Toward Complex and Structured Goals in Reinforcement Learning

Guy Davidson
guy.davidson@nyu.edu
Center for Data Science
New York University

Todd M. Gureckis
todd.gureckis@nyu.edu
Department of Psychology
New York University

Abstract

Goals play a central role in the study of agentic behavior. But what is a goal, and how should we best represent them? The traditional reinforcement learning answer is that all goals are expressible as the maximization of future rewards. While parsimonious, such a definition seems insufficient when viewed from both the perspective of humans specifying goals to machines and autotelic agents that self-propose tasks to explore and learn. We offer a critical perspective on the distillation of all goals directly into reward functions. We identify key features we believe goal representations ought to have, and then offer a proposal we believe meets those considerations.

The concept of "goals" is central to the study of reinforcement learning (RL) and agentic behavior more generally. But what is a goal? The traditional answer within the RL community is that all goals are expressible in an identical fashion — the maximization of the (discounted) sum of future (scalar) rewards (the *reward hypothesis*, Sutton, 2004). In this sense, goals are simply preferences over state-action histories (Bowling et al., 2023). While potentially universal and parsimonious, such a granular definition seems insufficient for several reasons.

First, consider training a household cleaning robot through reinforcement learning. The robot should act safely, without getting stuck, avoid collisions with objects and beings, be efficient, and not run out of battery. Instead of representing a reward value for each state transition, people specify and discuss such goals in more abstract and communicable terms. How can we create machines that can learn to achieve goals specified in these more abstract but intuitive ways? Second, consider the ways that people construct goals for themselves. For instance, children create playful goals for themselves during play, such as making a "truck-carrier truck" or "taking my stuffed animals to the bookstore." These self-proposed goals help children learn how to structure problem spaces and search for solutions (Chu & Schulz, 2020; Lillard, 2015; Andersen et al., 2023). How can we create agents that propose rich, structured, and creative goals for themselves, fostering self-guided learning?

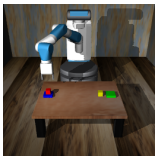
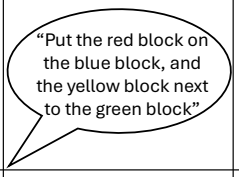
In this paper, we argue that thinking about goals merely as low-level reward functions to be maximized is insufficient. Instead, we propose a re-imagining of the role and representation of goals within RL. We argue that if we want to develop agents that accomplish diverse tasks across different environments, we need agents that can propose and pursue rich, complex, and creative goals. Our proposal draws inspiration from the cognitive science of how humans think about and create goals. Although several of the ideas summarized here have important antecedents in the prior RL (Oudeyer et al., 2007; Colas et al., 2022) and cognitive science (Molinaro & Collins, 2023; Chu et al., 2024) literature, we aim to provide a succinct and compact argument for this cluster of emerging ideas.

1 Key desiderata in the representation of goals

We begin by proposing key desiderata for the representation of goals in humans and machines: (1) abstraction, (2) temporal extension, (3) compositionality, and (4) grounding in behavior. These are semantic aspects of goals that we believe goal representations should support and make accessible

to agents; representation formats that fail to promote these qualities will struggle to represent the breadth and complexity of possible goals. We review how existing goal representation approaches fare under these properties (Figure 1 summarizes our findings). We focus our discussion on methods that explicitly present an agent with a goal, such as methods that leverage target positions (place the agent or manipulator in a particular position, e.g. Plappert et al., 2018) or image-based observations (match the agent's observation to a goal image, e.g. Florensa et al., 2018; Warde-Farley et al., 2018; Nair et al., 2018). We also discuss language-based approaches, including ones where the environment specifies a natural language task (Hill et al., 2019), approaches procedurally generating goals from minimal, limited grammars (Colas et al., 2020; Akakzia et al., 2021), language-based exploration approaches (Du et al., 2023b), and methods that use large multimodal models to marry language and image observations (Rocamonde et al., 2024; Baumli et al., 2023). We then consider an alternative approach that seems to meet these desiderata that relies on more explicit, symbolic program representations of goals. Our purpose in this comparison is to highlight the relative merits of different representations and stimulate a discussion of how the field should think about goals.

