The Design and Composition of Dynamic Structural Causal Decision Processes

SEBASTIAN BENTHALL*, International Computer Science Institute, USA and New York University, USA
ALAN LUJAN, Johns Hopkins University, USA

We present two new mathematical models of decision-making agents. Our approach is motivated by the needs of modeling the economics of computing systems. These systems are composed of subsystems, and in them, cognitive resources and discounting may be endogenous to the model. Structural Causal Decision Models (SCDMs) expand on Structural Causal Influence Models. Like SCIMs, they explicitly represent the causal relationships between model variables and the payoffs of agent decisions. Additionally, agent decisions can be constrained by their causal antecedents, and SCDMs can have open root variables for which no probability distribution or structural equation is given. We show that SCDMs have a well-defined and computationally useful property of composability. Building on SCDMs, we then define a Structural Causal Decision Process (SCDP) as a recurring SCDM with a discount variable. SCDPs benefit from the useful composition properties of SCDMs. Moreover, SCDPs are strictly more expressive than POMDPs because they do not assume rational belief formation. Indeed, an SCDP can endogenously model the memory-formation process, and is thus useful for modeling resource rational agents in dynamic settings. SCDPs are also capable of modeling variable discounting, a tool used widely in social scientific modeling. We pose that SCDPs are a useful framework for policy simulation for the digitial economy, mechanism design for information systems, and digital twin modeling of cyberinfrastructure.

Computing systems today are composed of many subsystems, which can include technical components as well as natural and artificially intelligent agents. Interactions between these components and agents consist of flows of data, which is then stored and used intelligently by other actors. Realistic models of these systems must be decomposable and reflect endogenous limits on cognitive resources. This paper presents a formal framework for constructing such models. We position this as a new method for flexibly representing potentially intelligent agents in an economy[6]. Our approach is also consistent with structural causal modeling [44] approaches from computer science and statistics. We will also perform computational complexity analysis to demonstrate the value of composability to the performance of solution algorithms.

We anticipate applications of this framework to include:

- **Policy simulation.** Exploring the consequences of public policy on the digital economy.
- **System design.** Modeling data flows in dynamic systems that include intelligent agents, with parameters that can be controlled for mechanism design.
- **Digital twin modeling.** Realistic modeling that combines strategic actors with physics-informed neural networks (PINN) for scenario-based risk analysis.

The framework developed in this paper stands on much prior work. Most closely, it builds to something similar to a dynamic influence model [39]. However, we build on more recent work on structural causal games [23] as well as the modeling concerns of computational economics, and more recent advances in deep learning, to arrive at a new design.

In this paper, we will focus on the definition of composable, causal environments for a single agent. We see this as the first step. Future work will expand the definition to multi-agent environments, and address algorithms for learning the equilibrium strategies of the agents and using the models for statistical analysis.

[*] Corresponding author.

Authors' addresses: Sebastian Benthall, International Computer Science Institute, USA and New York University, USA; Alan Lujan, Johns Hopkins University, Krieger School of Arts and Sciences, USA.

0.1 Contributions

Section 1 situates this article in prior literature and recent developments in computational economics and social science, AI safety, composable world models, and variable discounting.

Section 2 introduces the Structural Causal Decision Model, a variation on Structural Causal Influence Models (see Definition 18, [23]) that supports constraints on decision variables and root variables with no probability distribution. We will show how SCDMs can be composed and decomposed, and how, in a special case of sequential decomposition, this readily improves the efficiency of solving for optimal decision rules. This is illustrated with 2-period consumption saving problems, one of which introduces habit-formation.

Section 3 expands on the SCDM to show how it can be converted into a dynamic stochastic optimization problem. The new construct, a Structural Causal Decision Process (SCDP), can also be decomposed to improve solution efficiency. This is illustrated with a dynamic model that includes both consumption and portfolio allocation.

Section 4 discusses the expressive potential of SCDPs, showing that they are more expressive than both MDPs and POMDPs. We demonstrate that in exceeding the expressiveness of POMDPs, SCDPs are able to model resource rationality by endogenizing cognitive constraints and tradeoffs of agents.

Section 5 presents an adjustment to SCDPs that enables them to represent varying discount factors. We show that an SCDP can express models from macroeconomics where the discount factor, or patience, of consumers varies as a Markov process.

In sum, this article builds up a new general formalism for representing dynamic optimization problems that is expressed in terms of Pearlian causation, has useful properties of compositionality, and can model resource rationality and variable or abstract discounting. We intend SCDMs and SCDPs to be a new computational class of models that is well-suited to social scientific modeling of complex sociotechnical systems, such as human-AI hybrid ecosystems. We present the formal framework and motivation for these models in anticipation of ongoing work that expands them to represent multi-agent systems, and develops efficient algorithms for fitting these models to data and solving them for optimal decision rules.

In tandem with the development of this mathematical framework, we have developed an open source scientific software toolkit that implements it. The implementing project is called `scikit-agent`. It's documentation[1] and source code[2] are available online. At the time of this writing, this software is pre-beta. As this work develops, we will publish a version *0.1* of the software, as well as a roadmap of anticipated features.

1 PRIOR WORK

This work aims to synthesize recent developments in several areas of social and computer science as a foundation for asking and answering new kinds of questions about computing systems.

1.1 Economics and computational social science

Agent-based modeling (ABM) has intrigued social scientists for decades. Computational tools and ideas from physics informed computational sociology and enabled early work on studying "artificial societies" [17]. ABM has since been understood to be a bridge between disciplines [5]. On the other hand, economics has typically viewed ABM with some skepticism, preferring models that are more explicitly decision-theoretic [9].

[1] https://scikit-agent.github.io/scikit-agent/

[2] https://github.com/scikit-agent/scikit-agent

However, in their recent reformulation of the relationship between ABM and economics, Axtell and Farmer [6] paint an inspiring picture of its current and future role within the field. The availability of ground truth micro-data and cheaper compute has made mathematical assumptions about aggregate or emergent level phenomena, which has been the mainstay of many earlier methods, less appealing than direct simulation of large populations of simulated agents. These simulations allow the modeler to relax strict and often unrealistic assumptions of "rational expectations" and address cases of bounded rationality. That lets modelers achieve new levels of detail and realism. Central banks stand out as an example of vital institutions that use fine-grained, multi-sector ABMs to study economic outcomes. A single ABM can address several different aspects of the economy, including interbank credit markets, financial markets like the stock market, housing markets, and even the effects of climate change [10].

The use of ABMs for policy simulation is not without challenges. When, because of model complexity, agents can only approximately optimize their behavior, this can lead to unrobust learning, which can be mitigated by careful training procedures [2]. We maintain that given a formally well-defined decision-theoretic and structural model, these challenges may be surmountable.

1.2 Computer science and AI safety

We draw on techniques from the field of artificial intelligence. While economics has been slow to take up Pearl's causal graphical modeling approach Pearl [43] , this is common in computer science. Causal modeling has been used extensively in modeling the social effects of automated systems in such fields as fairness [15, 38] and privacy [8]. More recent work has combined this causal modeling of the impact of AI systems with explicit modeling of the intelligence – or goal-orientedness – of the AI systems themselves to addres the problem of aligning AI systems with human users.

The first synthesis of Pearlian causation and game theory was done by Koller and Milch [27] with the introduction of Multi-Agent Influence Diagrams (MAIDs). This work shows that Bayesian networks can be augmented with decision and utility variables that are assigned to different agents, and that this provides a compact way to represent multi-agent games that would be intractable to model in extensive form. New formal work has refined and expanded on the MAID framework and better aligned it with structural causal modeling, introducing Structural Causal Influence Models (SCIMs) and Structural Causal Games (SCG) [22, 23]. There is a software implementation, `PyCID`, that encapsulates some of the work of this team of researchers [21].

These new frameworks have been used to model AI safety issues, such as *reward tampering* where an AI system is able to directly modify its reward function [19], proposing AI system designs that would prevent the use of proxies for protected categories in classification tasks, and prevent a recommendation system from polarizing the politics of its users [18], as well as operationalizing what it means for an AI system to be honest, deceptive [26], or manipulative [13].

Beyond its usefulness for modeling AI and sociotechnical systems, there is a strong case to be made that causal representations of the world are necessary for robustness to distributional shifts [45]. In general, causal learning and modeling is considered by many to be the gold standard of machine intelligence, and we intend our framework to adhere to this standard.

1.3 Composability

A further motive for our approach is the desire for *composability* in modeling and simulation. By this we mean that it is useful if models can be decomposed into components that can be analyzed independently, and if they can be composed into new models. This is motivated by concerns that arise in both optimization, and in simulation.

In optimization. Composability can improve the tractability of an optimization problem embedded in the decision-theoretic model. One of the attractive properties of the Multi-Agent Influence Diagram [27] paradigm is that it allows a model to be decomposed into subgames which can be more efficiently solved to discover Nash equilibrium strategy profiles. In computational economics, Lujan [35] shows how a complex consumption-saving problem can be decomposed into a multi-stage problem, which improves the computational efficiency of solving it.

Several lines of prior work have explored how to use the decomposability of the transition function of a Markov Decision Process (MDP) to improve training performance. These approaches include factoring the subgraphs of a dynamic Bayesian network [11], clustering the state space based on the transition function [37], and decomposing the problem into a hierarchy of MDP subtasks [16].

In simulation. Beyond that concern, composability is a useful property when using models and simulation to understand and design complex systems. Complex systems are defined as those that consist of multiple distinct parts which interact to produce emergent properties that cannot be easily reduced to individual behavior. So modeling these systems generally must proceed from the bottom up. Crucially, the microstructure of these components can be empirically validated separately from the emergent outcomes. Thus, compositional systems for modeling and simulation have a long history [4, 7, 25], especially for the purpose of system design [42], digital twins [51], and studying complex systems [53, 57]. Recently, AI researchers have turned their attention to composable world models [46] and environments for LLMs [55, 56] and robot testing [41]. Researchers have grounded composability of simulations in terms of causality [54] and interconnection structure [40] when simulating physical systems, suggesting the viability of a Pearlian approach that is both causal and graphical.

1.4 Resource rationality

One recent development in the modeling of cognition – whether human or artificial – is the emergence of *resource rational analysis* [34], a reformulation of earlier versions of *rational analysis* in cognitive science [3, 14]. Rational analysis attempts to provide teleological explanations of cognitive operations as being a rational action given the incentive structure of the cognitive system. It interprets cognitive behavior as being directed towards goals even when it appears at first to be suboptimal. It is used as a way of narrowing down potential hypotheses to those which have evolutionary plausibility. One of the considerations in rational analysis is the availability of cognitive resources to solve problems; limited resources are one reason why people or animals may employ heuristics to solve a complex cognitive problem.

Resource rational analysis restates the core theses of rational analysis while repositioning them on a decision-theoretic framework in which the cost of cognitive resources are explicitly represented in the utility function of the agent. Agents choose to optimize their decision rule or policy taking into account both expected results and cognitive costs. This is proposed as a useful tool for modeling realistic agents. The concept of resource rationality has been taken up by recent proposals about AI safety and alignment, because of the challenges of aligning an AI system with humans who have limited cognitive abilities and perhaps not even rational preference structures [1, 33]. One motive for the modeling framework presented in this article is that it provides a way to model dynamic optimization problems in which cognitive resources are endogenously chosen and costly.

1.5 Variable discounting

In dynamic programming and reinforcement learning settings, there is normally a constant discount factor γ or $\beta \in (0, 1)$. However, in social scientific modeling, there are many uses of variable discount factors that may be subject to exogenous or even endogenous processes. We briefly discuss this work to motivate the inclusion of variable discounting in the SCDP framework we introduce.

