Kanwisher. My investigations on Abduction may extend Josh's research boundary of modeling human core knowledge; Abduction may provide Laura with a perspective to model the development of thinking about ideas; with Rebecca I can study how social environment shapes human communications and pragmatics; I may also explore if there exists brain circuits that function like Abduction, collaborating with Nancy.

I hope to take the first significant step towards understanding and modeling Abduction, in company with BCS.