The Cognitive Foundations of Economic Exchange: A Modular Framework Grounded in Behavioral Evidence

Anonymous Authors[1]

Abstract

A key challenge in multi-agent AI is modeling social cooperation under realistic behavioral constraints. Many foundational concepts in economics and ethics—such as "trust" or "morality"—are often defined informally, without operational criteria or cognitive grounding, which limits their testability and implementation in artificial agents. Drawing on converging empirical evidence from primate behavior, infant cognition, and economic anthropology, we propose a conceptual framework composed of three cognitively minimal mechanisms: individual recognition, reciprocal credence, and cost–return sensitivity. This framework reframes trust as a graded cognitive expectation, providing a simulateable basis for reciprocal exchange in artificial agents, and enabling the bottom-up emergence of scalable cooperation and institutional dynamics.

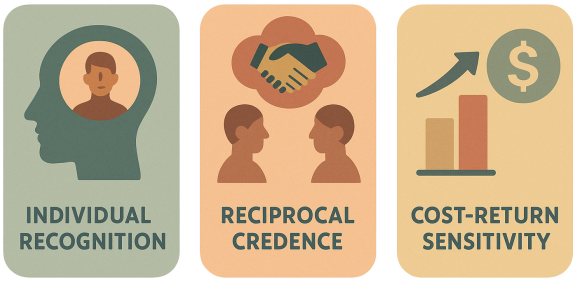

Figure 1. Three Core Cognitive Mechanisms—Individual Recognition, Reciprocal Credence, and Cost–Return Sensitivity—as Behavioral Primitives for Simulating Reciprocal Exchange in Artificial Agents.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

1. Introduction

A persistent gap in economic simulation is the lack of behavioral grounding: most models begin with institutions—markets, contracts, payoff matrices—while assuming the cognitive and social mechanisms that make such systems viable. This limits both the explanatory power and simulateability of artificial economic agents.

This issue is mirrored in broader AI+Society research. Concepts like "trust," "value," and "cooperation" are often treated as abstract labels, operationalized only at the outcome level. Without grounding these constructs in observable mechanisms, models risk simulating the surface of economic life without modeling its origins.

In this paper, we argue that the foundations of economic exchange—and by extension, scalable cooperation—can be traced to three cognitively minimal, biologically grounded mechanisms: *individual recognition*, *reciprocal credence*, and *cost–return sensitivity*. These capacities are observed across human and nonhuman primates, and support sustained reciprocal exchange without requiring symbolic trust or formal enforcement.

Anthropological accounts have long shown that early human exchange was not rooted in barter or equivalence, but in socially embedded reciprocity. Systems like the Kula ring (Malinowski, 2013) emphasized obligation and alliance over transaction. Similarly, primatology has documented robust reciprocal behaviors among chimpanzees and bonobos—grooming (De Waal, 1997), food sharing, and coalition—indicating that the behavioral foundations of reciprocal exchange are observable across species and precede formal market systems.

We propose a modular framework built from these three mechanisms, designed to serve as behavioral primitives in economic simulation. This approach not only bridges cognitive science and institutional theory, but also enables bottom-up modeling of how debt, credit, and monetary behavior may emerge from repeated social interaction.

Our contribution. We offer a theoretical realignment: a cognitively grounded account of reciprocal exchange that clarifies how economic systems can emerge from social

behavior rather than institutional design. Specifically, we:

- Identify three minimal mechanisms—individual recognition, reciprocal credence, and cost–return sensitivity—as simulateable behavioral primitives for modeling the emergence of economic exchange;
- Synthesize evidence from primate behavior, infant cognition, and economic anthropology to support these mechanisms across species and developmental stages;
- Reframe "trust" as a scalar, simulateable expectation—reciprocal credence—rather than a moral abstraction;
- Reinterpret key findings from behavioral economics—such as inequity aversion, reciprocal cooperation, and altruistic punishment—as consistent expressions of underlying biological mechanisms, rather than anomalies to rational choice theory.

This work reframes economic modeling not as a top-down implementation of rules, but as a bottom-up emergence of institutions from cognitive interaction.