Time inconsistency and hyperbolic discounting. The seminal work of Strotz [49] challenged the orthodoxy of exponential discounting by showing that time-variant discount rates imply dynamically inconsistent preferences: plans that are optimal today may no longer appear optimal tomorrow, even absent new information. This insight launched a large literature in behavioral economics.

Laibson [32] introduced quasi-hyperbolic (or β-δ) discounting, which approximates hyperbolic discounting in discrete time while retaining analytical tractability. Under this specification, agents discount the immediate future more heavily than distant periods, capturing the "present bias" observed in experimental settings. A key result is that sophisticated agents with quasi-hyperbolic preferences undersave relative to the exponential benchmark. Harris and Laibson [24] extended this analysis, proving existence of equilibrium and deriving a generalized Euler equation for hyperbolic consumers.

State-dependent discounting. More recently, Stachurski and Zhang [48] generalized discrete-time infinite-horizon dynamic programming to allow the discount factor to depend on the state. Rather than requiring $\beta < 1$ uniformly, they impose that the discount factor process be strictly less than one *on average in the long run.* Under this condition, the standard optimality results are recovered: Bellman's principle of optimality holds, and both value function iteration and policy function iteration converge. Their framework accommodates recursive preferences and unbounded rewards, with applications to asset pricing and interest rate dynamics.

Quantitative macroeconomics. Variable discounting plays an important role in heterogeneous-agent macroeconomics. Krusell and Smith [30] introduced discount factor heterogeneity to match the extreme concentration of wealth observed in the data: agents with higher β accumulate more assets in the long run. Krusell and Smith [31] embedded quasi-geometric discounting in the neoclassical growth model and showed that the resulting intrapersonal game admits a continuum of equilibria, complicating both theory and computation. Krusell et al. [29] analyzed welfare and policy implications in this setting. More recently, Cao [12] proved existence of recursive equilibrium in the Krusell-Smith economy with state-dependent discount factors, extending the theoretical foundations for this class of models.

Having situated and motivated our modeling approach, we will now proceed to present its core features. We will draw examples from computational economics, employ causal modeling tools from computer science, and demonstrate how our system enables composability.

2 STRUCTURAL CAUSAL DECISION MODELS (SCDM)

We have motivated this work with reference to a broad scope of social scientific modeling. While this work is in service to that agenda of modeling multi-agent systems, in this paper we will focus on the special case of a single agent.

2.1 Example: Two-period consumption model

Before presenting the formal definition of a Structural Causal Decision Model, we illustrate the core concepts with a canonical example from economics: the Fisher two-period consumption problem [20].

Consider a consumer who lives for two periods and seeks to maximize lifetime utility:

$$\max_{c_1, c_2} u(c_1) + \beta u(c_2)$$

where c_t denotes consumption in period t, $u(\cdot)$ is a strictly increasing, strictly concave utility function, and $\beta \in (0, 1)$ is a time preference factor.

The consumer begins period 1 with resources b_1 (for "bank balances")[3] and income y_1. Resources not consumed become end-of-period assets a_1:

$$a_1 = b_1 + y_1 - c_1.$$

These assets earn a gross return $R = 1 + r$, yielding period 2 resources $b_2 = R \cdot a_1$. In the final period, the consumer receives income y_2. They then choose $c_2 < b_2 + y_2$ to maximize utility. The optimal policy at this stage is, quite trivially, to consume all available resources $c_2 = b_2 + y_2$.

At first glance, this problem appears to involve three interrelated choices: how much to consume today (c_1), how much to save (a_1), and how much to consume tomorrow (c_2). However, inspection of the structural equations shows that these apparent decisions collapse into a single degree of freedom. Once the consumer chooses c_1, the asset equation determines a_1, which in turn determines b_2. The decision rule for c_2 is very simple to derive and compute.

This observation motivates the formal framework we develop below. We distinguish between *state variables* that describe the system at a point in time (b_1, b_2), *decision variables* under the agent's control (c_1, c_2), and *utility variables* that enter the objective ($u(c_1), u(c_2)$). These are linked by *structural equations* (deterministic relationships such as $a_1 = b_1 + y_1 - c_1$) and subject to *constraints* on feasible decisions ($0 \leq c_1 \leq b_1 + y_1; 0 \leq c_2 \leq b_2 + y_2$).

The structure also exhibits natural decomposition, illustrated in Figure 1. The variable a_1 serves as a "bridge" connecting two stages: period 1, where the agent chooses c_1 given b_1, and period 2, where c_2 is determined given b_2. This decomposition has computational significance: because c_2 is fully determined by the bridge variable, the second-stage "problem" can be solved first, yielding a continuation value function $v_2(b_2)$; this value function then informs the first-stage optimization. The sequential structure reduces a joint optimization over (c_1, c_2) to a sequence of simpler problems, foreshadowing the decomposition techniques we develop in later sections.

2.2 Structural Causal Decision Models

This section introduces a mathematical construct, which we will call a Structural Causal Decision Model (SCDM). This is closely akin to a different structure, the Structural Causal Influence Model of Hammond et al. [23], Koller and Milch [27], with a few key differences. (SCIMs are defined in Definition 18 in the Appendix.) Like the SCIM, it builds on both the prior work on influence models [47], which represent decision-theoretic problems with a graph, and Pearlian graphical causal modeling [43]. And SCDM augments an SCIM by addign explicit constraints on the optimization problem, as well as open 'inputs' to the model, which we will call 'root variables'.

We will use $\mathbf{Pa}_V$ to denote the parents of a variable V, given a graph; $\mathbf{pa}_V$ is a potential value of the parents of that variable. "dom" refers to the function that returns the domain (the set of all possible values) of a variable.

Definition 1 (Structural Causal Decision Model (SCDM)). A *structural equation influence model* $(\mathbf{V}, \mathbf{Z}, \mathcal{E}, (\mathbf{X}, \mathbf{D}, \mathbf{U}), \mathbf{Pr}, \mathbf{f}, \boldsymbol{\theta})$ consists of:

- A set of endogenous variables $\mathbf{V}$
- A set of exogenous variables $\mathbf{Z}$.
- An edge set $\mathcal{E}$ such that the graph $\mathcal{G} = (\mathbf{V} \cup \mathbf{Z}, \mathcal{E})$ is a DAG over $\mathbf{V} \cup \mathbf{Z}$, in which $\forall Z \in \mathbf{Z}, \mathbf{Pa}_Z = \varnothing$.
- $\mathbf{V}$ is partitioned into:

[3] We use b_t here to align with standard notation in the consumption literature; later sections use w_t for wealth when discussing more general dynamic models.

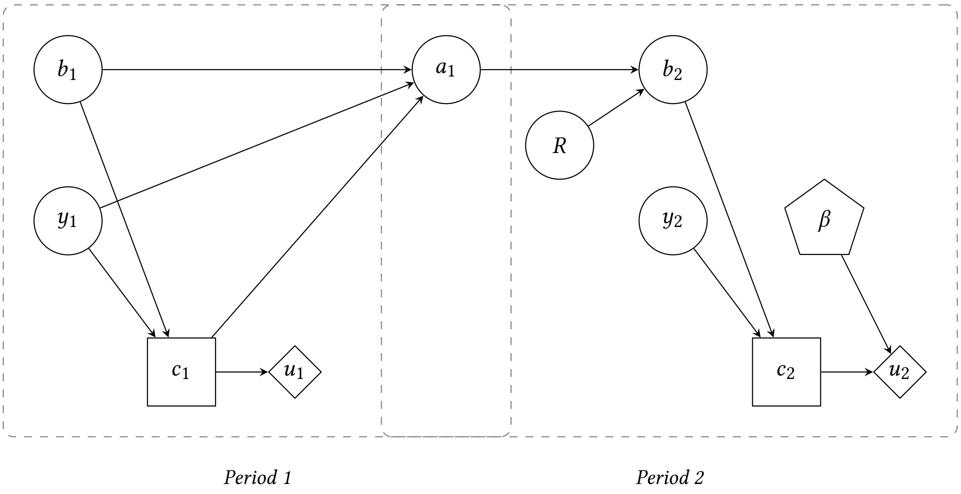

Fig. 1. Influence diagram for the two-period consumption problem. Circles denote state variables, rectangles denote decision variables, diamonds denote utility, pentagons denote discount factors. The bridge variable a_1 (end-of-period savings) connects the two periods; the boxes overlap on this node to indicate it belongs to both components. The edge from β to u_2 indicates discounting of future utility.

  - **X**, state variables
  - **D**, decision variables[4]
  - **U**, utility variables
- A parameterized probability distribution $\mathbf{Pr}(\boldsymbol{\theta}_E)$ over all exogenous variables **Z**.
- $\mathbf{f} = \{f_V : \mathrm{dom}(\mathbf{Pa}_V) \times \mathrm{dom}(\boldsymbol{\theta}) \to \mathrm{dom}(V) | V \in \mathbf{V} \setminus \mathbf{D} \text{ and } \mathbf{Pa}_V \neq \emptyset\}$ are structural equations governing all variables in **V** that are not decision nodes or parentless (root) nodes.
- $\boldsymbol{\Gamma} = \bigcup_{D \in \mathbf{D}} \{\Gamma_D\}$ where $\Gamma_D : \mathrm{dom}(\mathbf{Pa}_D) \times \mathrm{dom}(\boldsymbol{\theta}) \to \mathbb{P}(\mathrm{dom}(D))$ are constraints on the actions allowable at the decision variables.
- Parameters $\boldsymbol{\theta}$

SCDM is a departure from SCIM (Definition 18, Hammond et al. [23]) in just a few ways:

Constraints. We have introduced the decision constraints **Γ**, which are not included in the original definition. Introducing these constraints makes it easier to synthesize the SCIM and SCG constructs with the binding constraints of many control theory problems. An example constraint is the budget constraint ($c_1 \leq b_1$) in the consumption saving problem.

Structural functions. Rather than expressing the relationships between variables as "deterministic conditional probability distributions", we define deterministic functions **f** for each endogenous non-decision variable. Conversion to a conditional probability distribution is straightforward, and so this is largely a matter of style. However, unlike an SCIM, it is possible to have a state variable over which there is no *unconditional* probability distribution or governing structural equation. These are the variables V without parents.

[4] The influence diagram literature rarely intersects with the control theory literature. Decision variables and control variables are roughly synonymous.

Limiting noise. In a classic Pearlian structural causal model, each deterministic variable has a noise term ϵ_V. While SCDMs can express such a thing, as a matter of style we indicate each exogenous noise variable separately.

Continuous valued models. While both SCIMs and SCDMs are general with respect to whether variables are continuous or discrete, we have designed SCDMs to be well-suited to models with continuous state, shock, and decision values. The structural functions $f \in \mathbf{f}$ may be differentiable or invertible, which opens up more efficient solution methods [35]

We will not ground this form of modeling in measure theory in this paper. We will assume that utility variables $\mathbf{U}$ range over the real numbers for the purposes of presentation, though the structure is more general than this.

2.2.1 *Roots and leaves.*

We allow an SCDM to have state variables X with no parents, $\mathbf{Pa}_X = \varnothing$ (this is a departure from SCIM models). There are also variables with no descendants.

Definition 2 (Roots and leaves)**.** Given an SCDM $\mathcal{M} = (\mathbf{V}, \mathbf{Z}, \mathcal{E}, (\mathbf{X}, \mathbf{D}, \mathbf{U}), \mathbf{Pr}, \mathbf{f}, \boldsymbol{\theta})$, let the *roots* of the SCDM be $\tilde{\mathbf{X}} = \{X | X \in \mathbf{X} \text{ and } \mathbf{Pa}_X = \varnothing\}$, i.e., the chance variables with no parents. Let the *leaves* of the SCDM be the variables $V \in \mathbf{V}$ with no descendants.

