RIVERS OF CONVERSATIONS
RELATING SYSTEMS THINKING & DESIGN
RSD13 | OCTOBER 2024

Some Design Considerations for Concept-Forming and Concept-Sharing in Discursive Agents

Thomas Manning

In the last decade, there has been a growing interest in advancements made in artificial intelligence via modern natural language processing techniques utilising artificial neural networks in the form of large language models. This has led to the emergence of conversational technologies like ChatGPT, which mimic aspects of the conversational processes that occur between humans. However, it is contended that the model of conversation proposed above is inadequate as a model of human-like conversation. Already, far simpler models are found in the context of autonomous agents and multiagent systems theory (AAMAS), whilst more primitive are better representations of the types of primordial interactions found before the development of conversation proper. But whilst these developments are on the right track regarding replicating the emergence of actual conversational systems, there already exists a preexisting methodology so far not utilised in the context of AAMAS that could be used which better replicate the kind of conversations we humans have with each other. Thus, conversation theory (CT)—which is a cybernetic theory of concept-forming and concept-sharing via conversation—is suggested in this paper as a systems-oriented and cybernetic design methodology that could be used to design and model conversational interactions in AAMAS. Unlike contemporary approaches to the design of conversational and conceptual systems, conversation theory is specifically concerned with the following: How the formation of coherent structures of self-reproducing topics—i.e., the concept—via conversation permits the convergence of shared understandings through positing and testing for analogy relations holding between one participant's

concept with another's concept. One key benefit of utilising this approach for conversational design in the context of researchers working in AAMAS is that its insights are built upon (currently underutilised) literature from educational psychology and early childhood studies. Because of this, it is held in this text that conversation theory has untapped potential as a methodology for designing and analysing conversational interactions that involve mutual conceptualisation between discursive agents.

KEYWORDS: conversation theory, discursive agents, concepts, autonomous agents and multiagent systems

RSD TOPIC(S): Learning & Education, Methods, Methodology, & Theory

Introduction

The contents of this paper propose a cybernetic and systems-oriented design framework for concept-forming and concept-sharing agents for those working in the context of autonomous agents and multiagent systems theory (AAMAS) in the form of conversation theory (CT). This framework acts as an approach that can be used for the contemporary design and modelling of conversational systems, one which incorporates a model of co-evolution and joint-knowledge construction within a given interaction when performing some shared task, goal, or activity. Through this process, we can design to modulate the complexity of future interactions and the contents of a conversation, creating autonomous systems that reach an agreement regarding the design of some artefact.

The design considerations discussed here are derived from conversation theory, a cybernetic theory of discursive practice that was mainly applied to the sociotechnical development of human-computer educational technologies in the 1960s and 1970s. The theory was also incorporated into Nicholas Negroponte's book *Soft Architecture Machines* as a potential systemic design paradigm for generating a machine which could become a co-participant in an architectural design process (Negroponte & Pask, 1976). An implication of such a vision is that we could potentially use CT to create interactions which correspond to the good designer ideal of Sevaldson (2013), whereby such systems use different ways of thinking when conversing to reach a “resolution” for some problem during a creative process, whilst also keeping the general problem open to be revised (p. 1769).

Conversation theory has influenced conversational approaches to thinking about the essence of a design process (Glanville, 2008; Dubberly & Pangaro, 2009; Sweeting, 2019). Such views see conversation as permitting a systemic design paradigm, which could modulate learning, mutual understanding, and conceptual evolution between different participants, who shift between observing, acting, and modelling the complexity of the world around them during a design process. While CT has already been applied to the domains of search engine, curriculum, and educational technology design (See. Tilak et al., 2023a, 2023b), the paper will focus on how it can be deployed as a cybernetic design methodology for researchers in AAMAS seeking to develop conversational technologies. This is done in the belief that developing innovative conversational techniques in AAMAS will have later repercussions for the field of sociotechnical design when the required conversational technologies have matured to the point where we might be able to develop sociotechnical and collaborative co-designing technologies that were originally envisioned in Negroponte’s Architecture Machine Group (Negroponte & Pask, 1976).