Goal Representation Approaches	**Directly Encoded via Reward Function**	**Target Observation or Embedding**	**Natural Language**	**Program**

Legend
E : easy
M : moderate
D : difficult
Goal Desiderata
				`(and`
				`(on blue red)`
				`(adjacent yellow green)`
				`)`

Abstract Goals ("be efficient", "be safe", §1.1)	Challenging reward function engineering D	Impossible (?) D	Trivial to specify, but how do you ground? M	How to write programs for abstract goals? D
Goal Component Abstraction (lift from specific to general, §1.1)	Nontrivial reward function engineering M	Requires embedding abstractions, e.g. ("any block on any block") D	Abstraction in natural language is trivial, grounding harder E	Trivial for abstractions defined in program grammar E
Temporal Extension (goals requiring temporal reasoning, §1.2)	Only to the extent states encode temporally-extended information D	Only to the extent states encode temporally-extended information D	Natural language allows specifying temporally extended goals E	Program grammar dependent, but can be supported E
Compositionality (combine goals to represent new ones, §1.3)	Compose mathematically not necessarily linguistically M	Requires learning compositional operators over embeddings D	Natural language is highly compositional E	Naturally compositional to the extent defined by their grammar E
Grounding (determine goal achievement, §1.4)	By construction E	State matching or metric over embeddings E	Easy if environment is textual, hard if not, main language challenge D	Requires environment-specific program interpreter D

Figure 1: **Toward Complex and Structured Goals in Reinforcement Learning.** We summarize our observations across four broad categories of goal representations and several desiderata of goal semantics. Across the several considerations we examine (table rows), we note that prevailing goal representations (either implicitly encoded to the agent as a reward function or specified as a target observation or embedding) facilitate goal grounding but struggle with other desiderata. Conversely, goals represented in natural language or as programs (§2) enable many representational considerations but are substantially harder to ground to behavior.

1.1 Abstraction

Abstraction is a core property of human concept representations (Murphy, 2004; Hampton, 2003; Yee, 2019). In addition to concrete and tangible goals, we can also consider goals at varying levels of abstraction. While the representation and pursuit of abstract goals have received less attention (Gollwitzer & Moskowitz, 1996; Chu et al., 2024), there is evidence people abstractly represent the values of different options in light of changing goals (De Martino & Cortese, 2023). Human goals can be highly specific, such as stacking a particular set of blocks on a desk. People can also abstract away towards a goal like stacking any available object on any available surface. At the opposite end of the spectrum, goals can be entirely abstract, such as "do something fun", "act safely" or "learn

new things." We consider two separate questions regarding abstractions: **(1)** to what degree do different goal representations facilitate representing these types of abstract goals? and **(2)** to what degree do different goal representations enable flexibly abstracting components of a specific goal, such as moving from "stack blue blocks on the desk" to "stack as many objects as you can?"

Representing abstract goals: Humans can conceive of and act in accordance with both specific and abstract goals. For example, a person can aim to reach a desired state (e.g., the finish line of a marathon) or more abstract goals (e.g., living a healthy lifestyle). Some abstract goals can partly be mapped onto traditional reward functions. For example, "be efficient" mapped onto a small per-step penalty, or "be safe" into a penalty for incapacitation. Such mappings, however, can be limited. They require researchers to pre-specify important behaviors and map them onto rewards. For instance, a penalty for incapacitation might capture some safety-related objectives but miss other ones. They might also struggle to generalize between environments, as an appropriate per-step penalty in one domain might be extreme in another. Other abstract goals may be more challenging to map onto rewards (e.g., "help the other agent but don't trivialize the task for them"). Fundamentally, this approach requires researchers to consider edge cases and map them onto rewards, as opposed to specifying principles for a system and allowing it to discover how to best embody them.

Abstracting goal components: Consider an example domain of object manipulation using a robotic arm. Any particular configuration of objects, such as placing the blue block on the red block, could be specified using a visual observation (or, for that matter, a symbolic state representation). If we train a robot to accomplish many different stacking configurations, we might hope its policy generalizes to novel ones. However, even if it does, we would not be able to abstractly specify the goal of stacking any two objects on each other (or, say, stacking a non-block object on a block).

Language, of course, offers a way forward. Early approaches leveraged minimal language-generating grammars to represent limited goals in specific domains (Colas et al., 2020; Akakzia et al., 2021). More recently, advancements in large language models (LLMs) facilitate specifying goals using the flexibility of natural language. In section subsection 1.4, we will fully consider the challenge of grounding a goal representation to identify its achievement; for the time being, we remark that language-based approaches are split into two main categories. One class of approaches leverages environments where state representations are either symbolic (facilitating captioning) or include a natural language component. These include work from Hill et al. (2019) emitting language instructions from the environment and language-for-exploration approaches such as ELLM (Du et al., 2023b) and LMA3 (Colas et al., 2023). A second class of approaches uses multimodal models to bind between language and image observations, using VLMs to detect success (Du et al., 2023a) or offer rewards (Rocamonde et al., 2024; Baumli et al., 2023). While these approaches show great promise and appear to scale well, current work has not explicitly measured their ability to supervise goals with specific objects compared to their abstract counterparts.