Note that while exogenous shock variables in $\mathbf{Z}$ have no parents, they are not considered to be root variables in this sense of the term. For an SCDM, root variables will have no governing structural equation or probability distribution.

The SCDM can be seen as a stochastic function from root values $\tilde{\mathbf{x}} \in \text{dom}(\tilde{\mathbf{X}})$ to the values of all other variables, including the utility variables. The root nodes begin the flow of causation through the model. At the decision variables, an agent chooses a rule that governs the flow of values.

2.2.2 *Decisions.*

SCDMs are in the family of *influence models* because they define a decision-theoretic problem for the agent.

Definition 3 (Decision rule)**.** Given an SCDM $(\mathbf{V}, \mathbf{Z}, \mathcal{E}, (\mathbf{X}, \mathbf{D}, \mathbf{U}), \mathbf{Pr}, \mathbf{f}, \boldsymbol{\theta})$, a decision rule π_D for a decision variable $D \in \mathbf{D}$ is a function $\pi_D(\mathbf{Pa_D})$ of the form $\pi_D : \text{dom}(\mathbf{Pa}_D) \rightarrow \text{dom}(D)$, such that $\forall \mathbf{p} \in \text{dom}(\mathbf{Pa}_D), \pi_D(\mathbf{p}) \in \Gamma_D(\mathbf{p})$.[5] A set of decision rules π for all decision variables $\mathbf{D}$ is called a *policy profile.*[6]

A policy profile and values for the root variables provide all that is needed to *induce* an SCDM into a fully specified Structural Causal Model (see Definition 16), which characterizes the joint distribution over all model variables. This is done by turning each decision variable D into a state variable governed by the corresponding decision rule π_D, and assigning values to the root variables.

For any variable $V \in \mathbf{V}$, we can construct the deterministic function $f^*_V(\tilde{\mathbf{x}}, \mathbf{z}, \pi)$ for the value of that variable given the realization of root variables, shocks, and decision rules.

Under usual rationality assumptions, the agent's rational behavior is to choose the decision rules that maximize expected utility. We can define the *root value function* of the SCDM to be the maximum possible expected utility, given root values.

Definition 4 (Root value function of an SCDM)**.** We define the *root value function* of an SCDM $\mathcal{M} = (\mathbf{V}, \mathbf{Z}, \mathcal{E}, (\mathbf{X}, \mathbf{D}, \mathbf{U}), \mathbf{Pr}, \mathbf{f}, \boldsymbol{\theta})$ to be:

[5] In other words, the decisions according to the rule must fall within the decision constraints.

[6] Mixed strategy decision rules can be introduced into SCDMs by including an exogenous noise variable as a parent of the decision nodes.

$$v_{\mathcal{M}}(\tilde{\mathbf{x}}) = \max_{\pi} E_{\mathbf{Z}} \left[\sum_{U \in \mathbf{U}} f^{*}_{U}(\tilde{\mathbf{X}} = \tilde{\mathbf{x}}, \mathbf{z}, \boldsymbol{\pi}) \right] \tag{1}$$

The function $v_{\mathcal{M}} : \mathrm{dom}(\tilde{\mathbf{X}}) \to \mathbb{R}$ is over the values of the root variables of $\mathcal{M}$ and reflects the optimal expected value of an agent subject conditioned on the given values of the root variables. Note that the choice of policy profile π occurs, logically, before the realization of the shocks.

Note that this is an optimization over decision rules, not an optimization over "actions", or values of decision variables. This is a significant distinction, because the space of possible decision rules $\boldsymbol{\pi}$ is not the same as the space of possible functions from root values to decisions ($\mathrm{dom}(\tilde{\mathbf{X}}) \to \mathrm{dom}(\mathbf{D})$). We draw a distinction between the optimal decision rules – which are subject to the limited information available to the agent – and the optimal actions which would be taken by an omniscient agent under the circumstances of the root values:

$$\mathbf{d}^{*}_{\mathcal{M}}(\tilde{\mathbf{x}}) = \arg \max_{\mathbf{d} \in \mathrm{dom}(\mathbf{D})} E_{\mathbf{Z}} \left[\sum_{U \in \mathbf{U}} f^{*}_{U}(\tilde{\mathbf{X}} = \tilde{\mathbf{x}}, \mathbf{z}, \mathbf{D} = \mathbf{d}) \right] \tag{2}$$

SCDMs precisely define the information available at each decision node and for a given model, it is possible that a policy profile that guarantees optimal decision values does not exist.

2.2.3 Composition. We have discussed the motivations for model composition in Section 1.3. Composability enables the reuse and independent empirical validation of model components. We also anticipate ways in which composability can be leveraged to improve learning algorithms. Sometimes composability can be leveraged for computational performance.

In this paper, we will focus on a special case of composition: cases where an SCDM can be sequentially decomposed, meaning that it can be decomposed into two components, such that one follows 'after' the other. This has the computational benefit that the problem of selecting the optimal policy profile can be divided into nested subgames such that the last subgame can be solved first, and so on, through backwards induction.

The overall idea of composing SCDMs is quite simple. The DAG of an SCDM can be partitioned into subgraphs, each defining its own SCDM. Equivalently, this means that two SCDMs can be composed into a new, combined, model. We use the fact that SCDMs have variables for which no values or probability distributions are assigned – the roots $\tilde{\mathbf{X}}$. These variables are the points where the two components can 'attach'.

Definition 5 (SCDM Composition). Consider two SCDMs

$$\mathcal{M}_1 = (\mathbf{V}_1, \mathbf{Z}_1, \mathcal{E}_1, (\mathbf{X}_1, \mathbf{D}_1, \mathbf{U}_1), \mathbf{Pr}_1, \mathbf{f}_1, \boldsymbol{\theta}_1)$$

$$\mathcal{M}_2 = (\mathbf{V}_2, \mathbf{Z}_2, \mathcal{E}_2, (\mathbf{X}_2, \mathbf{D}_2, \mathbf{U}_2), \mathbf{Pr}_2, \mathbf{f}_2, \boldsymbol{\theta}_2)$$

such that $\mathbf{V}_1 \cap \mathbf{V}_2 = \mathbf{Y}_1 \subseteq \tilde{\mathbf{X}}_2$. Then we define the *composition* of the two models $\mathcal{M}_1 \circ \mathcal{M}_2$ as the tuple:

- $\mathbf{V}_1 \cup \mathbf{V}_2$
- $\mathbf{Z}_1 \cup \mathbf{Z}_2$
- $\mathcal{E}_1 \cup \mathcal{E}_2$
- $(\mathbf{X}_1 \cup \mathbf{X}_2, \mathbf{D}_1 \cup \mathbf{D}_2, \mathbf{U}_1 \cup \mathbf{U}_2)$
- $\mathbf{Pr}_1 \cdot \mathbf{Pr}_2$
- $\mathbf{f}_1 \cup \mathbf{f}_2$
- $\boldsymbol{\theta}_1 \cup \boldsymbol{\theta}_2$

We will call $\mathbf{Y} = \mathbf{V}_1 \cap \mathbf{V}_2$ the *bridge* of $\mathcal{M}_1 \circ \mathcal{M}_2$. We impose the restriction that for any $Y \in \mathbf{Y}$, then Y is a root variable for either $\mathcal{M}_1$ or $\mathcal{M}_2$.

A composition of two SCDMs is an SCDM. A *decomposition* of an SCDM is a partitioning of an SCDM into two, such that $\mathcal{M}_0 = \mathcal{M}_1 \circ \mathcal{M}_2$. We will refer to this as $\mathbf{Y}$, dropping the subscript, when context makes this unambiguous. Since bridge variables belong to both components by definition, they can be labeled by either period; in diagrams, we place each variable in the box corresponding to its subscript.

This is a general definition of compositions which does not rely on any assumptions about the model structure. For example, a model can be a composition of two other models that are entirely disjoint. In the next section, we will focus on a special case of model composition that we will call sequential decomposition.

2.3 Solving sequentially decomposed SCDMs

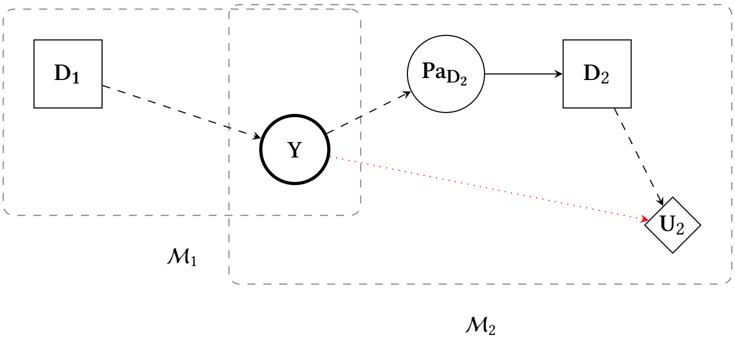

Fig. 2. A model $\mathcal{M}$ is composed of $\mathcal{M}_1 \circ \mathcal{M}_2$. Indirect paths are represented by dotted edges. If all paths from the bridge nodes $\mathbf{Y}$ to reward nodes in the second component $\mathbf{U_2}$ have a member of $\mathbf{Pa}_{D_2} \cup \mathbf{D_2}$ on it, then $\mathbf{Y}$ is d-separated from $\mathbf{U_2}$ given those nodes. Under that condition the composition is orthomodular or, equivalently, sequential. An indirect path from Y to U_2 which is not interrupted by $\mathbf{Pa}_{D_2} \cup \mathbf{D_2}$ (shown in red) breaks this orthomodularity condition.

In this section, we discuss solving a decomposed SCIM. In prior work on influence diagrams, there is a distinction drawn between *perfect recall*, when each decision node has direct information about all past observations and decisions, and *sufficient recall*, when the agent has enough access to prior observations and decisions to make an optimal choice [50]. A related notion is that of *strategic reliance* (see Definition 19). Roughly, if D_2 does not strategically rely on D_1, then the choice of optimal decision rule for D_2 does not depend on the choice of optimal decision for D_1. Koller and Milch [27] show that identifying the network of strategic reliance (see Definition 19) between decision nodes in an influence diagram allows the analyst to divide the problem of finding the optimal policy profile into efficient subgames. In this section, we will consider a special case of sequential decomposition and how it affords the breaking down of the model's solution problem into smaller subproblems.

Consider $\mathcal{M}_0 = \mathcal{M}_1 \circ \mathcal{M}_2$ In order to get computational power out of decomposing an SCIM, the decomposition has to separate the decision variables such that $\mathbf{D_2}$ does not strategically rely on $\mathbf{D_1}$. Under these conditions, the optimal decision rules for $\mathbf{D_2}$ will not depend on the decision rules for $\mathbf{D_1}$, and as we will see this enables the problem to be solved through a simple backwards induction procedure.

We can guarantee this by imposing a graphical criterion on the decomposition. It relies on the definition of d-separation (see Definition 21).

Definition 6 (Orthomodularity). Given $\mathcal{M}_0 = \mathcal{M}_1 \circ \mathcal{M}_2$ with bridge $\mathbf{Y}$, the decomposition is *orthomodular* if and only if $\mathbf{Y}$ and $\mathbf{U_2}$ are d-separated given $\mathbf{D_2} \cup \mathbf{Pa_{D_2}}$.

Note that we are presenting orthomodularity as a sufficient, but not necessary, condition for the separability of component subgames of an SCDM.

Theorem 7. Given an orthomodular decomposition $\mathcal{M}_0 = \mathcal{M}_1 \circ \mathcal{M}_2$, then there is no $D_2 \in \mathbf{D_2}$ that strategically relies on $D_1 \in \mathbf{D_1}$.