In any case, an opportunity arises here for such a cybernetic and systems-oriented design approach, as there has been a growing interest in the study and application of conceptual, social, or discursive practices in the context of machine learning (Lake et al., 2015; Lazaridou et al., 2016; Mao et al., 2019; Lazaridou & Baroni, 2020; Ding et al., 2023; Du, Li, Torralba, Tenenbaum, & Mordatch, 2023; Duenez-Guzman et al., 2023; Tenenbaum et al., 2023; Hsu et al., 2024). This belief of CT’s potential utility as a design methodology is based on two observations regarding recent trajectories of research in the context of machine learning.

The first trajectory relates to research on the *emergence of natural language* in the context of multiagent systems (Lazaridou et al., 2016; Lazaridou & Baroni, 2020; Du et al., 2023). For example, the referential game proposed by Lazaridou provides a model of natural language emergence: It involves two networks containing a sender network and a receiver network. The sender network selects an image from a shared sample of images and emits a symbol corresponding to that image. The receiver network then maps the symbol onto one of two randomised images derived from a shared sample during a given interaction. If it selects the correct one, both networks are rewarded (Lazaridou et al., 2016; Lazaridou & Baroni, 2020). This process has been suggested by Lazaridou and associates to be one pathway through which we may model the emergence of natural language. Another approach that has been taken more recently regarding studying behaviours related to the emergence of natural language involves the use of large language models in the context of multiagent debates to yield a consensus among participants (Du et al., 2023).

The second trajectory has involved *emulating conceptualisation processes* more generally in the context of machine learning. In this context, the term concept has loosely come to be understood by some as the meaning of a word or sentence (Bergen & Feldman, 2008; Ding et al., 2023). We could likewise conceive a concept in general as the meaning we give a denoted thing. The meaning of such-and-such is taken to be simulated or imagined by humans, who respond based on what they assume the correct course of action is (Bergen & Feldman, 2008). Despite the simplicity of this definition, there is an ongoing debate about whether deep learning models are appropriate to mimic human-like intelligence when we can learn a concept in only a small number of iterations rather than through a vast data set (Lake et al., 2015).

Now, contemporary neural symbolic approaches to interpreting the semantic contents of questions, visual concepts, and words have displayed some notable similarities to the way we humans reason. Mao et al. (2019) for example, have suggested such conceptual learning via neural symbolic approaches produces incremental learning in a way that is similar to humans. Recently, meanwhile, Ding et al. (2023) have designed robot agents that utilise visual concepts to create semantic maps and execute actions based on such considerations. They argue to have modelled their robotic agent's behaviour on what they purport to be behaviours found in infant learning and development. This is amicable to the iterative learning proposed by Mao et al. (2019), as well as aspects of established literature in educational psychology and early childhood studies, which provides weight to their assertion (Vygotsky, 1934/1962; Bruner, 1983). However, there are questions about the extent it is appropriate to use neuro-symbolic approaches to mimic human learning when many of the capacities found in pre-verbal infants seem innate rather than learned in short iterations (See. Spelke, 2023). However, the general sentiment on modelling such systems more closely on human development is a welcome one.

Ding et al. (2023) suggest that modelling robotic agents on such considerations would reduce labour-intensive training on large data sets. This sentiment is growing in AAMAS and related fields, as well as the need to take design considerations derived from human development models in the context of machine learning seriously (Tenenbaum et al., 2023). If such data-intensive processing were to occur, we would need to design agents to engage in the kind of cognitive offloading we humans place onto our environments that we do in our everyday discursive practices, e.g., signs, writing systems, numeracy systems, etc. (Vygotsky, 1934/1962, 1978). This would provide a better guide to the kind of social generation of human-like intelligence that has been discussed recently (Duenez-Guzman et al., 2023).