Ethical Statement Importantly, this work does not rely on evolutionary explanations. While we draw on behavioral evidence from primates and human infants, our goal is not to claim innate or adaptive origins of exchange. Rather, these cases serve as empirical constraints to identify minimal cognitive mechanisms sufficient for reciprocal behavior. Our account is grounded in behavioral plausibility, not evolutionary teleology.

2. Related Work

2.1. Agent-Based Social Simulation

Recent advances in multi-agent language models have produced systems capable of planning, negotiation, and scripted cooperation (Park et al., 2023; Li et al., 2023). However, these models typically lack behavioral grounding: they simulate interaction as token-level dialogue without modeling the internal mechanisms that make social exchange stable—such as partner memory, reciprocity tracking, or cost–benefit sensitivity.

While some frameworks include persistent memory modules or heuristic-based reasoning, these components often lack principled constraints. Agents either memorize indiscriminately or update state in ways that are unstable or opaque, making it difficult to simulate long-term reciprocal dynamics. As a result, apparent cooperation often emerges from prompt bias or hardcoded behavior, not from simulateable social inference.

This gap has two key consequences: (1) It limits the explanatory value of multi-agent simulations as models of human social behavior, and (2) it prevents systematic exploration of how early exchange mechanisms scale into institutional forms. Our framework addresses this by identifying minimal, biologically plausible mechanisms that can be embedded in LLM agents to support grounded, dynamic reciprocity over time.

2.2. Reciprocal Behavior in Nonhuman Primates

Research in primatology provides critical insight into the behavioral foundations of reciprocity. Species such as chimpanzees and bonobos engage in reciprocal behaviors across contexts like food sharing, grooming, and coalition support. For instance, de Waal (De Waal, 1997) illustrated how chimpanzees form long-term social alliances maintained through reciprocal exchange, and Brosnan et al. (Brosnan & De Waal, 2003) demonstrated that primates are sensitive to fairness and inequity in outcomes.

While these studies have richly documented the behavioral forms of reciprocity, they often stop short of linking such behaviors to the broader architecture of economic interaction. The question of how these mechanisms scale or evolve into structured systems of exchange remains relatively underexplored.

Our framework builds on these findings by situating primate reciprocity within a cognitive-behavioral model. We argue that capacities such as individual recognition, memory of interaction history, and fairness sensitivity form a biologically grounded substrate upon which more complex human economic systems could emerge.

2.3. Social Exchange Without Markets

Anthropological accounts have long challenged the barter-origin myth that underpins classical economic theory. Foundational works such as The Gift (Mauss, 2024) and Stone Age Economics (Sahlins, 2013) argue that early exchange was not driven by market logic, but embedded in webs of social relationships, mutual obligation, and symbolic prestige.

These accounts have been instrumental in shifting attention away from transactional equivalence and toward socially embedded forms of value. Yet, they often bracket the question of what kinds of cognitive and behavioral mechanisms make such systems feasible. In other words: what allows individuals to engage in sustained, deferred, and socially contingent exchange?

Our framework complements these anthropological insights by identifying the minimal behavioral architecture that enables reciprocity to stabilize across time and social partners. We argue that reciprocity is not merely a cultural inven-

tion, but a behaviorally tractable structure observable across species.

2.4. From Behavioral Anomalies to Cognitive Substrates

Behavioral economics has significantly advanced our understanding of social preferences, particularly in contexts of fairness, cooperation, and reciprocity. Paradigms such as the Ultimatum Game (Güth et al., 1982), Trust Game (Berg et al., 1995), and Public Goods Game (Fehr & Gächter, 2002) reveal systematic deviations from self-interested rationality, with participants rejecting unfair offers, rewarding cooperation, and punishing free-riders.

These empirical patterns have led to influential models of "inequity aversion" (Fehr & Schmidt, 1999) and "strong reciprocity" (Gintis, 2000), which posit internalized fairness norms and preferences for norm enforcement. However, these models typically treat the relevant capacities—such as fairness sensitivity or social tracking—as given, rather than as targets of explanation.

In contrast, our framework treats these preferences as emergent outcomes of specific cognitive-social mechanisms: individual recognition, memory of interaction, and cost–return evaluation. By shifting the focus from outcomes to underlying capacities, we aim to bridge behavioral economic observations with a biologically grounded account of exchange behavior.