Proof. By Definition 5, all paths between $\mathbf{D_1}$ and $\mathbf{U_{D_2}}$ must include at least one bridge node Y, and $\mathbf{D_1} \cap \mathbf{V_2} = \emptyset$.

All paths from $\hat{\mathbf{D_1}}$, added parents of $\mathbf{D_1}$, to $\mathbf{U_{D_2}}$ will include a bridge node $Y \in \mathbf{Y}$.

By Definition 6, $\mathbf{Y}$ and U_2 are d-separated given $\mathbf{D_2} \cup \mathbf{Pa_{D_2}}$.

Therefore there are no active paths from $\mathbf{D_1}$ to U_2 given $\mathbf{D_2} \cup \mathbf{Pa_{D_2}}$.

By Definition 20, $\mathbf{D_1}$ is not s-reachable from $\mathbf{D_2}$.

By Theorem 22, $\mathbf{D_2}$ does not strategically rely on $\mathbf{D_1}$. □

We now revisit the problem of solving for the optimal policy profile of an SCIM given the tool of sequential decomposition.

Definition 8 (Bridge value function). Given an orthomodular decomposition $\mathcal{M}_0 = \mathcal{M}_1 \circ \mathcal{M}_2$, the *bridge value function* of $\mathcal{M}_2$ is:

$$v_{\mathcal{M}_2}(\mathbf{y}) = \max_{\boldsymbol{\pi}_2} E_{\mathbf{Z}} \left[\sum_{U \in \mathbf{U}_2} f^*_U(\mathbf{y}, \mathbf{z}, \boldsymbol{\pi}_2) \right] \tag{3}$$

By Theorem 7, we know that in the optimal profile policy for the composed model $\mathcal{M}_0$, the optimal policy profile for π_2 does not depend on the choice of π_1. This means that we can use the bridge value function as a *continuation value function* in the solution for the total model.

$$\begin{aligned} v_{\mathcal{M}_0}(\tilde{\mathbf{x}}) &= \max_{\boldsymbol{\pi}_0} E_{\mathbf{Z}_0} \left[\sum_{U \in \mathbf{U_0}} f^*_U(\tilde{\mathbf{x}}, \mathbf{z_0}, \boldsymbol{\pi}) \right] \\ &= \max_{\boldsymbol{\pi}_1} E_{\mathbf{Z}_1} \left[\sum_{U_1 \in \mathbf{U_1}} f^*_{U_1}(\tilde{\mathbf{x}}, \mathbf{z_1}, \boldsymbol{\pi_1}) + \max_{\boldsymbol{\pi}_2} E_{\mathbf{Z}_2} \left(\sum_{U_2 \in \mathbf{U}_2} f^*_{U_2}(f_{\mathbf{y}}(\tilde{\mathbf{x}}, \mathbf{z_1}, \boldsymbol{\pi_1}), \mathbf{z}_2, \boldsymbol{\pi_2} \right) \right] \\ &= \max_{\boldsymbol{\pi}_1} E_{\mathbf{Z}_1} \left[\sum_{U_1 \in \mathbf{U_1}} f^*_{U_1}(\tilde{\mathbf{x}}, \mathbf{z_1}, \boldsymbol{\pi_1}) + v_{\mathcal{M}_2}(f_{\mathbf{y}}(\tilde{\mathbf{x}}, \mathbf{z_1}, \boldsymbol{\pi_1}), \boldsymbol{\pi_1}) \right] \end{aligned} \tag{4}$$

2.4 Computational benefits of decomposition

Consider the problem of finding the optimal policy profile π_0 for model $\mathcal{M}_0 = \mathcal{M}_1 \circ \mathcal{M}_2$, with orthomodular decomposition. Let $|\pi_i|$ be the number of possible total policy profiles in model $\mathcal{M}_i$. $|\pi_0| = |\pi_1||\pi_2|$. Searching for the optimal π_0 in the most naive way involves a search over $|\pi_0|$ possibilities that is $O(|\pi_0|)$.

Because the model is sequential, the subgame $\mathcal{M}_2$ can be solved for every case of the bridge $\mathbf{Y}$ with a worst-case search. Once solved, the bridge value function can be solved in constant time $O(1)$ (assuming the structural equations $\mathbf{f}$ are all $O(1)$). Thus, solving the decomposed model in sequence can be done in $O(|\pi_2||Y| + |\pi_1|)$. This is a significant improvement.

2.5 Example: Two-period consumption with habit formation

Having seen how the standard two-period consumption problem decomposes sequentially, we now consider a variation that *requires an expanded bridge*: habit formation. This example demonstrates how certain structural features necessitate richer sufficient statistics for decomposition.

Consider a consumer who lives for two periods and seeks to maximize lifetime utility:

$$\max_{c_1,c_2} u(c_1, h_1) + \beta u(c_2, h_2),$$

where c_t denotes consumption in period t, h_t denotes the habit stock in period t, $u(\cdot, \cdot)$ is a utility function that depends on both consumption and habits, and $\beta \in (0, 1)$ is a time preference factor.

The term “habit formation” here refers not to behavioral routines but to preferences in which past consumption affects current utility. Utility is decreasing in the habit stock (that is, $\partial u/\partial h < 0$): higher past consumption raises the reference point against which current consumption is compared, making any given level of current consumption less satisfying.

The budget dynamics are identical to the standard problem:

$$a_1 = b_1 + y_1 - c_1,$$

$$b_2 = R \cdot a_1,$$

$$c_2 = b_2 + y_2.$$

However, we now add an equation describing how habits evolve. In the simplest formulation:

$$h_2 = c_1.$$

That is, the habit stock in period 2 equals consumption in period 1. We take the initial habit h_1 as an exogenous parameter reflecting the consumer’s consumption history prior to period 1; it enters the period 1 utility function and, as we will see, becomes part of the state space for the optimization problem.

Why the bridge must expand. The critical difference from the standard model lies in how information flows between periods. In the standard problem, period 2 utility $u(c_2)$ depends only on c_2, which is determined by the bridge variable a_1 (equivalently b_2) and period 2 income y_2. The single bridge variable a_1 serves as a sufficient statistic: it captures all the information from period 1 that is relevant for period 2.

With habit formation, period 2 utility $u(c_2, h_2)$ depends on both c_2 and $h_2 = c_1$. The consumption choice c_1 now affects period 2 utility through two distinct channels: (i) indirectly through $a_1 \longrightarrow b_2 \longrightarrow c_2$, and (ii) directly through $h_2 = c_1$. This means that a_1 alone is no longer a sufficient statistic; period 2 outcomes depend on information from period 1 that is not captured by a_1.

The solution is to *expand the bridge* to include both channels. If we take $\mathrm{Y} = \{a_1, c_1\}$ as the bridge (or equivalently $\{b_2, h_2\}$ on the period 2 side), then all connections between periods pass through the bridge. With this expanded bridge, c_2 becomes a genuine decision variable: the consumer in period 2 observes both their wealth b_2 and their habit stock h_2, and chooses consumption accordingly.

Under this formulation, the orthomodularity condition of Definition 6 is satisfied. With c_2 as a decision variable depending on h_2, the path $c_1 \longrightarrow h_2 \longrightarrow u_2$ is blocked: h_2 is now a parent of the decision c_2, so conditioning on $\mathbf{D}_2 \cup \mathbf{Pa}_{\mathbf{D}_2} = \{c_2, b_2, y_2, h_2\}$ includes h_2, which blocks the path. The expanded bridge $\{b_2, h_2\}$ and $\mathbf{U}_2$ are d-separated given the period 2 decision and its parents, restoring the backwards induction strategy of Theorem 7.

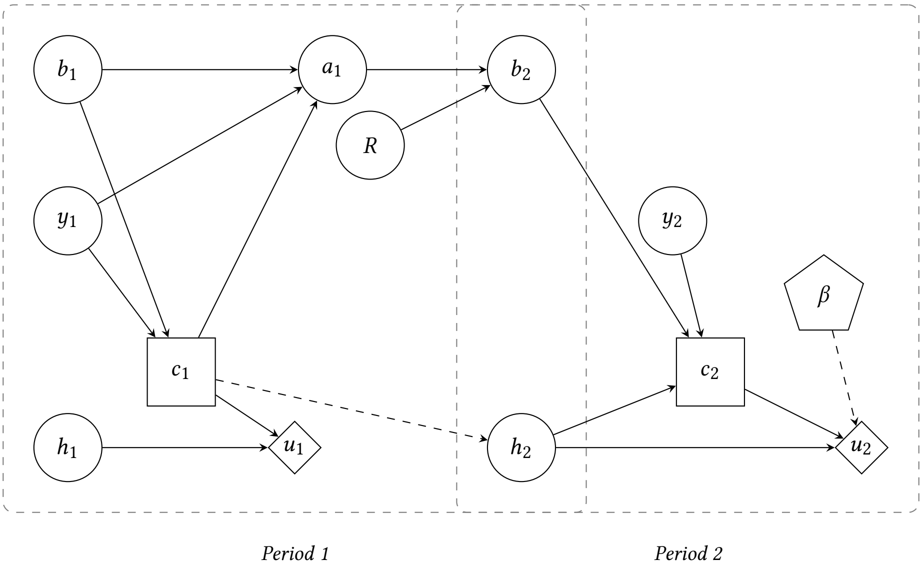

Fig. 3. Influence diagram for the two-period consumption problem with habit formation. The dashed arrow from c_1 to h_2 represents the habit formation mechanism: period 1 consumption directly affects period 2 habits. The expanded bridge $\{b_2, h_2\}$ connects the two periods; the boxes overlap on these nodes to indicate they belong to both components. With this expanded bridge, c_2 becomes a genuine decision variable (rectangle) that depends on both wealth b_2 and habits h_2. Pentagons denote discount factors.

Implications for solution methods. The expanded bridge has practical consequences for solution methods. In the standard problem, we evaluate period 2 for each possible value of the single bridge variable a_1, obtaining a one-dimensional continuation value function $v_2(b_2)$. With habits, the continuation value function is two-dimensional, defined over the expanded bridge:

$$v_2(b_2, h_2) = \max_{c_2} u(c_2, h_2) \quad \text{subject to } c_2 \leq b_2 + y_2.$$

The consumer in period 2 now makes a genuine choice: given wealth b_2 and habit stock h_2, they choose c_2 to maximize utility. This continuation value function can be computed independently of period 1, and then used in the period 1 problem:

$$v_1(b_1, h_1) = \max_{c_1} \left\{ u(c_1, h_1) + \beta\, v_2\big(R(b_1 + y_1 - c_1), c_1\big) \right\}. \tag{5}$$

The first argument of v_2 is the wealth channel (via $a_1 \longrightarrow b_2$); the second is the habit channel (via $c_1 \longrightarrow h_2$). The problem decomposes sequentially, but over a higher-dimensional bridge than the standard model.

General lessons. This example illustrates a broader point about model structure and decomposability. The standard consumption problem decomposes with a single bridge variable because asset position is a sufficient statistic for all period 1 information relevant to period 2. Habit formation requires expanding the bridge: past consumption matters for future utility beyond its effect on wealth, so both a_1 and c_1 (or equivalently b_2 and h_2) must be included. More generally, when past decisions affect future outcomes through multiple channels, the bridge must expand to capture all relevant sufficient statistics. The framework accommodates such cases: decomposition does not fail, but the dimensionality of the bridge increases.

3 STRUCTURAL CAUSAL DECISION PROCESSES (SCDP)

We now build on the prior definitions of SCDMs and composition to develop a new form of composable model for a dynamic decision process, an SCDP. Like the SCDMs, a decomposable SCDP can be solved more efficiently through separation into subgames.