While it is believed here that these developments in machine learning are important, these two trajectories of research—dealing with the emergence of natural language, and emulating processes related to conceptualization — currently show little overlap with each other. There are, however, growing undercurrents of interest as exemplified by a recent Santa Fe Institute workshop (Spelke, 2023; Tenenbaum et al., 2023).

To this end, the aim here is to provide cybernetic and systems-oriented design considerations for the kind of concept-sharing and concept-forming we humans engage in when we converse with each other for those working in the field of autonomous agents and multiagent systems. Any system satisfying the condition of being able to concept-share and concept-form during conversation shall be called a *discursive agent* for this paper. This is done to distinguish it from other conversational agents—such as chatbots—with lack this capacity. It is claimed that without the ability to converse in the sense of understanding and clarifying what we mean by a given topic of discourse, any development in discursive approaches to developing autonomous agents and multiagent systems will be limited regarding emulating human-like intellect. Conversation theory will now be introduced as a—potential—theoretical tool for researchers and designers who wish to emulate the joint-conceptualization processes found in human activity.

Methodology

Conversation theory (CT) is utilised here to articulate cybernetic and systems-orientated design considerations for discursive agents. The reason CT has been chosen for this task is that it is primarily a theory of how discursive agents engage in concept-forming and concept-sharing through means of conversation to achieve some goal (Manning, 2023a, 2023b). Pask (1976; Scott, 2021) credits Vygotsky, Piaget, and others for influencing his approach to designing educational technologies. Specifically, the paired experiments of the Soviet School of Psychology, where participants would construct explanations or responses based on “how” and “why” considerations (Pask, 1976, pp. 19-20). Such experiments were used to examine concept-forming and concept-sharing in conversational participants.

The theory has seen a resurgence of interest, but I believe it is of considerable use in offering design specifications for researchers working with autonomous agents and multiagent systems (AAMAS). However, certain terminology, such as the word “concept”, have slightly different connotations to those being used in contemporary machine learning, so readers are warned not to view the vocabulary used as necessarily synonymous to those of other researchers (See. (Lake et al., 2015; Mao et al., 2019).

Now, conversation theory is a cybernetic theory of conversation that focuses on how given a potential topic of discussion, asynchronous conversational agents can synchronise their understandings of that topic to form a common understanding as to its significance. More specifically, it is a theory concerned with the roles conversational participants take to better understand one another’s perspectives and test them to see if there is an agreement as to what each other understands of a given topic (Pask, 1975a, 1975b, 1976; Scott, 2021). Given the emphasis on asynchronous participants interacting in such a way to synchronise with each other, the theory may be effectively utilised by those in AAMAS who have already started dealing with these concerns in the past decade (Lazaridou et al., 2016; Lazaridou & Baroni, 2020; Du et al., 2023).

One notable benefit this theory has to those working in this area is that it provides a formalised account of conceptualisation, which necessitates discursive agents to give commands, questions, and explanations to each other to perform some activity or to explain what has been done to satisfy this activity. In conversing, each participant is forced to continuously reconceptualise what they purport each other's understandings of a given topic to be. As this process unfolds, and more clarifications are given regarding the nature of a goal or subgoal that is to be achieved, agreements are formed as to the significance of an idea or action within a given interaction. This represents a synchronisation of previous asynchronous autonomous interacting agents, as well as such participants within the conversation co-designing a shared understanding with each other when engaged in some goal or task. While there has been work regarding such processes using language models in multiagent debates, issues related to computational expense and amnesia have been noted as potential limitations to these ends (Du et al., 2023). The paper will now extract design considerations from CT for those working in the AAMAS community.

Analysis

This section has compiled core systems-oriented design considerations from CT for those working in the AAMAS community. These considerations are not exhaustive, and the literature detailing conversation theory and its sister theories is vast and profound. However, as a concise introduction for those working in AAMAS, this account should suffice.