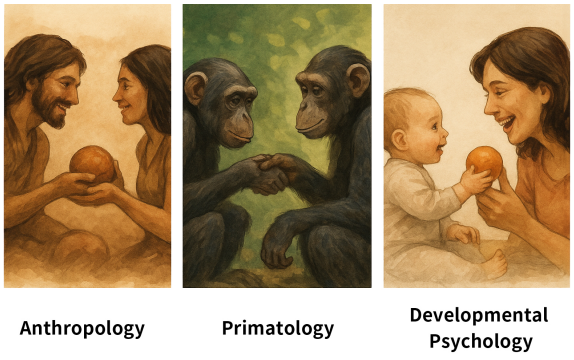

Figure 2. Evidence from anthropology, primatology, and developmental psychology suggests that the behavioral foundations of economic exchange lie in reciprocity—not barter or market logic.

3. Background: Behavioral Origins of Exchange

3.1. The Myth of Barter and the Reciprocal Foundations of Exchange

The textbook account of economic origins often begins with barter: the notion that early humans exchanged goods directly in the absence of currency, with markets and money arising later to reduce transactional friction. This narrative, while intuitively appealing, has little support in either ethnographic or historical records.

As Sahlins argued in *Stone Age Economics* (Sahlins, 2013), early exchange systems were not structured around equivalence, but around relational modes of reciprocity—generalized, balanced, and negative. These forms differ not in rationality but in social distance and temporal orientation: generalized reciprocity, typical among kin, involved open-ended giving without expectation of immediate return; balanced reciprocity, among peers, entailed roughly equivalent exchanges over time; and negative reciprocity, often among rivals or strangers, reflected opportunistic or exploitative behavior.

A paradigmatic ethnographic example is the Kula ring, described by Malinowski (Malinowski, 2013), in which ceremonial armbands and necklaces circulate across Trobriand Islands through long-term, prestige-based exchanges. These objects held little utilitarian value, yet were exchanged with strict directional patterns, social memory, and partner exclusivity. What appears as a symbolic ritual is, upon closer inspection, a complex system of relational tracking, reputational assessment, and obligation extension—precisely the kinds of cognitive demands that structure early forms of exchange.

Mauss's *The Gift* (Mauss, 2024) further formalized this insight by showing how gift-giving in archaic societies is embedded in a triadic obligation: to give, to receive, and to reciprocate. The gift is not merely a transfer of value, but a means of sustaining long-term social bonds. When viewed from this angle, early exchange is not the product of impersonal rational choice, but of biologically and socially grounded behaviors—tracking individuals, remembering interactions, maintaining alliances, and adjusting for asymmetrical returns.

These foundations suggest that exchange predates markets not just historically, but cognitively. What emerges from these systems is not merely institutional scaffolding, but a structured behavioral logic—one that we formalize in the following section.

3.2. The Three Core Mechanisms of Biological Reciprocity

If the earliest forms of economic interaction were not driven by impersonal barter, then what behavioral dynamics made sustained exchange possible? We argue that reciprocal exchange emerged not from a need to equalize value, but from the ability to maintain accountability and continuity across repeated social interactions.

These early exchanges were not market transactions, but embedded social patterns shaped by memory, mutual respon-

siveness, and partner-specific interaction history. Exchange, in this view, arises not from price formation, but from behavioral infrastructure: the cognitive and social capacities that allow individuals to track others, assess cooperative value, and anticipate reciprocity.

To formalize these behavioral foundations, we propose a cognitively grounded framework composed of three simulateable mechanisms:

- **Individual recognition**: enabling agents to identify and re-engage specific social partners over time;
- **Reciprocal credence**: a scalar, updateable expectation that cooperation will be returned;
- **Cost–return sensitivity**: allowing agents to regulate cooperative behavior based on dynamic payoff asymmetries.

These three capacities are supported by converging evidence from nonhuman primates and human infants, suggesting they are both biologically plausible and developmentally early. Importantly, they are not merely descriptive—they are simulateable as modular components in multi-agent systems.

While primate studies have richly documented reciprocal behaviors—grooming, food sharing, coalition support—the cognitive prerequisites for sustaining such behavior are often left implicit. As (Schino & Aureli, 2010) note, reciprocity is frequently treated as a local response pattern, rather than a structured process involving memory and expectation. Yet reciprocal exchange, by definition, requires continuity. Without recognition or anticipation of future interaction, trust-like behavior cannot functionally emerge.