1.2 Temporal extension

One particular kind of abstraction that merits separate discussion is abstraction over time. Humans find it natural to consider goals that require more than a single moment in time to evaluate and achieve. For example, some goals might be achieved with a single well-executed action but require longer to evaluate, such as making a basketball trick shot ("over the second rafter, off the floor, nothing but net", as Michael Jordan does in a 1993 McDonald's commercial). At the other extreme, human goals may extend over arbitrarily long time horizons; a man in the UK set out to park in each and every spot in a large parking lot required over six years to do so (Yuhas, 2021 thanks to Chu et al., 2024). As the parking lot had 211 spots, naively representing the full space of previously parked spots would require 2^{211} states, an unfathomably large number, and yet we can intuitively understand and pursue such a goal. If we can represent and evaluate goals with sequential preferences over arbitrary time horizons, can we endow artificial agents with a similar ability?

Goal representations in reinforcement learning, to date, have largely eschewed this complexity. If we specify a goal through an observation, e.g. a robotic manipulator position or configuration of

objects, that format does not enable temporal extension. While trajectory demonstrations have long been used for learning (Schaal, 1996) or as specifications to infer a reward function from (Ng & Russell, 2000), prior work does not directly specify goals as trajectories. Similarly, language-based approaches tend to compare a language-specified goal to an embedding of the current environment state. As improvements in VLMs facilitate reasoning over longer contexts, nothing in principle limits them from being used to offer reward over longer sets of observations; Du et al. (2023a) demonstrate this by fine-tuning Flamingo (Alayrac et al., 2022) to score success on short clips. In these settings, goals are usually represented by natural language descriptions — which would also challenge agents to learn to condition long-term policies on a single goal description or to generate suitable subgoals to assist in pursuing the goal. A final class of approaches explicitly represents temporal tasks. These methods use temporal logic (Littman et al., 2017; Icarte et al., 2018b; 2022), latent-space subgoals (Fang et al., 2022), or combine both by specifying tasks textually in temporal logic (Leon et al., 2022). These approaches offer promising templates for defining extended goals; it is unclear at present if they also facilitate other desiderata, such as abstraction and compositional specification.

1.3 Compositionality

The human ability to compositionally combine concepts to represent and understand the world around us is well-studied (Murphy, 2004; Ward, 1994; Frankland & Greene, 2020; Lake & Baroni, 2023). This compositionality extends to our goals: once we can set a goal for ourselves like "build a block tower" and "throw balls onto the desk", we can extend this to novel goals like "build a block tower on the desk" or "throw balls at the block tower." RL approaches often evaluate the ability of their agents to learn compositional *policies* (e.g., holding out some goal configurations and evaluating agents on the held-out goals). Here we consider how goal representations can facilitate the ability to *generate* compositional goals as is required for *autotelic* agents that learn by creating new tasks for themselves (Colas et al., 2022; Akakzia et al., 2021).

We begin by examining representing world states, such as object configurations, through their (usually image-based) observation. Generating observations that match compositions is trivial: if we can place the red cube on the blue cube, and separately the blue cube on the yellow cube, we can also stack all three cubes. However, systematically composing these combinations is harder; it is unclear how to compose observations of "red on blue" and "blue on yellow" to produce an observation that would guide toward "red on blue on yellow." The compositionality of natural language (Goldberg, 2015) facilitates leveraging natural language for compositional goal descriptions. While natural language goal representations likely compose well, the ability to detect goal achievement may not be a trivial consequence: for instance, while SuccessVQA (Du et al., 2023a) detects goal satisfaction for new agents on old tasks, it falls to near-chance performance on held-out goals. Logical operations, such as conjunctions and negation, offer another type of compositional test. While Hill et al. (2019) demonstrate compositional generalization to held-out objects, their language-based approach fails to generalize to negations of previous instructions; the explicitly logical method proposed by Leon et al. (2022) fares better. In summary, compositionality offers three nested challenges: generating compositional goals, rewarding their achievement, and training agents to pursue them.

1.4 Grounding in behavior

Human goal representations enable many different types of processes. For example, people can act instrumentally toward goals, be it in a model-free fashion or by planning how to achieve them. They can observe another agent acting and infer their likely goal (Jara-Ettinger, 2019). They can also evaluate their own or another agent's behavior with respect to a stated goal and identify whether or not and how successful they are in achieving it. This last ability is critical to training agents to pursue these goals: if we cannot ground a goal to behaviors that accomplish it, we cannot provide a reward signal or other feedback. Therefore, we evaluate the various goal representations we survey on how concrete their semantics are, and how much effort is required to ground them to behaviors.