Conversation

At its heart, conversation theory is about how discursive agents *Z* design and calibrate their understandings and activities through conversation and learning. This calibration involves attempting to purport an agreement as to the significance of a given topic and to iteratively test to see if the agreement still holds (Pask, 1975b; Manning, 2023a). The conversational language L is the language through which conversation in conversation theory is said to occur (Pask, 1975b, 1976). It is a language that prioritises the semantic aspects of language above its syntax and permits discursive agents to issue commands and questions to each other. The conversational language L is demarcated into discourse involving how one might achieve some goal L_0 and discourse involving why it works or is coherent with some initial aim or framework of assessment L_1. Thus, $L = \langle L_0, L_1 \rangle$. It is important in conversation theory not only for a conversational participant to utilise discourse regarding how to direct an activity or the steps it took to derive some goal but also why what it did is consistent with pre-existing parameters of assessments such that it means some activity or solution does not contradict previous parameters.

The conversational language L also demarcates discursive content stores π, and descriptions or utterances D along these lines, such that $\pi = \langle \pi_0, \pi_1 \rangle$ and $D = \langle D_0, D_1 \rangle$ hold. Finally, the conversational language presumes there is a modeling facility *MF* whereby two participants may design and edit models of their joint workings through the means of conversation.

Utterances D in conversation theory exist as three modalities of speech acts: Commands, questions, and explanations (Pask, 1975b; Manning, 2023a). Rescher's work on the logic of commands is incorporated into this conception of discursive practice, whereby the speaker obliges the addressee Z to do such-and-such under the condition | of so-and-so (Rescher, 1966; Pask, 1975a). This takes the form:

$$< Z!\, X \mid Y >$$

In this sense, A command places a prescription upon some discursive agent Z! to execute some action, given some precondition; a question places a prescription upon some discursive agent to explain, given some precondition; an explanation is a procedure that yields a topic, and finally, an execution is a process that builds some model M_i in some *MF*. These are specified formally as:

$$COMM\ i \quad = \quad \langle Z!\ EXEC\ i \mid PRECON \rangle$$

$$
\begin{aligned}
\mathit{EQUEST}\ i &= \langle Z!\ \mathit{EXPL}\ i \mid \mathit{PRECON} \rangle \\
\mathit{EXPL}\ i &= \mathit{PROC}(R_i) \\
\mathit{EXEC}\ i &= \mathit{PROC}(M_i)
\end{aligned}
$$

A basic conversational form that permits the giving and asking of reasons and prescribing actions is argued—under considerations from early childhood studies—to correspond to these types under the condition of inducing the desired behaviours from the recipient *Z* (Pask, 1975b; Bruner, 1983; Brandom, 1994; Manning, 2023a, 2023b).

Contemporary literature using question-answer models primarily treats a conceptual agent as a recipient of a request with no intentionality of their own (Gordon et al., 2018; Wijmans et al., 2019; Chen et al., 2023; Hong et al., 2023). The biggest limitation of this approach regarding emulating human behaviour in artificial constructs is that no clarifications are asked by the discursive agent as to what is meant: They simply choose the most appropriate response. For a discursive agent to be said to teach or learn, it may, at times, need to switch roles from learner to teacher and teacher to learner (Pask, 1972). It must not only be able to provide answers when interrogated but interrogate the meaning of the interrogatee. This involves clarifying the interrogatee's concept of their concept to get clarification on what the interrogatee means, which was a feature designed into the paired experiments of the Soviet School, where a respondent could appeal for help and the experimenter provide a demonstration in turn until an agreement was formed around a given topic (Pask, 1976, p. 20).