By defining these mechanisms as behavioral primitives, we offer a bottom-up model of economic emergence—one that grounds institutional systems not in rules, but in biologically plausible, cognitively minimal substrates of social interaction.

4. Theoretical Framework: Simulateable Primitives for Economic Exchange

Existing efforts in agent-based cooperation often rely on engineered incentives, fixed strategy spaces, or symbolic approximations of concepts like "trust." What remains missing is a biologically grounded, simulateable account of the behavioral capacities that make reciprocal exchange possible.

We propose three minimal mechanisms—individual recognition, reciprocal credence, and cost–return sensitivity—as a compositional framework for simulating the emergence of economic exchange. These mechanisms are:

- **Cognitively minimal**: each is supported by empirical data from nonhuman primates and human infants;
- **Behaviorally tractable**: each has observable correlates in real-world cooperation;
- **Simulateable**: each can be implemented in agent-based systems as a decision module or update function.

This framework allows social exchange to be modeled from the bottom up—without relying on predefined markets, institutions, or payoff matrices.

4.1. Individual Recognition

We define individual recognition as the ability to identify and distinguish specific agents across time, enabling the tracking of prior social interactions. This capacity underlies the first and most basic source of reciprocal credence: memory of positive experiences with known others. Without recognition, no information about past cooperation can be meaningfully retained or used to guide future behavior.

Among nonhuman primates, this capacity is especially robust. A recent study shows that chimpanzees and bonobos remember familiar individuals after decades of separation, and preferentially attend to those with whom they had positive social histories (Lewis et al., 2023). These findings show that individual recognition in apes is long-term and identity-specific, enabling the tracking of social partners across time.

In nonhuman primates, individual recognition enables partner-specific cooperation. In rope-pulling tasks, chimpanzees adjusted their actions based on the identity and prior behavior of their partners—pausing, glancing, or gesturing to initiate coordination (Hirata & Fuwa, 2007). Dyads with higher food-sharing tolerance were also more successful in joint tasks, indicating that cooperative memory is structured around known social partners(Melis et al., 2006b).

In humans, this ability emerges early. Infants as young as 14 months selectively help or share with individuals who have previously acted prosocially (Dunfield et al., 2011). Such findings suggest that recognition, even in its early form, is tied to social history and used to guide future cooperative behavior.

4.2. Reciprocal Credence

The term "trust" is widely used across disciplines, yet rarely defined with precision. In both everyday language and scientific literature, it is invoked to describe moral commitment, emotional closeness, institutional reliability, and behavioral expectation—often interchangeably. This conceptual ambiguity makes trust difficult to operationalize, especially in comparative, developmental, or artificial settings.

To address this gap, we introduce the concept of *reciprocal credence*: a cognitively grounded, graded expectation that another agent will respond to cooperation with cooperation. Reciprocal credence does not assume moral virtue or social obligation. Instead, it arises from informational cues—typically based on one of two elements: (1) past prosocial behavior directed toward the self or others, or (2) contextual signals that make cooperative return likely. It is not a feeling, but a function: a decision-weighting variable that governs the willingness to enter or continue reciprocal exchange.

Unlike categorical trust, reciprocal credence is inherently scalar: it reflects how likely an agent believes cooperation will be reciprocated, not whether it will be or not. Treating reciprocal credence as a scalar variable is essential to modeling cooperation as a dynamic, context-sensitive behavior.

We distinguish three primary sources of *reciprocal credence*:

1. **Direct positive interaction history**, where an individual has experienced prior cooperative behavior, prosocial engagement, or emotional bonding with another.
2. **Reputational inference**, where the individual observes or learns about others' cooperative behavior.
3. **Role-based expectation**, where social roles or institutional norms lead to default expectations of reciprocity.

Each source varies in reliability, stability, and cognitive cost, but all contribute to the emergence of trust-like behavior without requiring moral commitment. Together, these sources provide the informational basis for *reciprocal credence*—a flexible, graded expectation that others will act in ways conducive to cooperation.