Image- or single-state-based goals are the most trivial to ground. If we can generate a state representation or observation, and define a distance metric over state (or latent) space, we can reward goals using this metric. While not all distance metrics and state representations offer an equally good signal towards goal achievement (Akella et al., 2023), some representations even allow interpolating in latent space to plan toward a goal (Eysenbach et al., 2024). Grounding language-based goals depends on both the environment and the complexity of the goal. With environments that admit a language state description, grounding can be implemented as a similarity comparison between embeddings of the state and goal without additional engineering efforts. Symbolic states allow hard-coding a state captioning system; Du et al. (2023b) offer evidence that agents using a hard-coded captioner fare better than those using a learned one, indicating that caption quality can be a bottleneck. Environments using visual observations and language goals use multimodal models to align between the modalities. These approaches appear to benefit from increasing model scale (Rocamonde et al., 2024; Baumli et al., 2023), but how such approaches fare with increasingly elaborate goals is an open question; evidence from video-based tasks suggests models' ability to reason about complex queries over long videos remains limited compared to humans' (Rawal et al., 2024).

2 Program representations of goals

Given these desiderata, we consider another approach to representing goals, by treating them as symbolic programs. Specifically, we examine a proposal to model human goal generation as synthesizing reward-producing programs (Davidson et al., 2024). These are symbolic programs explicitly representing goal semantics, supporting compositional recombination, that are interpretable to detect partial or complete goal achievement. Goals have long been implicitly specified to agents using programs as the reward functions implemented in artificial simulation environments. Here, we consider explicitly endowing agents with access to symbolic program goal representations. We consider programs under the set of desiderata surveyed above and then discuss several implications.

2.1 Goal program desiderata

Abstraction: Abstracting within components of programs is rather natural. Taking as an example the LISP-like syntax adopted by Davidson et al. (2024), modifying a goal to act over different objects requires modifying a variable declaration, as objects are referenced via variable quantification. Abstracting a goal to a superordinate relation might require further modification; but as programs explicitly denote semantics, defining a grammar that facilitates abstraction is achievable.

Temporal extension: (stateful) program representations could encode goals with arbitrarily long time horizons. For instance, the representation used by Davidson et al. (2024) is interpreted into a state machine that then acts as a reward function, inspired by reward machines (Icarte et al., 2018a; 2022). Such programs have not yet been used to train agents, though, leaving the question open of how to best embed them as goals on which to condition an agent's behavior.

Compositionality: Programs are naturally compositional to the extent defined by their syntax—program-based approaches trivially allow combining multiple reward signals (e.g. Eureka and Dr. Eureka's (Ma et al., 2023; 2024) generated reward functions), and the program goals generated by Davidson et al. (2024) allow composing separate preferences to structure an overall goal.

Grounding in behavior: Grounding programs depends on the environment's state representation and the nature of the programs. For example, Eureka and Dr. Eureka generate Python reward functions (and do not require a custom interpreter) and operate directly on environment states (aided by inspecting environment source code). At the other end, Davidson et al. (2024) implemented a custom interpreter to parse their domain-specific language into state machines, which also operate over symbolic state representations. Grounding is simplified by approaches that interpret into the same language used to specify environments, and by symbolic states; conversely, if an environment does not admit a symbolic representation at all, grounding programs to it may be challenging.

2.2 Goal programs and goal-conditioned policies

One key property separating some of the approaches reviewed in this paper from others is whether or not the agent's policy is goal-conditioned. We view this distinction as crucial because to condition the policy on a goal, it needs to be represented in a suitable fashion (usually as an embedding in some high-dimensional space). This naturally encourages representations that either match the observations the agent learns to embed as they interact with the environment, or textual observations embeddable using a language model. Program-based goals (implemented by reward functions) have been used in non-goal-conditioned settings by the Eureka family of methods, and more broadly, every simulated environment requires specifying a function to compute reward. Conversely, no prior works embedded a goal program in a manner that enables an agent to condition its policy on the program, and it is unclear if single embeddings would fully capture the rich semantics of programs.

2.3 The agent-environment boundary and the many uses of goal representations

The goal-conditioning distinction relates to the broader question of where to draw the agent-environment boundary. A narrow view of a reinforcement learning agent is one where the agent's only interaction with the concept of a goal is through the environment-defined reward function. Introducing explicit goals to the agent has several benefits. Goal-conditioned agents allow learning a single policy capable of pursuing distinct goals, hopefully facilitating generalization to novel goals with minimal further learning. Agents may also propose their own goals to better explore their environment and develop baseline skills, as proposed in Colas et al.'s (2022) *autotelic* framework.