3.1 Example: Consumer choice with portfolio allocation

We consider two components of a dynamic stochastic optimization problem. A consumer earns and consumes, anticipating the availability of resources for future consumption. They also make a portfolio allocation decision which exposes their savings to a more or less risky rate of return. This model is diagramed in Figure 4.

Consumption-saving problem. Consider a simple consumption-saving problem with borrowing constraint [36]. The consumer aims to maximize, through choice of their level of consumption c_t:

$$\max_{\{c_t\}} E\left[\sum_{t=0}^{\infty} \beta^t u(c_t)\right]$$

subject to the following transition equations and constraints:

$$\begin{aligned} y &= \mu_y + \sigma_y \epsilon_y;\ \epsilon_y \sim \mathcal{N}(0,1) && (6)\\ m &= w + e^y && (7)\\ 0 \leq c(n) &\leq m && (8)\\ u &= \ln c && (9)\\ a &= m - c && (10)\\ w' &= ra && (11) \end{aligned}$$

where w is the level of wealth, r is a rate of return on savings, and y is an income shock, and u is utility from consumption.

Portfolio allocation. We can make the consumption-saving problem more complex by introducing a risky return and portfolio allocation choice α. Whereas in the earlier problem the rate of return r was given, now it is an endogenously scaled mixture of a constant r_f and a stochastic shock r_r. The consumer chooses α, the share of their wealth to invest in the risky asset.

$$\begin{aligned} 0 \leq \alpha(a) &\leq 1 && (12)\\ r_r &= e^{\mu_r + \sigma_r \epsilon_r};\ \epsilon_r \sim \mathcal{N}(0,1) && (13)\\ r &= (1-\alpha) r_f + \alpha r_r && (14) \end{aligned}$$

A critical piece of the interpretation of these equations is that the portfolio allocation choice α is made with knowledge of the post-consumption wealth a but without the information of the realization of r_r.

Composed model. . The composed model is shown in Figure 4. The variables a and r are shared by both models and thus are the *bridge*. Because the allocation component has no utility nodes, the composition is trivially orthomodular. The single period problem can be solved first by solving for α given the continuation value of w', and then by solving for optimal c.

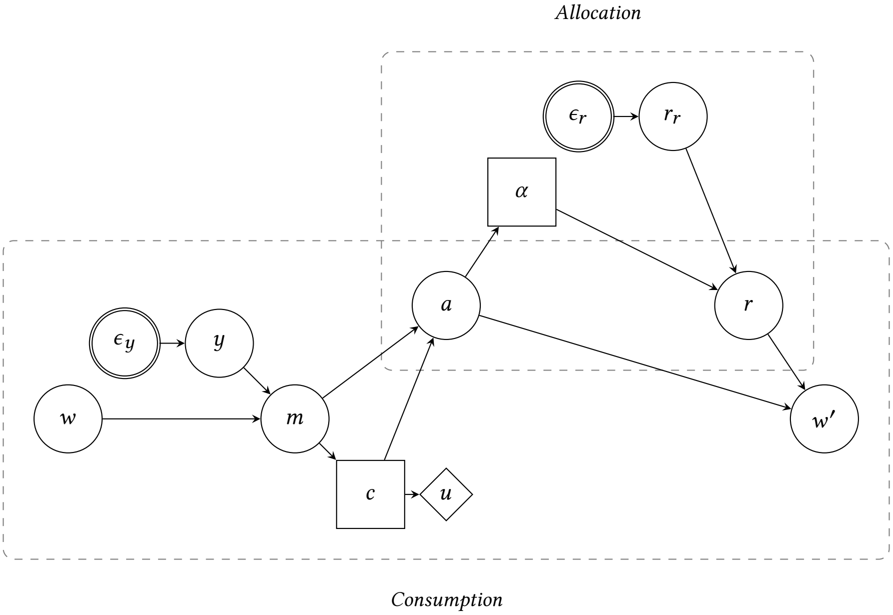

Fig. 4. The composed consumption and portfolio allocation dynamic problem.

3.2 Dynamic Optimization

We begin with a simple definition of a dynamic stochastic optimization problem, drawn from [36].

Definition 9 (Dynamic stochastic optimization problem (DSOP))**.** Given discrete time steps $t \in 0, 1, ..., T$ with T potentially infinite, and given exogenous shocks $z = \{z_t \sim P_{Z_t}\}_{t=0}^{T}$ and endogenous state $x = \{x_t\}_{t=0}^{T}$ and actions $a = \{a_t\}_{t=0}^{T}$ such that:

- $a_t \in A_t(x_t, z_t)$
- $x_{t+1} = g_t(x_t, z_t, a_t)$

the agent maximizes expected lifetime value

$$v_0 = \max_{a,x} E_z \left[\sum_{t=0}^{T-1} \beta^t r(x_t, z_t, a_t) \right] \tag{15}$$

where $\beta \in [0, 1)$ is the discount factor.

As is well-known, this sort of optimization problem can be represented by a recursive Bellman equation. We will use a somewhat adjusted form.

Definition 10 (Shifted Bellman Equations)**.** The Shifted Bellman Equation is for $v_t(x_t)$. We let $v_T(x_t) = 0$. For $t < T$,

$$v_t(x_t) = E_{z_t} \left[\max_{a_t \in A_t(x_t, z_t)} \{ r_t(x_t, z_t, a_t) + \beta_t v_{t+1}(x_{t+1}) \} \right] \tag{16}$$

where $x_{t+1} = g_t(x_t, z_t, a_t)$.

The benefit of Equation 16 is that it formulates the “solution” of a single time period t in terms of a small number of self-contained elements. We will give this bundle of elements a name – a period block, or P-block.

Definition 11 (P-block)**.** A P-block is a tuple $\mathcal{B}_t^P = (\mathcal{X}, \mathcal{Z}, P_{\mathcal{Z}}, \mathcal{A}, A, g, r, \beta_t)$:

- A state space $\mathcal{X}$.
- A space of exogenous shocks $\mathcal{Z}$.
- A probability distribution over shocks $P_{\mathcal{Z}}$.
- A space of actions $\mathcal{A}$
- A set of action constraints $A : \mathcal{X} \times \mathcal{Z} \to \mathbb{P}(\mathcal{A})$
- A deterministic transition function $g : \mathcal{X} \times \mathcal{Z} \times \mathcal{A} \to \mathcal{X}$
- A reward function $r : \mathcal{X} \times \mathcal{Z} \times \mathcal{A} \to \mathbb{R}$
- A discount function $\beta : \mathcal{X} \times \mathcal{Z} \times \mathcal{A} \to \mathbb{R}$, which is most often valued as a constant.

We have allowed a functional or variable discount factor here. We will explore what the implications of that are in Section 5. For purposes of presentation here, we will continue to show β as a constant.

Lemma 12. *Given, for $t = \{0, 1, ..., T-1\}$, P-Block $\mathcal{B}_t^P$ then the optimization problem in shifted Bellman* form*

$$v_t(x_t) = E_{z_t}\left[\max_{a_t \in A_t(x_t, z_t)} \{r_t(x_t, z_t, a_t) + \beta_t v_t(g(x_t, z_t, a_t))\}\right] \tag{17}$$

is a DSOP.

Proof. Equation 17 implies Equation 16, which is an alternative form for a DSOP. □

3.2.1 Equivalence of SCIMs and P-blocks. We have defined two mathematical objects, a P-block (11) which represents a single 'period' of a DSOP (9), and an SCDM (1). While these two constructs come from different intellectual fields, it is simple to produce a P-block from an SCDM.

Theorem 13. Given

- an SCDM $\mathcal{M} = (\mathbf{V}, \mathbf{Z}, \mathcal{E}, (\mathbf{X}, \mathbf{D}, \mathbf{U}), \mathbf{Pr}, \mathbf{f}, \boldsymbol{\theta})$
- a discount variable $\beta \in \mathbf{V}$
- a mapping from root variables to end-of-period-variables $w : \tilde{\mathbf{X}} \to \mathbf{V}$ such that $\forall \tilde{X} \in \tilde{\mathbf{X}}, \text{dom}(w(\tilde{X})) = \text{dom}(\tilde{X}))$ and $\mathbf{W} = \{w(\tilde{X}) | \tilde{X} \in \tilde{\mathbf{X}}\}$

there is a corresponding P-block $\mathcal{B}_t^P = (\mathcal{X}, \mathcal{Z}, P_{\mathcal{Z}}, \mathcal{A}, A, g, r, \beta_t)$.

Proof. By construction.
Let $\mathcal{X} \subseteq \tilde{\mathbf{X}}$, the roots.
Let $\mathcal{Z} = \mathbf{Z}$.
Let $P_{\mathcal{Z}} = Pr(\mathbf{Z}, \theta)$
Let $\mathcal{A} = \mathbf{D}$.
Let $A = \Gamma$.
Let $g(x, z, a) = \mathbf{f}_{\mathbf{W}}(x, z, a)$, the structural functions evaluated at the end-of-state variables $\mathbf{W}$.
Let $r(x, z, a) = \sum_{U \in \mathbf{U}} f_U^*(x, z, a)$, the sum of utility functions.
Let $\beta = f_\beta^*$. □

Thus, an SCIM can be made 'dynamic' by adding a discount factor and by making an explicit mapping from the available model variables to the state variables for the next period. We now have a method of creating dynamic stochastic optimization problems from an SCDM. We will call these *Structural Causal Decision Processes* (SCDP).

Definition 14 (Structural Causal Decision Process)**.** A *structural causal decision process* $(\mathcal{M}, \beta, w)$ is a SCIM $\mathcal{M} = (\mathbf{V}, \mathbf{Z}, \mathcal{E}, (\mathbf{X}, \mathbf{D}, \mathbf{U}), \mathbf{Pr}, \mathbf{f}, \boldsymbol{\theta})$, a discount variable $\beta \in \mathbf{V}$, and an end-state mapping $w : \tilde{\mathbf{X}} \to \mathbf{V}$, composed, via P-block, into a DSOP.

3.2.2 Dynamic Models and Sequential Decomposition. We have shown how sequential decomposition reduces the complexity of solving static models in Section 2.4. Here, we will show that these same computational advantages carry through to the dynamic case.

Consider a P-block $\mathcal{B}^P$ constructed from an SCDM which sequentially decomposes $\mathcal{M}_0 = \mathcal{M}_1 \circ \mathcal{M}_2$, and an algorithm that attempts to derive the value function by iteratively recomputing the value function over all states $\mathcal{X}$, taking expectation over all shocks $\mathcal{Z}$, and optimizing over all actions $\mathcal{A}$. Naively, this results in an $O(|\mathcal{X}||\mathcal{Z}||\mathcal{A}|)$ step for each iteration.

However, because the SCDM is sequentially decomposable, it is possible to solve the optimization problem embedded in the update more efficiently. Whereas $|\mathcal{A}| = |\boldsymbol{\pi}_0|$, we have seen that the second subgame can be solved separately from the first. The complexity of the optimization is $O(|\boldsymbol{\pi}_2||\mathrm{Y}| + |\boldsymbol{\pi}_1|)$, resulting in a total complexity for an iteration of $O(|\tilde{\mathrm{X}}||\mathcal{Z}|(|\boldsymbol{\pi}_2||\mathrm{Y}| + |\boldsymbol{\pi}_1|))$, a significantly tighter bound.

4 SCDPS BEYOND POMDPS TOWARD RESOURCE RATIONALITY

We will discuss the relationship between an SCDP and more commonly known formalisms, MDP and POMDP. While an SCDP has many elements in common with MDPs and POMDPs, it is not exactly the same thing as these. The class of SCDPs is not the same as the class of MDPs because in an SCDP, it is possible that the agent will have limited information when making their decisions. The class of SCDPs is not the same as the class of POMDPs because it is not guaranteed that the agent in the SCDP has memory of sufficient statistics with which to solve for optimal behavior. Rather, the SCDP framework is flexible enough to model the memory capacity of agents directly.