Empirical studies support the differential emergence of these sources across species and developmental stages. In chimpanzees, reciprocal cooperation is consistently observed when individuals have direct histories of prosocial interaction, with little evidence for reputation-based inference or role-based expectation. Similarly, human infants display early sensitivity to agents who have helped them directly, yet show limited use of third-party observation or role generalization. These patterns suggest that direct positive interaction is the foundational layer of reciprocal credence, while the other two sources emerge developmentally and culturally.

Across species, direct prosocial interaction serves as the most robust and foundational basis for reciprocal credence. In wild chimpanzees, grooming is exchanged not merely in-the-moment, but over extended time horizons. Longitudinal data reveal that grooming given and received is significantly correlated across dyads, even with delays spanning several days—indicating that chimpanzees maintain memory of past interactions and calibrate future behavior accordingly (Gomes et al., 2009). Experimental studies further show that chimpanzees prefer to collaborate with previous cooperative partners, suggesting they encode individual-specific social history and deploy it strategically in cooperative contexts (Melis et al., 2006a).

Human infants exhibit a comparable early capacity. By three years of age, children are less likely to help individuals who have previously acted with harmful intent—even when the harm was directed at a third party, and even if no harm was successfully done. This selective avoidance indicates that infants track not only outcomes but also the moral valence of agents' prior actions when deciding to engage prosocially (Vaish et al., 2010).

Together, these findings support the primacy of direct positive interaction in generating reciprocal credence. Unlike inference or abstraction, this source relies on lived experience—anchoring cooperative expectations in observable history, not hypothetical roles or indirect cues.

A note on "trust". It is also important to distinguish social trust from what might be better described as functional dependency. Contemporary systems such as ChatGPT or Google are often described as being "trusted," when in reality they are simply relied upon due to their consistent output, broad adoption, and lack of failure. This perceived "trust" is better understood as a product of usability, network effects, and repeated exposure — not reciprocal social accountability.

A similar semantic conflation is common in financial discourse, where "trust in money" or "trust in banks" is frequently cited as foundational to economic stability. However, what is usually meant in such contexts is not interpersonal trust, but confidence in the continuity of institutional enforcement, liquidity guarantees, or the absence of systemic failure.

4.3. Cost–return sensitivity

We define *cost–return sensitivity* as the capacity to evaluate whether engaging in cooperative behavior yields a net benefit over time. Unlike categorical heuristics (e.g., always cooperate or always defect), this sensitivity enables agents to condition their participation on the payoff structure of ongoing interactions. It allows withdrawal from exploitative relationships and reinforcement of beneficial ones, serving as a regulatory mechanism for reciprocal dynamics.

This capacity need not involve explicit calculation; approximate assessments—such as tracking recent imbalances in giving and receiving—are sufficient to support adaptive behavior. For example, chimpanzees are less likely to groom or share with partners who have failed to reciprocate in the

past, even if those costs are not immediate or symmetrical. Similarly, infants gradually reduce helping behavior toward agents who consistently fail to reciprocate attention or aid.

Such sensitivity ensures that reciprocal exchange remains robust under uncertainty, preventing exploitation while maintaining flexibility in dynamic social environments.

In nonhuman primates, cost–return sensitivity manifests in both prosocial and retaliatory contexts. Chimpanzees engage in grooming and service-like exchanges where helping behavior depends on previous benefits received from others (De Waal, 1997). They are also capable of negative reciprocity: retaliating against individuals who have stolen from them, even when such retaliation yields no direct benefit—suggesting an expectation of equitable return and a sensitivity to intentional harm (Jensen et al., 2007).

In humans, early traces of this mechanism appear in infancy. By 18 months, children selectively direct helping or sharing behavior toward agents who previously acted fairly or cooperatively (House et al., 2013). These decisions are not based on abstract principles of justice, but on perceived patterns of cost and return—who gave, who withheld, and under what circumstances.

Taken together, these findings suggest that cost–return sensitivity enables organisms to calibrate cooperative investment, maintaining reciprocity in the absence of formal institutions or explicit contracts. As a building block of economic cognition, it governs the micro-dynamics of exchange long before the emergence of price or property.

5. Possible Implementation via Memory Design

While each proposed mechanism is tractable in isolation, integrating them into a unified architecture remains an open challenge—particularly due to current limitations in memory design for LLM-based agents. Most existing multi-agent frameworks lack persistent, structured memory, resulting in shallow social modeling and limited capacity for behavioral adaptation.