Our desiderata for goal representations are motivated in part by the many ways in which people use their goal representations. We can use a goal to guide our behavior or to plan toward achieving it. We can also attempt to infer another agent's goal from watching them behave (not unlike inverse RL), or evaluate an agent's progress toward a known goal (playing the part of a success detector or reward function). From the perspective of reinforcement learning, perhaps the agent-environment boundary also passes somewhere within the human mind, and we can consider the RL agent to only model the policy learning aspect (which may be neurally separate from other aspects of decision-making, see Niv, 2009, for a review). However, if we are interested in building agents that can propose and pursue their own goals and maybe even infer ours and assist us in solving them, we should consider how to represent goals in a manner that facilitates these separate yet related skills.

3 Discussion

We offer an argument to revisit the role and representation of goals in reinforcement learning. We present several desiderata for goal representations, inspired by the many functions of goals in human cognition, and review how current RL methods fare under these considerations. We then discuss a recent proposal to explicitly represent goals as symbolic, evaluable programs and situate it with respect to the agent-environment boundary question. We conclude with a few discussion points:

We chose to highlight a specific set of desiderata, which we do not view as all-encompassing. We hope this work fosters a discussion about what properties are helpful in constructing effective goals. For instance, communication is another key property of goals, one that might be particularly crucial in multi-agent or human-machine collaboration settings. Another desirable quality might be the ability to generate goals to maximize a more generalized information-seeking principle, such as learning progress or empowerment—though it remains to be seen if this is a property of the goal representation or of the goal-generating algorithm.

We note that representing abstract goals to agents remains an open challenge. One path to scale up manually implemented auxiliary rewards is to build on approaches such as Eureka and Dr. Eureka (Ma et al., 2023; 2024), which use LLMs to synthesize task-specific reward functions (coupled with safety-related instructions). Programs, which naturally offer abstraction mechanisms, offer one alternative toward explicitly specifying abstract goals to agents. Another way forward might come

from AI alignment research, which develops methods designed to impart desired behaviors to LLMs. For instance, constitutional AI (Bai et al., 2022) tries to improve the ability of models to act in accordance with a set of principles; however, developing the supervision signal over agent interactions in arbitrary environments, as opposed to textual LLM outputs, may require substantial work.

If we propose sequential preferences over temporally extended goals as a desirable property, how does that interact with the Markov assumption? A state representation that incorporates physical information about a single moment in time may not suffice to reason about temporal goals. Consider the challenge of making a shot "over the second rafter, off the floor, nothing but net." We could detect each of these conditions as they happen (e.g., the ball is over a particular rafter), but we could not define this goal as a reward function over singular moments. While this particular goal might be resolved by treating a short time horizon as the state, we believe that for any (finite) horizon, we could construct a goal requiring reasoning beyond it. However, for any temporally extended goal, we envision there exists some auxiliary state construction that transforms the goal to be Markovian, in line with the reward machines proposed by Icarte et al. (2018a). We highlight two open questions. First, how can these approaches be married to goal-conditioned RL to train agents to construct and solve many such goals? Second, what can we learn from human cognition about solving temporally extended goals? Do people construct task-specific MDPs, or how else do we pursue such behavior?

Finally, though we draw inspiration from human psychology, we note that psychology, too, offers limited definitions of goals. Goals are prevalent across psychological research (Dweck, 1992; Austin & Vancouver, 1996), having been studied from perspectives such as motivation (Hyland, 1988; Eccles & Wigfield, 2002; Brown, 2007), personality and social psychology (Fishbach & Ferguson, 2007; Pervin, 2015), and learning and decision-making (Moskowitz & Grant, 2009; Molinaro & Collins, 2023; Chu et al., 2024). For all these efforts, goals are often discussed absent a technical definition, and when one is provided, it is often vague or simplified, e.g., defining goals as future objects to be approached or avoided (Elliot & Fryer, 2008). Future work should strive to reconcile between the richness of goals and goal-directed behavior and the narrow scope of definitions offered.

References

Ahmed Akakzia, Cédric Colas, Pierre-Yves Oudeyer, Mohamed Chetouani, and Olivier Sigaud. Grounding language to autonomously-acquired skills via goal generation. In *ICLR 2021-Ninth International Conference on Learning Representation*, 2021.

Ravi Tej Akella, Benjamin Eysenbach, Jeff Schneider, and Russ Salakhutdinov. Distributional distance classifiers for goal-conditioned reinforcement learning. 2023.

Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, Antoine Miech, Iain Barr, Yana Hasson, Karel Lenc, Arthur Mensch, Katherine Millican, Malcolm Reynolds, Roman Ring, Eliza Rutherford, Serkan Cabi, Tengda Han, Zhitao Gong, Sina Samangooei, Marianne Monteiro, Jacob L Menick, Sebastian Borgeaud, Andy Brock, Aida Nematzadeh, Sahand Sharifzadeh, Mikoł aj Bińkowski, Ricardo Barreira, Oriol Vinyals, Andrew Zisserman, and Karén Simonyan. Flamingo: a visual language model for few-shot learning. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 23716–23736. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/960a172bc7fbf0177ccccbb411a7d800-Paper-Conference.pdf.