4.1 Example: latent income process

Figure 5 depicts a valid SCDP, governed by the following equations:

$$\begin{aligned}
\epsilon_a &\sim Normal(0,1)\\
a' &= a + e^{\epsilon_a}\\
c &= b + a\\
0 &< d(c) < c\\
\epsilon_b &\sim Normal(0,1)\\
b' &= c - d + e^{\epsilon}_{b}\\
u &= \log d
\end{aligned}$$

This is similar to the consumer problems we have discussed before, except now the agent is subject to an unobserved exogenous income process (a, ϵ, a'). Each period, they experience both a transitory and a permanent (evolving over time) income shock, but they do not know whether the income is from the transitory or permanent source. If the agent were able to observe b at d, their decision rule would in effect have full information of a, but this is not the case.

Given these equations, the agent has no past memory with which to formulate beliefs about their unobserved environment. Rather, they can infer the latent value of a only through their observation c directly. Any inferences about the environment will be implicit in the decision rule π_d.

However, if we add variables and connections, we can construct the capacity for the agent to have a Bayesian filter with which to retain past information and make better decisions.

$$\begin{aligned}
m'(m,c) &\in \{0,1\}\\
0 < d(c,m') &< c
\end{aligned}$$

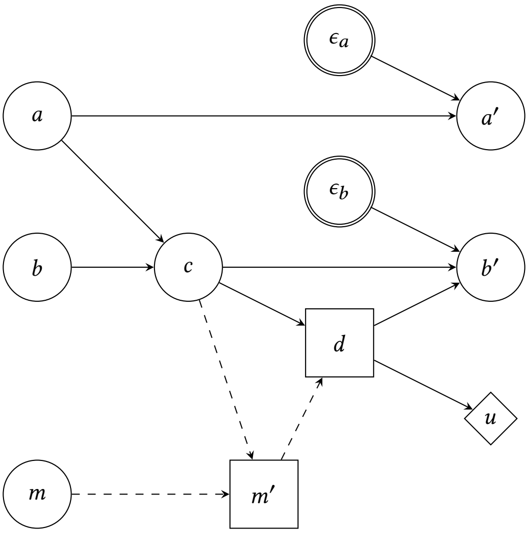

Fig. 5. An SCDP with latent state and shocks. The agent observes c at decision d, but not the exogenous income process $(a, \epsilon_a, a', \epsilon_b)$. Double circles denote exogenous noise variables. An SCDP need not have a Bayesian filter, which is typically considered part of a POMDP, but it is possible to build one in. The variables m, m' and connecting dashed edges represent that (optional) filter.

In this extended model, the agent chooses two decision rules. π_d governs the consumption decision. $\pi_{m'}$ governs the retention of information from observations c. We would expect that the agent equipped with this additional memory unit would be able to capture greater lifetime reward. But for the sake of exposition, we have restricted the 'belief states' of this agent to only two states, 0 or 1.

4.2 SCDP and MDP

An MDP is a tuple (S, A, P_a, R_a), where S are the states, A_s are the actions available at each state $s \in S$, $P_a(s, s')$ are the transition probabilities, and $R(s, a)$ is the immediate reward after having taken action a in state s. There are many variations on this formalism. This is very similar to the definition of the DSOP in Definition 9, in that it sets up all the elements needed for the dynamic optimization problem, but for the discount factor β.

However, canonically, the policy function in an MDP is $\pi(s)$, a function of state values. This implies that the agent has full information about the state variables when they make their decision – it is a 'fully observed' model.

In contrast, in an SCDP, decision rules π_D are functions of the $\mathbf{Pa}_D$, which may not include all the variables in $\tilde{\mathbf{X}}$. SCDPs are expressive enough to allow for partial-observability or latent variables in a model. That includes the case in which decisions are made sequentially but with exogenous shocks interleaved in between them.

4.3 SCDP and POMDP

POMDPs are an extension to MDPs that separate states from observations [28]. It is a tuple $(S, A, \Omega, T, O, R, \beta)$, where S are states, A are actions, Ω are a set of possible observations, $T(s'|s, a)$ is the transition probability distribution, $O(o|s', a)$ is the observation probability distribution, $R(s, a)$ is the reward function, and β is the discount factor.

POMDPs are typically analyzed as implying an internal belief state MDP (B, A, τ, r, β), which is a kind of subjective mirror to the real environment described by the POMDP. A belief transition function $\tau(b'|a, b)$ operates as a Bayesian filter, aggregating information from prior observations and storing what is ideally a sufficient statistic in the belief state b, and $r(a, b)$ computes the expectations of reward R given belief in b. The agent 'solves' the POMDP by constructing and solving the belief state MDP.

This formation of the belief state MDP implies that there is an information flow from past beliefs to the information set of current decisions which is *not required* for an SCDP. The SCDP makes explicit which variables are in the information sets for its decision variable(s). For example, in the first version of the model we introduce in Section 4.1, the agent does not remember prior information. They are limited to their immediate observation c but have no way of tracking the latent state beyond that signal. In the second version, the filter m serves as the memory for the agent, subject to the agents' decision rule π_m.

So, SCDPs are a more general form of model than POMDPs. There are other ways in which SCDPs are more expressive than POMDPs. For example, a single SCDP can have multiple decision variables, each with a different information set, with different linking 'memory' states.

4.4 Expressing resource rational and other realistic agents

Our contention is that this additional expressiveness that SCDPs have, beyond the POMDP, is useful when considering how to model realistic settings with agents. This is because the rationality assumptions implied by the POMDP model are not realistic for many social scientifically valid agents that act with limited information, irrational decision rules, and operate in complex environments that are composed out of many contexts. (See the discussion of resource rationality in Section 1.4.)

Minor variations of the example model can expand the memory available to the agent ($m \in \mathbb{R}$), change the way that memory works (by making $m'(m, x)$ a fixed function rather than a flexible decision rule), or by composing it with additional states (as in the portfolio allocation module in Section 3.1). We do not guarantee that all environments developed in this way are efficiently soluble. Rather, we are concerned with how to flexibly describe agent environments based on their realistic informational and computational constraints.

4.5 Example: costly memory

We present one more example that illustrates how SCDPs can represent the design of artificial agents in a way that exceeds the representational power of POMDPs. It varies the model in Section 4.1 only slightly by allowing the agent to make a decision about *how well they remember* past information. This is a model of an agent which can retain data about past observations, but at a cost. This reflects the contemporary situation of firms which may have to pay for cloud storage for their data.

We retain the equations of the consumer with a latent income process.

$$
\begin{aligned}
\epsilon_a &\sim Normal(0, 1) \\
a' &= a + e^{\epsilon_a} \\
c &= b + a \\
0 < d(c, q) &< c \\
\epsilon_b &\sim Normal(0, 1) \\
b' &= c - d + e_b^{\epsilon} \\
u &= \log d
\end{aligned}
$$

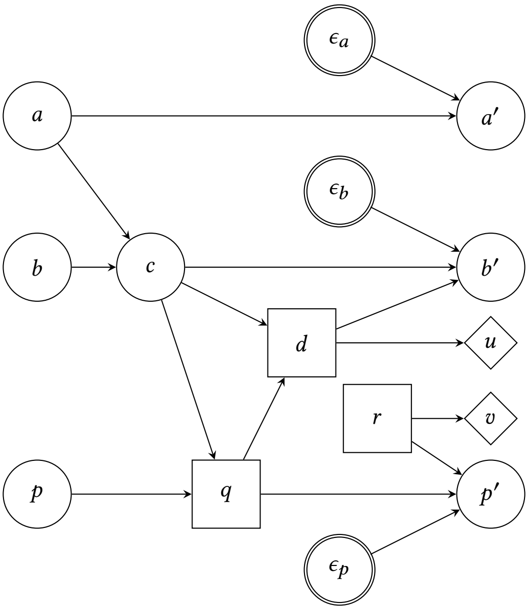

Fig. 6. An SCDP in which the agent consumes at d, updates their beliefs about the world at q, and chooses how much to remember at r. The agent experiences joy with consumption u and consternation at remembering carefully v. Their resources vary over time due to transitory shocks ϵ_b as well as a latent stochastic income process which is not directly observed.

However, we alter the memory process (p, q) so that it is noisy ϵ_p, with the noise modulated by a tunable rate of remembering r which introduces a cost v.

$$\begin{aligned} q(p, c) &\in \mathbb{R} \\ r &> 0 \\ v &= -1/r \\ \epsilon_p &\sim Normal(0, 1) \\ p' &= q + \epsilon_p r \end{aligned}$$

In this model, the agent now has three decision variables, each with a different information set.

- At q, the agent updates their estimate of the state of the latent income process based on the new information at c.
- At d, the agent makes a consumption decision, expending scarce resources in order to experience reward at u.
- At r, the agent makes a global decision about how much information they will retain from period to period, as r limits the drift of memory caused by ϵ_p. The agent can choose to have more accurate memory, but at the cost of utility at v.

This is well-formulated as an SCDP. It is not possible to model this problem as a POMDP. Unlike in the POMDP, we have endogenized the agent's memory and belief formation into the model, making these subject to strategies and cognitive costs.

5 VARIABLE DISCOUNT FACTORS IN SCDPS

In Definition 14, we discussed that the discount factor for an SCDP is a choice of variable $\beta \in \mathbf{V}$. In most presentations of dynamic programming problems, the discount factor β (or γ) is a constant; this is reflected in Definition 9. However, in economics there is a widespread use of dynamic, state dependent, and non-exponential discounting, which we discussed in Section 1.5.

Variable discounting poses no fundamental obstacle for the SCDM representation. Because the discount factor is a selected state variable $\beta \in \mathbf{V}$, it may have its own structural equation governing its evolution. (It may also be assigned a constant.) For example, in the quasi-hyperbolic case, β takes different values depending on the temporal distance of the decision. In the state-dependent case of Stachurski and Zhang [48], $\beta = \beta(x)$ for some function of the state. The graphical structure of the SCDM then makes explicit how discounting interacts with other state variables, potentially revealing new decomposition opportunities or, conversely, identifying when variable discounting breaks orthomodularity.

5.1 Example: Stochastic discount factors

We now present an example with stochastic discount factors following Krusell and Smith [30]. In this formulation, the discount factor β is a separate state variable that follows its own Markov process, independent of the income shock, allowing patience to vary across agents and over time.

Consider a consumption-saving problem where the discount factor $\beta \in \{\underline{\beta}, \bar{\beta}\}$ transitions according to a Markov chain with transition matrix Π_β. The persistence of β is calibrated so that the average duration in each state corresponds to an agent's lifetime. The agent maximizes:

$$\max_{\{c_t\}} E\left[\sum_{t=0}^{\infty}\left(\prod_{s=0}^{t-1}\beta_s\right)u(c_t)\right]$$

subject to the following transition equations and constraints:

$$\begin{aligned} y' &= \rho y + \epsilon_y;\ \epsilon_y \sim \mathcal{N}(0, \sigma^2) && (18)\\ \beta' &= g_\beta(\beta, \epsilon_\beta);\ \epsilon_\beta \sim F_\beta && (19)\\ 0 \leq c(a, y, \beta) &\leq a + y && (20)\\ u &= u(c) && (21)\\ a' &= a + y - c && (22) \end{aligned}$$

where a is the level of assets, y is income, β is the discount factor, u is utility from consumption, and g_β maps the current discount factor and shock to the next period's value according to the Markov transition Π_β. The income shock follows an AR(1) process, while $\beta'|\beta$ evolves independently. We assume $R = 1$ for simplicity.