Concepts

Concepts conceived of in conversation theory are a special type of organisation that is not necessarily synonymous with how it is conceived of in contemporary machine learning. It consists of a finite set of topic relations R_i that can be discussed in conversation. Given an initial hypothesis where R_H is the head topic that forms that hypothesis, the concept does two things: Firstly, it enacts procedures to satisfy that hypothesis, and secondly, tests to see if the set of procedures coheres with a given set of parameters of R_H (Pask, 1975b). Thus, it can be said to involve the reactivation of some subset of configurations (neural or mechanical) to derive itself (Bergen &

Feldman, 2008, p. 321). These steps loosely correspond to L_0, L_1 forms of discourse, where we are concerned with how to do something and why it works. In this sense, a concept can be conceptualised as a problem solver (Pask, 1975b, p. 244).

In attempting to satisfy a given topic, a concept can be conceived as follows:

$$CON(R_H) = R_H$$

Such that a concept of a topic relation forming the hypothesis is substitutable with the satisfaction of said hypothesis, the concept itself is defined by definition—i.e., $\triangleq$—as a process that is considered to consist of the ordered pair of a program and its interpretation. The latter exists in the form of the executable compilation of that program in some processor:

$$CON \triangleq PROC = \langle PROG, INTER \rangle$$

Where the program or code of some input x is specified as $\pi(x)$, and the interpretation of some input is specified as $\lambda(x)$. The input of the interpretation is treated here as a program (or perhaps a series of programs) that may be executed in some processor once compiled (Pask, 1981, p. 273). This entails the following identities:

$$INTER \triangleq \lambda(x)$$

$$PROG \triangleq \pi(x)$$

Since a concept is defined as a process that is reducible to a topic relation, it entails the organisational closure of a given topic relation:

$$EX(PROC(R_H)) = EX(\lambda(\pi(R_H)) \Rightarrow R_H$$

This is what some have described as the looping-through*ness* of a concept, whereby the concept continuously reconstitutes itself as a concept (Pangaro, 2003, pp. 13-19).

One important caveat to this, is that this process of deriving a hypothesis in the form of a topic relation must be a coherent process. The sense of coherence used is derived from the philosophy of Nicholas Rescher (1973) and refers to the mutual reciprocity of topics towards each other. This necessitates in conversation theory that any topic relation R_i must necessarily be inferred and be inferred by other topic relations R_j within a given concept. Thus, $R_i \vdash R_j$ and $R_j \vdash R_i$ within a given concept (Pask, 1975b; Manning, 2023a).

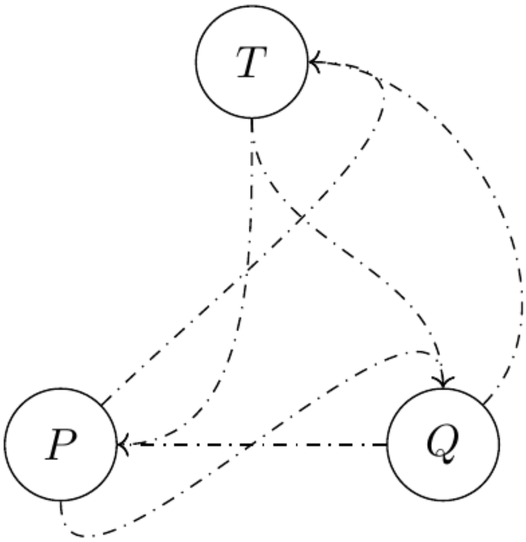

Figure 1: A depiction of a coherent structure of mutually reciprocal topics. Modified by Thomas Manning from a diagram sourced from Nick Green (2006), licensed under the Creative Commons Attribution-Share Alike 3.0.