Marc Malmdorf Andersen, Julian Kiverstein, Mark Miller, and Andreas Roepstorff. Play in predictive minds: A cognitive theory of play. *Psychological Review*, 130:462–479, 6 2023. ISSN 19391471. doi: 10.1037/REV0000369.

James T. Austin and Jeffrey B. Vancouver. Goal constructs in psychology: Structure, process, and content. *Psychological Bulletin*, 120:338–375, 11 1996. ISSN 0033-2909. doi: 10.1037/0033-2909.120.3.338. URL /record/1996-01405-002.

Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, Carol Chen, Catherine Olsson, Christopher Olah, Danny Hernandez, Dawn Drain, Deep Ganguli, Dustin Li, Eli Tran-Johnson, Ethan Perez, Jamie Kerr, Jared Mueller, Jeffrey Ladish, Joshua Landau, Kamal Ndousse, Kamile Lukosuite, Liane Lovitt, Michael Sellitto, Nelson Elhage, Nicholas Schiefer, Noemi Mercado, Nova DasSarma, Robert Lasenby, Robin Larson, Sam Ringer, Scott Johnston, Shauna Kravec, Sheer El Showk, Stanislav Fort, Tamera Lanham, Timothy Telleen-Lawton, Tom Conerly, Tom Henighan, Tristan Hume, Samuel R Bowman, Zac Hatfield-Dodds, Ben Mann, Dario Amodei, Nicholas Joseph, Sam McCandlish, Tom Brown, and Jared Kaplan. Constitutional AI: Harmlessness from AI feedback. December 2022.

Kate Baumli, Satinder Baveja, Feryal Behbahani, Harris Chan, Gheorghe Comanici, Sebastian Flennerhag, Maxime Gazeau, Kristian Holsheimer, Dan Horgan, Michael Laskin, Clare Lyle, Hussain Masoom, Kay McKinney, Volodymyr Mnih, Alexander Neitz, Fabio Pardo, Jack Parker-Holder, John Quan, Tim Rocktäschel, Himanshu Sahni, Tom Schaul, Yannick Schroecker, Stephen Spencer, Richie Steigerwald, Luyu Wang, and Lei Zhang. Vision-Language models as a source of rewards. December 2023.

Michael Bowling, John D Martin, David Abel, and Will Dabney. Settling the reward hypothesis. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 3003–3020. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/bowling23a.html.

L.V. Brown. *Psychology of Motivation*. Nova Science Publishers, 2007. ISBN 9781600215988. URL https://books.google.com/books?id=hzPCuKfpXLMC.

Junyi Chu and Laura E. Schulz. Play, curiosity, and cognition. *Annual Review of Developmental Psychology*, 2:317–343, 12 2020. ISSN 2640-7922. doi: 10.1146/ANNUREV-DEVPSYCH-070120-014806. URL https://www.annualreviews.org/doi/abs/10.1146/annurev-devpsych-070120-014806.

Junyi Chu, Joshua B Tenenbaum, and Laura E Schulz. In praise of folly: flexible goals and human cognition. *Trends Cogn. Sci.*, April 2024.

Cédric Colas, Laetitia Teodorescu, Pierre-Yves Oudeyer, Xingdi Yuan, and Marc-Alexandre Côté. Augmenting autotelic agents with large language models. In *Conference on Lifelong Learning Agents*, pp. 205–226. PMLR, 2023.

Cédric Colas, Tristan Karch, Nicolas Lair, Jean Michel Dussoux, Clément Moulin-Frier, Peter Ford Dominey, and Pierre Yves Oudeyer. Language as a cognitive tool to imagine goals in curiosity-driven exploration. *Advances in Neural Information Processing Systems*, 2020-Decem, 2 2020. ISSN 10495258. URL https://arxiv.org/abs/2002.09253v4.

Cédric Colas, Tristan Karch, Olivier Sigaud, and Pierre-Yves Oudeyer. Autotelic agents with intrinsically motivated goal-conditioned reinforcement learning: a short survey. *Journal of Artificial Intelligence Research*, 74:1159–1199, 07 2022. URL https://arxiv.org/abs/2012.09830v5.

Guy Davidson, Graham Todd, Julian Togelius, Todd M. Gureckis, and Brenden M. Lake. Goals as Reward-Producing Programs. *arXiv preprint arXiv:2405.13242*, May 2024. URL https://arxiv.org/abs/2405.13242.