The Bellman equation for this problem is:

$$v(a, y, \beta) = \max_{c \in [0, a+y]} \left\{u(c) + \beta \cdot E_{y',\beta'}[v(a + y - c, y', \beta')]\right\} \tag{23}$$

The expectation is over both y' and β', which evolve independently according to their respective stochastic processes. Figure 7 shows the influence diagram. The agent observes the state (a, y, β) and chooses consumption c. The law of motion $a' = a + y - c$ determines next period's assets from current resources minus consumption. The bridge variables (a', y', β') appear in overlapping boxes to indicate they belong to both the current period (as outputs of the transition) and the next period (as initial states).

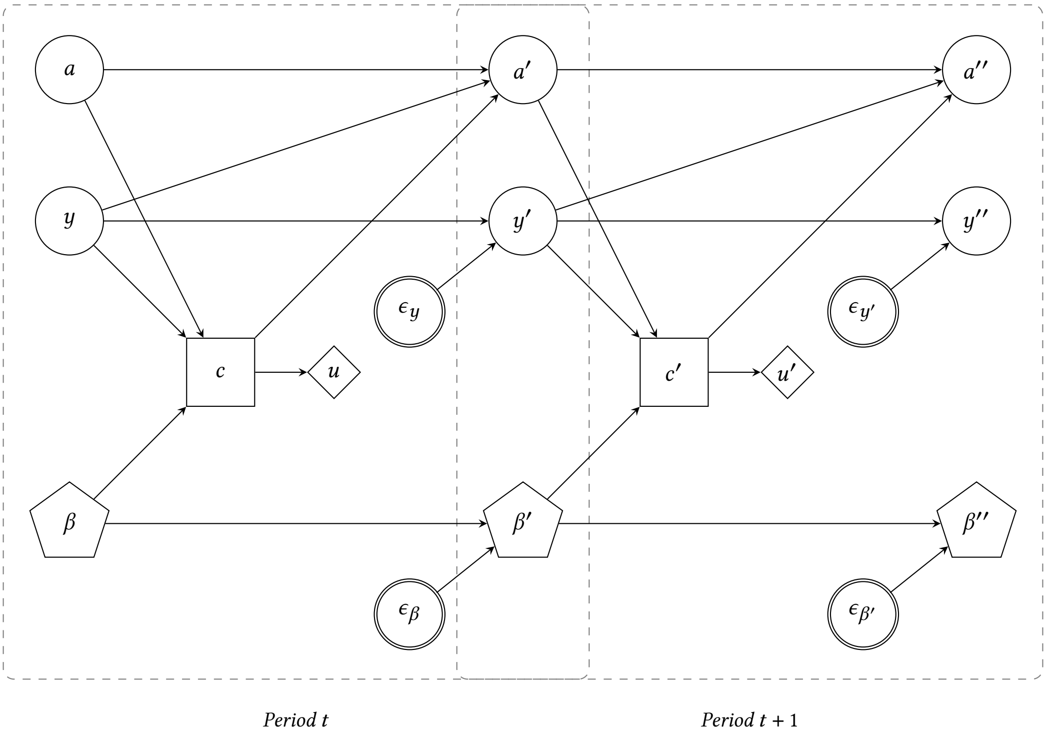

Fig. 7. Influence diagram for the consumption-saving problem with stochastic discount factors. The agent observes (a, y, β) and chooses consumption c. The bridge variables (a', y', β') connect to the next period; the boxes overlap on these nodes to indicate they belong to both components. Income and discount factor each evolve via Markov processes driven by exogenous shocks. Pentagons denote discount factors; double circles denote exogenous noise variables.

This model decomposes in the standard way. The state a' depends on (a, y, c) via the law of motion $a' = a+y-c$, while β' depends only on β through its stochastic transition, not on endogenous choices. Since both y and β evolve exogenously, they do not create additional channels between periods, and the orthomodularity condition remains satisfied. When $\beta = \bar{\beta}$ for all t, this reduces to the standard model. The stochastic formulation preserves the recursive structure that enables dynamic programming.

6 CONCLUSION

In this paper, we have approached the problem of modeling agents in computing systems with several desiderata in mind. These include:

- Graphical causal modeling precisely captures the structural relationships between variables.
- The agents are goal directed, but they may be imperfectly informed, have insufficient recall of their past, and can act suboptimally.
- The models are (de)composability, such that the analyst can capture efficiency gains due to the strategic independence of control variables.
- The models can represent latent, unobserved state and multiple decision variables with different information sets within a single model or dynamic period.
- Limitations to cognitive resources and value discounting are endogenous to the model.

We introduce new modeling tools, SCDMs and SCDPs, which satisfy all of the above criteria. These tools build on recent work in causal game theory, and extend it to the dynamic setting. SCDMs are distinguished from prior constructs like SCIMs mainly because of how they allow for root variables to be ungoverned by probability distributions or structural equations. These root nodes are what enable SCDMs to be composed and also converted into SCDPs. We show that SCDPs are more expressive than POMDPs, particularly with respect to their ability to model resource rationality and variable discount factors.

In future work, we will extend SCDMs and SCDPs into multi-agent systems. The groundwork has already been laid for this with causal game theory and Structural Causal Games. We will also develop efficient algorithms for solving these models and fitting them to data that take advantage of the causal structure. We are developing a scientific software library, `scikit-agent`, which implements this modeling framework and will later be a repository for implementations of these algorithms.

ACKNOWLEDGMENTS

SB's contributions have been supported by NSF awards 2131532, 2131533 and 2105301, as well as the Future of Life Foundation, the Microsoft Corporation, and the Econ-ARK project. AL's contributions have been supported by NSF award 2131533. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the sponsors.

REFERENCES

[1] [n. d.]. Beyond Preferences in AI Alignment, author=Zhi-Xuan, Tan and Carroll, Micah and Franklin, Matija and Ashton, Hal, journal=Philosophical Studies, volume=182, number=7, pages=1813–1863, year=2025, publisher=Springer. ([n. d.]).

[2] Akash Agrawal, Joel Dyer, Aldo Glielmo, and Michael J Wooldridge. 2025. Robust policy design in agent-based simulators using adversarial reinforcement learning. In *The First MARW: Multi-Agent AI in the Real World Workshop at AAAI 2025*.

[3] John R Anderson. 1991. Is human cognition adaptive? *Behavioral and brain sciences* 14, 3 (1991), 471–485.

[4] David I August, Sharad Malik, Li-Shiuan Peh, Vijay Pai, Manish Vachharajani, and Paul Willmann. 2005. Achieving structural and composable modeling of complex systems. *International Journal of Parallel Programming* 33, 2 (2005), 81–101.

[5] Robert Axelrod. 2006. Agent-based modeling as a bridge between disciplines. *Handbook of computational economics* 2 (2006), 1565–1584.

[6] Robert L Axtell and J Doyne Farmer. 2025. Agent-based modeling in economics and finance: Past, present, and future. *Journal of Economic Literature* 63, 1 (2025), 197–287.

[7] Osman Balci, James D Arthur, and William F Ormsby. 2011. Achieving reusability and composability with a simulation conceptual model. *Journal of Simulation* 5, 3 (2011), 157–165.

[8] Sebastian Benthall. 2019. Situated information flow theory. In *Proceedings of the 6th Annual Symposium on Hot Topics in the Science of Security*. 1–10.

[9] Lawrence Blume. 2015. Agent-based models for policy analysis. In *Assessing the Use of Agent-Based Models for Tobacco Regulation*. National Academies Press (US).

[10] András Borsos, Adrian Carro, Aldo Glielmo, Marc Hinterschweiger, Jagoda Kaszowska-Mojsa, and Arzu Uluc. 2025. Agent-Based Modelling at Central Banks: Recent Developments and New Challenges. (2025).

[11] Craig Boutilier, Richard Dearden, and Moisés Goldszmidt. 2000. Stochastic dynamic programming with factored representations. *Artificial intelligence* 121, 1-2 (2000), 49–107.

[12] Dan Cao. 2020. Recursive equilibrium in Krusell and Smith (1998). *Journal of Economic Theory* 186 (2020), 104978.

[13] Micah Carroll, Alan Chan, Henry Ashton, and David Krueger. 2023. Characterizing manipulation from AI systems. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–13.

[14] Nick Chater, Mike Oaksford, Nick Chater, and Mike Oaksford. 1999. Ten years of the rational analysis of cognition. *Trends in cognitive sciences* 3, 2 (1999), 57–65.

[15] Elliot Creager, David Madras, Toniann Pitassi, and Richard Zemel. 2020. Causal Modeling for Fairness in Dynamical Systems. arXiv:1909.09141 [cs.LG] https://arxiv.org/abs/1909.09141

[16] Thomas G Dietterich. 2000. Hierarchical reinforcement learning with the MAXQ value function decomposition. *Journal of artificial intelligence research* 13 (2000), 227–303.

[17] Joshua M Epstein and Robert Axtell. 1996. *Growing artificial societies: social science from the bottom up*. Brookings Institution Press.

[18] Tom Everitt, Ryan Carey, Eric D Langlois, Pedro A Ortega, and Shane Legg. 2021. Agent incentives: A causal perspective. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35. 11487–11495.

[19] Tom Everitt, Marcus Hutter, Ramana Kumar, and Victoria Krakovna. 2021. Reward tampering problems and solutions in reinforcement learning: A causal influence diagram perspective. *Synthese* 198, Suppl 27 (2021), 6435–6467.

[20] Irving Fisher. 1930. *The Theory of Interest*. MacMillan, New York.

[21] James Fox, Tom Everitt, Ryan Carey, Eric D Langlois, Alessandro Abate, and Michael J Wooldridge. 2021. PyCID: A Python Library for Causal Influence Diagrams.. In *SciPy*. 65–73.

[22] Lewis Hammond, James Fox, Tom Everitt, Alessandro Abate, and Michael Wooldridge. 2021. Equilibrium Refinements for Multi-Agent Influence Diagrams: Theory and Practice. *arXiv preprint arXiv:2102.05008* (2021).

[23] Lewis Hammond, James Fox, Tom Everitt, Ryan Carey, Alessandro Abate, and Michael Wooldridge. 2023. Reasoning about causality in games. *Artificial Intelligence* 320 (2023), 103919.

[24] Christopher Harris and David Laibson. 2001. Dynamic choices of hyperbolic consumers. *Econometrica* 69, 4 (2001), 935–957.

[25] Stephen Kasputis and Henry C Ng. 2000. Composable simulations. In *2000 Winter Simulation Conference Proceedings (Cat. No. 00CH37165)*, Vol. 2. IEEE, 1577–1584.

[26] Zachary Kenton, Ramana Kumar, Sebastian Farquhar, Jonathan Richens, Matt MacDermott, and Tom Everitt. 2023. Discovering agents. *Artificial Intelligence* 322 (2023), 103963.

[27] Daphne Koller and Brian Milch. 2003. Multi-agent influence diagrams for representing and solving games. *Games and economic behavior* 45, 1 (2003), 181–221.

[28] Vikram Krishnamurthy. 2016. *Partially observed Markov decision processes*. Cambridge university press.

[29] Per Krusell, Burhanettin Kuruşçu, and Anthony A Smith. 2002. Equilibrium welfare and government policy with quasi-geometric discounting. *Journal of Economic Theory* 105, 1 (2002), 42–72.

[30] Per Krusell and Anthony A Smith. 1998. Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy* 106, 5 (1998), 867–896.

[31] Per Krusell and Anthony A Smith. 2003. Consumption–savings decisions with quasi–geometric discounting. *Econometrica* 71, 1 (2003), 365–375.