We propose that each behavioral primitive—individual recognition, reciprocal credence, and cost–return sensitivity—can be realized through targeted extensions to agent memory and retrieval mechanisms:

- **Individual recognition** can be implemented using identity-specific memory slots or embedding-indexed retrieval systems, enabling agents to store and retrieve interaction histories with named partners. This supports continuity in agent behavior and the ability to condition future responses on partner identity.
- **Reciprocal credence** may be represented as an internal latent variable or scalar annotation associated with each agent in memory (e.g., "Agent X: credence = 0.7"), and updated through lightweight heuristics (e.g., recency-weighted prosocial acts). This allows LLM agents to infer, represent, and update partner expectations without requiring explicit reward functions.
- **Cost–return sensitivity** can be implemented through structured memory logs of past cooperative attempts and outcomes. Agents may track cumulative give–take asymmetries and adjust future investments accordingly. Prompts can surface this information as part of decision rationale (e.g., "Agent Y has taken without giving for three turns").

These strategies enable biologically inspired reciprocity to be modeled using memory-centric extensions, rather than reinforcement learning or symbolic planning. By shifting the burden of social behavior onto structured memory design and prompt-level reasoning, this framework opens a tractable path toward more socially competent LLM agents.

6. Implications and Discussion

6.1. Limitations of the Framework

Comparative Methodology. Comparative psychology and developmental psychology use fundamentally different experimental paradigms. While primate studies often capture interactive, ecologically valid behaviors, human infant research typically relies on simplified, highly constrained designs. This asymmetry limits the granularity with which we can compare capacities across species.

Institutional Complexity. Although we argue that modern exchange institutions build upon a biologically grounded behavioral substrate, we do not claim that such mechanisms alone are sufficient to explain the emergence of formal systems like currency, interest-bearing debt, or taxation. These likely require additional cultural, historical, and symbolic developments.

Social Scale and Enforcement. Our framework focuses on the dyadic and small-group foundations of reciprocal exchange, where individuals recognize and respond to the behavior of specific partners. It does not yet account for mechanisms that support large-scale, anonymous cooperation—such as third-party punishment, norm internalization, or reputation management across broader social networks. These are critical features of modern institutions, and extending the framework to address how such systems emerge from or layer upon basic social cognition remains a key direction for future work.

6.2. Theoretical Implications

From Institutions to Cognition. Our work reorients the study of exchange from an institutionalist or symbolic perspective toward a cognitive-behavioral foundation. Rather than treating markets, money, or debt as cultural inventions that enabled exchange, we argue that exchange itself is grounded in prior cognitive and social capacities—capacities that predate and scaffold institutionalization.

Rethinking Trust as System Robustness. The term "trust" is often invoked in discussions of economic exchange, yet it bundles together distinct ideas—from interpersonal reliability to confidence in systems. We argue that this vagueness limits its explanatory value.

To address this, we introduce *reciprocal credence*: a graded, biologically grounded expectation that cooperation will be returned. It emerges from recognition, interaction history, and cost–return sensitivity—not from moral obligation. In contrast, large-scale "trust" in systems like money or platforms is better understood as confidence in robustness, not social accountability. This reframing clarifies how reciprocal behavior scales without assuming institutional trust.

Cross-Species Foundations of Exchange. By aligning primate cooperation studies with human developmental data and ethnographic accounts of non-monetary exchange, we provide a cross-species bridge for modeling the origins of economic behavior. This integrative view challenges the idea that complex exchange systems are uniquely human, and instead locates their roots in broader social cognition.

Unifying Behavioral Economics through Cognitive Foundations. Many findings in behavioral economics—such as fairness preferences, reciprocal cooperation, and sensitivity to framing or loss—are often treated as deviations from rational choice theory. Yet evidence from primatology and developmental psychology suggests that these behaviors reflect consistent biological patterns: they emerge early in human development, recur across great apes, and follow predictable social-cognitive constraints. Rather than anomalies or patchwork heuristics, these tendencies may be better understood as structured expressions of an underlying behavioral logic—one shaped by the cognitive demands of navigating social exchange.