Benedetto De Martino and Aurelio Cortese. Goals, usefulness and abstraction in value-based choice. *Trends Cogn. Sci.*, 27(1):65–80, January 2023.

Yuqing Du, Ksenia Konyushkova, Misha Denil, Akhil Raju, Jessica Landon, Felix Hill, Nando de Freitas, and Serkan Cabi. Vision-Language models as success detectors. March 2023a.

Yuqing Du, Olivia Watkins, Zihan Wang, Cédric Colas, Trevor Darrell, Pieter Abbeel, Abhishek Gupta, and Jacob Andreas. Guiding pretraining in reinforcement learning with large language models. In *Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202*, 7 2023b. URL https://arxiv.org/abs/2302.06692v2.

Carol S. Dweck. Article commentary: The study of goals in psychology. *Psychological Science*, 3 (3):165–167, 1992. doi: 10.1111/j.1467-9280.1992.tb00019.x. URL https://doi.org/10.1111/j.1467-9280.1992.tb00019.x.

Jacquelynne S Eccles and Allan Wigfield. Motivational beliefs, values, and goals. *Annu. Rev. Psychol.*, 53:109–132, 2002.

Andrew J Elliot and James W Fryer. The goal construct in psychology. In James Y Shah (ed.), *Handbook of motivation science (pp*, volume 638, pp. 235–250. The Guilford Press, xviii, New York, NY, US, 2008.

Benjamin Eysenbach, Vivek Myers, Ruslan Salakhutdinov, and Sergey Levine. Inference via interpolation: Contrastive representations provably enable planning and inference. March 2024.

Kuan Fang, Patrick Yin, Ashvin Nair, and Sergey Levine. Planning to practice: Efficient online fine-tuning by composing goals in latent space. 5 2022. doi: 10.48550/arxiv.2205.08129. URL https://arxiv.org/abs/2205.08129v1.

Ayelet Fishbach and Melissa J Ferguson. The goal construct in social psychology. In Arie W Kruglanski and E Tory Higgins (eds.), *Social psychology: Handbook of basic principles*, volume 2, pp. 490–515. The Guilford Press, xiii, New York, NY, US, 2007.

Carlos Florensa, David Held, Xinyang Geng, and Pieter Abbeel. Automatic goal generation for reinforcement learning agents. In *International conference on machine learning*, pp. 1515–1528. PMLR, 2018.

Steven M Frankland and Joshua D Greene. Concepts and compositionality: In search of the brain's language of thought. *Annu. Rev. Psychol.*, 71:273–303, January 2020.

Adele E. Goldberg. *Compositionality*, pp. 419–433. Taylor and Francis Inc., July 2015. ISBN 9780415661737.

Peter M. Gollwitzer and Gordon B. Moskowitz. Goal effects on action and cognition. In E. Tory Higgins (ed.), *Social psychology : handbook of basic principles*, pp. 361–399. Guilford, New York, 1996. ISBN 1-57230-100-7.

James A Hampton. Abstraction and context in concept representation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 358(1435):1251–1259, July 2003.

Felix Hill, Andrew Lampinen, Rosalia Schneider, Stephen Clark, Matthew Botvinick, James L McClelland, and Adam Santoro. Environmental drivers of systematicity and generalization in a situated agent. In *International Conference on Learning Representations*, 2019.

Michael E Hyland. Motivational control theory: An integrative framework. *Journal of Personality and Social Psychology*, 55(4):642, 1988.

Rodrigo Toro Icarte, Toryn Klassen, Richard Valenzano, and Sheila McIlraith. Using reward machines for high-level task specification and decomposition in reinforcement learning. *Proceedings of the 35th International Conference on Machine Learning*, 80:2107–2116, 10–15 Jul 2018a. URL https://proceedings.mlr.press/v80/icarte18a.html.

Rodrigo Toro Icarte, Toryn Q Klassen, Richard Valenzano, and Sheila A McIlraith. Teaching multiple tasks to an rl agent using ltl, 2018b. URL www.ifaamas.org.

Rodrigo Toro Icarte, Toryn Q. Klassen, Richard Valenzano, and Sheila A. McIlraith. Reward machines: Exploiting reward function structure in reinforcement learning. *Journal of Artificial Intelligence Research 73 (2022)*, 73:173–208, 10 2022. URL http://arxiv.org/abs/2010.03950.

Julian Jara-Ettinger. Theory of mind as inverse reinforcement learning. *Current Opinion in Behavioral Sciences*, 29:105–110, October 2019.

Brenden M Lake and Marco Baroni. Human-like systematic generalization through a meta-learning neural network. *Nature*, 623(7985):115–121, November 2023.