[32] David Laibson. 1997. Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics* 112, 2 (1997), 443–478.

[33] Sydney Levine, Matija Franklin, Tan Zhi-Xuan, Secil Yanik Guyot, Lionel Wong, Daniel Kilov, Yejin Choi, Joshua B Tenenbaum, Noah Goodman, Seth Lazar, et al. 2025. Resource Rational Contractualism Should Guide AI Alignment. *arXiv preprint arXiv:2506.17434* (2025).

[34] Falk Lieder and Thomas L Griffiths. 2020. Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and brain sciences* 43 (2020), e1.

[35] Alan Lujan. 2026. EGM^n: The Sequential Endogenous Grid Method. (2026). Working Paper.

[36] Lilia Maliar, Serguei Maliar, and Pablo Winant. 2021. Deep learning for solving dynamic economic models. *Journal of Monetary Economics* 122 (2021), 76–101.

[37] Shie Mannor, Ishai Menache, Amit Hoze, and Uri Klein. 2004. Dynamic abstraction in reinforcement learning via clustering. In *Proceedings of the twenty-first international conference on Machine learning*. 71.

[38] Vishwali Mhasawade and Rumi Chunara. 2021. Causal Multi-Level Fairness. arXiv:2010.07343 [cs.LG] https://arxiv.org/abs/2010.07343

[39] Richard E Neapolitan et al. 2004. *Learning bayesian networks*. Vol. 38. Pearson Prentice Hall Upper Saddle River, NJ.

[40] Cyrus Neary and Ufuk Topcu. 2023. Compositional learning of dynamical system models using port-Hamiltonian neural networks. In *Learning for Dynamics and Control Conference*. PMLR, 679–691.

[41] Argentina Ortega, Samuel Parra, Sven Schneider, and Nico Hochgeschwender. 2024. Composable and executable scenarios for simulation-based testing of mobile robots. *Frontiers in Robotics and AI* 11 (2024), 1363281.

[42] Christiaan JJ Paredis, Antonio Diaz-Calderon, Rajarishi Sinha, and Pradeep K Khosla. 2001. Composable models for simulation-based design. *Engineering with Computers* 17, 2 (2001), 112–128.

[43] Judea Pearl. 1994. A probabilistic calculus of actions. In *Uncertainty in artificial intelligence*. Elsevier, 454–462.

[44] Judea Pearl. 2009. *Causality*. Cambridge university press.

[45] Jonathan Richens and Tom Everitt. 2024. Robust agents learn causal world models. *arXiv preprint arXiv:2402.10877* (2024).

[46] Atharva Sehgal, Arya Grayeli, Jennifer J Sun, and Swarat Chaudhuri. 2023. Neurosymbolic grounding for compositional world models. *arXiv preprint arXiv:2310.12690* (2023).

[47] Ross D Shachter. 1986. Evaluating influence diagrams. *Operations research* 34, 6 (1986), 871–882.
[48] John Stachurski and Junnan Zhang. 2021. Dynamic programming with state-dependent discounting. *Journal of Economic Theory* 192 (2021), 105190.
[49] Robert H Strotz. 1956. Myopia and inconsistency in dynamic utility maximization. *The Review of Economic Studies* 23, 3 (1956), 165–180.
[50] Chris Van Merwijk, Ryan Carey, and Tom Everitt. 2022. A complete criterion for value of information in soluble influence diagrams. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 10034–10041.
[51] Pieter van Schalkwyk and Dan Isaacs. 2023. Achieving scale through composable and lean digital twins. In *The Digital Twin*. Springer, 153–180.
[52] Thomas Verma and Judea Pearl. 1988. *Influence Diagrams and D-seperation*. University of California (Los Angeles). Computer Science Department.
[53] Neal Wagner. 2024. Comparing the complexity and efficiency of composable modeling techniques for multi-scale and multi-domain complex system modeling and simulation applications: a probabilistic analysis. *Systems* 12, 3 (2024), 96.
[54] Sifan Wang, Shyam Sankaran, and Paris Perdikaris. 2022. Respecting causality is all you need for training physics-informed neural networks. *arXiv preprint arXiv:2203.07404* (2022).
[55] Xinyue Wang and Biwei Huang. 2025. Modeling Unseen Environments with Language-guided Composable Causal Components in Reinforcement Learning. *arXiv preprint arXiv:2505.08361* (2025).
[56] Hongxin Zhang, Zeyuan Wang, Qiushi Lyu, Zheyuan Zhang, Sunli Chen, Tianmin Shu, Behzad Dariush, Kwonjoon Lee, Yilun Du, and Chuang Gan. 2024. COMBO: compositional world models for embodied multi-agent cooperation. *arXiv preprint arXiv:2404.10775* (2024).
[57] Feng Zhu, Yiping Yao, Jin Li, and Wenjie Tang. 2019. Reusability and composability analysis for an agent-based hierarchical modelling and simulation framework. *Simulation Modelling Practice and Theory* 90 (2019), 81–97.

A REFERENCE

A.1 Structural Causal Games

A Bayesian network is a graphical model of a joint distribution over random variables.

Definition 15 (Causal Bayesian Network (CBN) [44]). A causal Bayesian network (CBN) over set of random variables $\mathbf{V}$, parameters θ, and joint probability distribution $\Pr(\mathbf{V};\theta)$ is a structure $\mathcal{M} = (\mathbf{V}, \mathcal{E}, \theta)$ such that:

- $\mathcal{G} = (\mathbf{V}, \mathcal{E})$ is a directed acyclic graph with vertices $\mathbf{V}$ and edges $\mathcal{E}$, and
- $\Pr(\mathbf{v};\theta) = \prod_{V \in \mathbf{V}} Pr(v|\mathbf{pa}_V;\theta_V)$, where $\mathbf{pa}_V$ are the parents of V on $\mathcal{G}$.

The marginal probability distribution of each variable is conditional only on its parents. Together, these marginal conditional distributions constitute the original joint distribution $\Pr_\theta(V)$. While this distribution can be represented by many different (Markov-equivalent) Bayesian network structures, the edge structure $\mathcal{E}$ is considered *causal* with respect to interventions on the variable values [44].

A *structural causal model* (SCM) is a Bayesian Network that has all of its stochasticity in its root nodes. The exogenous variables – those which have no parents – are random variables. The values of the endogenous variables – which do have parent nodes – are governed by structural equations or, equivalently, deterministic conditional probability distributions.

Definition 16 (Structural Causal Model (SCM) [44]). A (Markovian) structural causal model is a CBN $\mathcal{M} = (\mathbf{W}, E, \boldsymbol{\theta})$ where $\mathbf{W} = (\mathbf{V} \cup \mathbf{Z})$, with exogenous variables $\mathbf{Z}$ and endogenous variables $\mathbf{V}$ such that for all $Z \in \mathbf{Z}$, $\mathbf{pa}_Z = \emptyset$ and for all $V \in \mathbf{V}$, $\mathbf{pa}_Z \neq \emptyset$. The parameters $\boldsymbol{\theta}$ assign deterministic distributions $\Pr(V|\mathbf{pa}_V;\theta_V)$ to each endogenous variable and a stochastic distribution $\Pr(\mathbf{Z};\boldsymbol{\theta}) = \prod_{Z \in \mathbf{Z}} \Pr(Z;\theta_Z)$ to the exogenous variables.

A *structural causal game* (SCG) is an SCM that reserves some of its variables as decision and utility variables, assigned to particular agents.

Definition 17 (Structural Causal Game [23, 27]). A structural causal game (SCG) is a structure $\mathcal{G} = (N, \mathbf{W}, E, \boldsymbol{\theta})$ where

- $N = \{1, \dots, n\}$ is a set of agents
- $(\mathbf{W}, E, \boldsymbol{\theta})$ is an SCM with endogenous variables $\mathbf{V} \subset W$.
- $\mathbf{V}$ is partitioned into nature variables $\mathbf{X}$, decision variables $\mathbf{D} = \cup_{i \in N} \mathbf{D}_i$, and utility variables $\mathbf{U} = \cup_{i \in N} \mathbf{U}_i$

A single-player variation of an SCG is the SCIM work of Hammond et al. [23]:

Definition 18 (Structural Causal Influence Model (SCIM) [23]). A *structural causal influence model* $(\mathbf{V}, \mathbf{E}, \mathcal{E}, (\mathbf{X}, \mathbf{D}, \mathbf{U}), \mathbf{Pr}, \boldsymbol{\theta})$ consists of:

- A set of endogenous variables $\mathbf{V}$
- A set of exogenous variables $\mathbf{E} = \{E_V\}_{V \in \mathbf{V}}$.
- A graph $\mathcal{G} = (\mathbf{E} \cup \mathbf{V}, \mathcal{E})$ is a DAG over $\mathbf{E}$ and $\mathbf{V}$ where $\mathbf{Pa}_V \cap \mathbf{E} = \{E_V\}$
- $\mathbf{V}$ is partitioned into:
  - $\mathbf{X}$, state variables
  - $\mathbf{D}$, decision variables[7]
  - $\mathbf{U}$, utility variables
- Distribution $\mathbf{Pr}$ and parameters $\boldsymbol{\theta}$ assign deterministic distributions $Pr(x|\mathbf{pa}_x, \theta_X)$ to each state variable [8], $Pr(u|\mathbf{pa}_u, \theta_U)$ to each utility variable, and a stochastic distribution $\mathbf{Pr}(\mathbf{E}, \boldsymbol{\theta}) = \prod_{E \in \mathbf{E}} Pr(E; \theta_E)$ to the exogenous variables.

The SCDM definition used in this paper is indebted to this prior work and is an extension of these constructs.

A.2 Strategic reliance

For multi-agent causal influence models (of which the SCIMs of Definition 18 are a subclass), Koller and Milch [27] identify graphical criteria for the strategic relevance of decision variables on each other, which allow for efficient algorithms for finding the Nash equilibria. SCIMs are a simpler problem, and the same logic applies.

Definition 19 (Strategic reliance [27]). A decision variable D *strategically relies* on a decision variable D' in an SCIM if there are two policy profiles π and π' such that π and π' differ only at D', but some decision rule for D is optimal for π and not for π'.

For an SCIM, let $U_D = \mathbf{U} \cap \mathrm{Desc}(D)$ be those utility variables that are descendants of decision variable D.

Definition 20 (S-reachability [27]). A node D' is *s-reachable* from a node D in an SCIM if there is some utility node $U \in U_D$ such that if a new parent $\hat{D}'$ were added to D', there would be an active path from $\hat{D}'$ to $Pa(U)$ given $Pa(D) \cup D$.

The definition of *active path* comes from earlier work on causal graphical models.

Definition 21 (d-separation; active path [52]). On a directed graph, two sets of nodes X and Y are d-separated given a third set Z if and only if there is no active bi-directed path from a node in X to a node in Y.

A path between two nodes is an *active path* given a set of nodes Z if every node with converging arrows (a collider) either is or has a descendant in Z and every other node along the path is not in Z.

[7] The influence diagram literature rarely intersects with the control theory literature. Decision variables and control variables are roughly synonymous.

[8] The use of probability distributions for the deterministic variable assignments are a consequence of this construct's being derived from work on Bayesian networks. We will address this directly in a moment.

Koller and Milch [27] then prove the connection between S-reachability and strategic reliance.

Theorem 22 (Soundness [27])**.** *If D and D' are two decision nodes in a SCIM and D' is not s-reachable from D in the MAID, then D does not rely on D'.*

Koller and Milch [27] outline a general algorithm for identifying the reliance structure of decision nodes in a multi-agent influence diagram and efficiently solving for their optimal and equilibrium strategy profiles.