6.3. Future Work

6.3.1. Operationalizing Economic Emergence in Agent Simulations

A key direction for future research is to instantiate the three proposed behavioral primitives—*individual recognition*, *reciprocal credence*, and *cost–return sensitivity*—as explicit, modular components in LLM-based multi-agent systems.

This involves not only defining each mechanism's internal representation (e.g., persistent memory slots for identity recognition, scalar credence values for reciprocity tracking, and interaction-weighted cost–return summaries), but also studying how they interact dynamically across time and partners.

We propose that future simulations systematically test:

- How memory constraints (e.g., limited identity tracking) affect the emergence of sustained cooperation;
- Whether agents with scalar credence variables spontaneously differentiate partners and regulate prosocial behavior;
- Under what conditions cost–return asymmetries trigger withdrawal, retaliation, or symbolic substitutes (e.g., tokens, credit markers).

Such experiments can help identify the minimal memory and inference structure necessary to support scalable exchange. More broadly, this line of work offers a concrete path toward simulating institutional emergence—from micro-level reciprocity to meso-level norms and proto-economic artifacts—grounded in biologically plausible cognition.

6.3.2. Simulating Institution Formation via Behavioral Primitives

A promising direction for future work is to simulate the emergence of economic institutions—such as debt, symbolic exchange, and trust networks—by embedding behavioral primitives into artificial agents. These primitives, derived from comparative psychology and developmental research, include individual recognition, reciprocal credence, and cost–return sensitivity—as the minimal substrate for modeling scalable cooperation. Memory and bonding may be embedded within the learning processes that govern credence formation.

With the rise of large language model (LLM)-based agents and embodied multi-agent simulators (e.g., Habitat, ThreeDWorld, AutoGen), it is now possible to explore whether agents endowed with these primitives can spontaneously develop systems of reciprocal exchange, role specialization, symbolic debt, or proto-monetary behaviors. Such experiments would help identify the minimal substrate for scalable cooperation and provide a generative framework for building socially competent artificial agents.

Importantly, this approach also enables us to test under what conditions behavioral reciprocity transitions into institutionalization: for example, whether limited memory

capacity leads to symbolic tokens, or whether increasing agent density necessitates generalized trust representations. Simulation environments offer a new way to bridge cognitive models of exchange with the design of future artificial social systems.

6.3.3. Exchange among unfamiliar agents

One unresolved but critical question concerns the possibility of reciprocal exchange among unfamiliar agents. While most existing studies—whether in primatology, developmental psychology, or economic anthropology—focus on cooperation within socially familiar contexts, little is known about how exchange unfolds in the absence of recognition, shared norms, or interaction history. Transactions between unfamiliar individuals, especially without institutional enforcement, may require distinct cognitive or communicative scaffolding that is not yet well understood. We highlight this not as a limitation of the current framework, but as an open frontier that future research must address.

7. Conclusion

This paper proposes a cognitively grounded framework for modeling the behavioral origins of economic exchange in multi-agent systems. Instead of relying on institutional assumptions or symbolic abstractions, we identify three simulateable primitives—individual recognition, reciprocal credence, and cost–return sensitivity—that support sustained reciprocal cooperation in the absence of formal mechanisms.

Our framework addresses a central modeling challenge: how to operationalize constructs such as "trust" or "morality," which are foundational in human societies but often lack consensus definitions or concrete implementation paths in computational systems. Drawing on evidence from primate behavior, infant cognition, and economic anthropology, we reframe these abstract concepts in terms of observable, cognitively minimal mechanisms—providing a tractable basis for implementation in artificial agents.

This perspective encourages a shift from top-down economic modeling to bottom-up simulation, where complex structures emerge from agent interaction rather than being imposed. Future work can embed these primitives into large-scale multi-agent environments to explore the spontaneous formation of scalable cooperation and exchange systems.

Impact Statement

This paper proposes a cognitively grounded framework for simulating the behavioral origins of economic exchange, with potential applications in multi-agent modeling, artificial social systems, and foundational research in AI + Society. While the work is theoretical and does not involve data collection or deployment, it provides biologically plausible primitives that could improve the interpretability and social competence of agent-based systems. We foresee no foreseeable negative societal consequences or ethical risks associated with this research.

References

Berg, J., Dickhaut, J., and McCabe, K. Trust, reciprocity, and social history. *Games and economic behavior*, 10(1): 122–142, 1995.