Borja G. Leon, Murray Shanahan, and Francesco Belardinelli. In a nutshell, the human asked for this: Latent goals for following temporal specifications. *ICLR 2022*, 2022. URL https://openreview.net/forum?id=rUwm9wCjURV.

Angeline S. Lillard. The development of play. In L. Liben and U. Mueller (eds.), *Handbook of Child Psychology and Developmental Science, Vol. 3: Cognitive Development*, volume 3, pp. 425–468. Wiley-Blackwell, 2015.

Michael L. Littman, Ufuk Topcu, Jie Fu, Charles Isbell, Min Wen, and James MacGlashan. Environment-independent task specifications via gltl. *arXiv*, 4 2017. URL http://arxiv.org/abs/1704.04341.

Yecheng Jason Ma, William Liang, Guanzhi Wang, De-An Huang, Osbert Bastani, Dinesh Jayaraman, Yuke Zhu, Linxi Fan, and Anima Anandkumar. Eureka: Human-Level reward design via coding large language models. October 2023.

Yecheng Jason Ma, William Liang, Hungju Wang, Sam Wang, Yuke Zhu, Linxi Fan, Osbert Bastani, and Dinesh Jayaraman. Dreureka: Language model guided sim-to-real transfer. 2024.

Gaia Molinaro and Anne G E Collins. A goal-centric outlook on learning. *Trends Cogn. Sci.*, 27 (12):1150–1164, December 2023.

Gordon B Moskowitz and Heidi Grant (eds.). *The psychology of goals*, volume 548. Guilford Press, New York, NY, US, 2009.

Gregory L Murphy. *The big book of concepts*. MIT Press, January 2004.

Ashvin V Nair, Vitchyr Pong, Murtaza Dalal, Shikhar Bahl, Steven Lin, and Sergey Levine. Visual reinforcement learning with imagined goals. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/7ec69dd44416c46745f6edd947b470cd-Paper.pdf.

Andrew Y. Ng and Stuart J. Russell. Algorithms for inverse reinforcement learning. In *Proceedings of the Seventeenth International Conference on Machine Learning*, ICML '00, pp. 663–670, San Francisco, CA, USA, 2000. Morgan Kaufmann Publishers Inc. ISBN 1558607072.

Yael Niv. Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53(3):139–154, 2009. ISSN 0022-2496. doi: https://doi.org/10.1016/j.jmp.2008.12.005. URL https://www.sciencedirect.com/science/article/pii/S0022249608001181. Special Issue: Dynamic Decision Making.

Pierre-Yves Oudeyer, Frdric Kaplan, and Verena V Hafner. Intrinsic motivation systems for autonomous mental development. *IEEE Trans. Evol. Comput.*, 11(2):265–286, April 2007.

L.A. Pervin. *Goal Concepts in Personality and Social Psychology*. Psychology Library Editions: Social Psychology. Taylor & Francis, 2015. ISBN 9781317510222. URL https://books.google.com/books?id=lIXwCQAAQBAJ.

Matthias Plappert, Marcin Andrychowicz, Alex Ray, Bob McGrew, Bowen Baker, Glenn Powell, Jonas Schneider, Josh Tobin, Maciek Chociej, Peter Welinder, Vikash Kumar, and Wojciech Zaremba. Multi-Goal reinforcement learning: Challenging robotics environments and request for research. February 2018.

Ruchit Rawal, Khalid Saifullah, Ronen Basri, David Jacobs, Gowthami Somepalli, and Tom Goldstein. CinePile: A long video question answering dataset and benchmark. May 2024.

Juan Rocamonde, Victoriano Montesinos, Elvis Nava, Ethan Perez, and David Lindner. Vision-language models are zero-shot reward models for reinforcement learning. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=N0I2RtD8je.

Stefan Schaal. Learning from demonstration. *Advances in neural information processing systems*, 9, 1996.

Rich S Sutton. The reward hypothesis. http://incompleteideas.net/rlai.cs.ualberta.ca/RLAI/rewardhypothesis.html, 2004.

T. B. Ward. Structured imagination: the role of category structure in exemplar generation. *Cognitive Psychology*, 27:1–40, 8 1994. ISSN 0010-0285. doi: 10.1006/COGP.1994.1010.

David Warde-Farley, Tom Van de Wiele, Tejas Kulkarni, Catalin Ionescu, Steven Hansen, and Volodymyr Mnih. Unsupervised control through non-parametric discriminative rewards. In *International Conference on Learning Representations*, 2018.

Eiling Yee. Abstraction and concepts: when, how, where, what and why? *Language, Cognition and Neuroscience*, 34(10):1257–1265, November 2019.

Alan Yuhas. Over 6 years and 211 spots, a british man conquers a parking lot. *The New York Times*, April 2021.