Brosnan, S. F. and De Waal, F. B. Monkeys reject unequal pay. *Nature*, 425(6955):297–299, 2003.

De Waal, F. B. The chimpanzee's service economy: Food for grooming. *Evolution and Human Behavior*, 18(6): 375–386, 1997.

Dunfield, K., Kuhlmeier, V. A., O'Connell, L., and Kelley, E. Examining the diversity of prosocial behavior: Helping, sharing, and comforting in infancy. *Infancy*, 16(3):227–247, 2011.

Fehr, E. and Gächter, S. Altruistic punishment in humans. *Nature*, 415(6868):137–140, 2002.

Fehr, E. and Schmidt, K. M. A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, 114(3):817–868, 1999.

Gintis, H. Strong reciprocity and human sociality. *Journal of theoretical biology*, 206(2):169–179, 2000.

Gomes, C. M., Mundry, R., and Boesch, C. Long-term reciprocation of grooming in wild west african chimpanzees. *Proceedings of the Royal Society B: Biological Sciences*, 276(1657):699–706, 2009.

Graeber, D. *Debt: The First 5,000 Years, Updated and Expanded*. Melville House, 2014.

Güth, W., Schmittberger, R., and Schwarze, B. An experimental analysis of ultimatum bargaining. *Journal of economic behavior & organization*, 3(4):367–388, 1982.

Hirata, S. and Fuwa, K. Chimpanzees (pan troglodytes) learn to act with other individuals in a cooperative task. *Primates*, 48:13–21, 2007.

House, B. R., Silk, J. B., Henrich, J., Barrett, H. C., Scelza, B. A., Boyette, A. H., Hewlett, B. S., McElreath, R., and Laurence, S. Ontogeny of prosocial behavior across diverse societies. *Proceedings of the National Academy of Sciences*, 110(36):14586–14591, 2013.

Jensen, K., Call, J., and Tomasello, M. Chimpanzees are vengeful but not spiteful. *Proceedings of the National Academy of Sciences*, 104(32):13046–13050, 2007.

Lewis, L. S., Wessling, E. G., Kano, F., Stevens, J. M., Call, J., and Krupenye, C. Bonobos and chimpanzees remember familiar conspecifics for decades. *Proceedings of the National Academy of Sciences*, 120(52):e2304903120, 2023.

Li, G., Hammoud, H., Itani, H., Khizbullin, D., and Ghanem, B. Camel: Communicative agents for" mind" exploration of large language model society. *Advances in Neural Information Processing Systems*, 36:51991–52008, 2023.

Malinowski, B. *Argonauts of the western Pacific: An account of native enterprise and adventure in the archipelagoes of Melanesian New Guinea [1922/1994]*. Routledge, 2013.

Mauss, M. *The gift: The form and reason for exchange in archaic societies*. Taylor & Francis, 2024.

Melis, A. P., Hare, B., and Tomasello, M. Chimpanzees recruit the best collaborators. *Science*, 311(5765):1297–1300, 2006a.

Melis, A. P., Hare, B., and Tomasello, M. Engineering cooperation in chimpanzees: tolerance constraints on cooperation. *Animal Behaviour*, 72(2):275–286, 2006b.

Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., and Bernstein, M. S. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, pp. 1–22, 2023.

Sahlins, M. *Stone age economics*. Routledge, 2013.

Schino, G. and Aureli, F. Primate reciprocity and its cognitive requirements. *Evolutionary Anthropology: Issues, News, and Reviews*, 19(4):130–135, 2010.

Schmelz, M., Grueneisen, S., Kabalak, A., Jost, J., and Tomasello, M. Chimpanzees return favors at a personal cost. *Proceedings of the National Academy of Sciences*, 114(28):7462–7467, 2017.

Vaish, A., Carpenter, M., and Tomasello, M. Young children selectively avoid helping people with harmful intentions. *Child development*, 81(6):1661–1669, 2010.

Wang, G., Xie, Y., Jiang, Y., Mandlekar, A., Xiao, C., Zhu, Y., Fan, L., and Anandkumar, A. Voyager: An open-ended embodied agent with large language models. *arXiv preprint arXiv:2305.16291*, 2023.