For example, let there exist topic relations that exist within some finite index *i* such that $\langle R_i \rangle = \langle T, P, Q \rangle$ holds. Now suppose $P, Q \in \Gamma$ such that the two topics belong $\in$ to a set of premises Γ. Suppose that Γ can derive—as represented by the symbol $\vdash$—the topic *T*. We could write the following derivation:

$$\Gamma \vdash T$$

This forms a derivation in the context of CT whose premises are defeasible and open to be revised through subsequent interactions of conceptual reproduction. Such a derivation by itself, however, is not enough to satisfy the criteria of being a concept in conversation theory: It lacks a formal coherent structure, as depicted in Figure 1. For this derivation to be considered a concept, then the following coherence conditionals—as articulated by Rescher (1973)—must hold.

$$\Gamma_1 \vdash T \mid \Gamma_1 = \langle P,Q \rangle$$

$$\Gamma_2 \vdash P \mid \Gamma_2 = \langle T,Q \rangle$$

$$\Gamma_3 \vdash Q \mid \Gamma_3 = \langle P,T \rangle$$

In a way, then, a concept may be considered a coherent cluster of derivations whereby the topic relations that form these inferences are mutually reciprocal with each other in the form of mutually satisfying derivations (as exemplified in Figure 1). This permits the kind of organisational closure understood in conversation theory's conception of a concept, by which topics loop through each other in the process of the productions and reproductions of those said concepts (Pangaro, 2003). In doing so, each topic may be informed by the others that aid in contextualising it.

An example of this might be the following:

$$\langle \textit{infinite\_monkeys, typewriters} \rangle \vdash \textit{the\_works\_of\_shakespeare}$$

Which may be read approximately as: "*The works of Shakespeare* could be recreated through infinite monkeys on typewriters". But this would infer that "*Typewriters* are used by infinite monkeys to recreate the works of Shakespeare" and "*Infinite monkeys* on typewriters recreate the works of Shakespeare". While the above example is descriptive, this is not to say we could not have prescriptive examples of this as well. We could also have sense distinctions, represented through different concepts of a topic: For example, "cool" in the sense of aesthetics and in the sense of temperature. If this were the case, then the concepts would only intersect at the point of the topic of "cool" but would otherwise be semantically incompatible with each other.

In CT, the meaning of a topic is determined by the contexts in which they are used, and it is the job of discursive agents to test to see how such topics are being used through testing to see if they form a coherent bundle of derivations. This may eventually disclose more topics that blend the discursive agents' concept of another's concept, which comes about in CT through the forming of agreements (Pask, 1976).

Agreements

The notion of analogy is fundamental in conversation theory for establishing common understandings between discursive agents (Pask, 1976). Not only must my concept of my concept be analogous to your concept of your concept, but my concept of my concept must be analogous to *your* concept of *my* concept and *vice versa* (Laing et al., 1964; Manning, 2023a). Discursive agents might purport agreements and, likewise, disagreements to hold when there are none regarding some topic of discussion. They converse to clarify what they mean through means of conversational language, which permits the asking of counter-questioning for clarification and teaching back what each other understands of each other's concepts (Pask, 1975b; Brandom, 1994, 2000). Let $Z = \langle A, B \rangle$ in this instance, be used to distinguish two discursive agents. Assuming:

$$CON_A(T) \Leftrightarrow CON_B(T)$$

Then, for all topics in *A* and *B*'s concepts of the topic relation *T*, an analogy relation $\Leftrightarrow$ must hold across *A* and *B*'s topics that constitute the topic relation *T* . This requires discursive agents to have the capacity to engage in interpersonal perception, through which participants may keep score of each other's beliefs and the beliefs others think we hold, and challenge them when applicable (Laing et al., 1964; Brandom, 1994, 2000). In conversation theory, we can have $A(T)$, for example, which stands for *A*'s view of a topic. We can also have $A(B(T))$, which stands for *A*'s view on *B*'s view of a topic. Finally, we can have $A(B(A(T)))$, which stands for *A*'s view on *B*'s view on *A*'s view of a topic. All three are necessary to map discursive agents' actual and purported mental models, to have some criteria to clarify their understanding in relation to others and update accordingly.

Discursive agents must not only keep track of how they constitute their own concept of a topic but also how they view this topic and how others view them to conceive this topic (Laing et al., 1964). In keeping score of each other's conceptions, *A* may request *B* to clarify what they mean through issuing commands or questions within the confines of a conversational language *L*. Once *B* has given an explanation, *A* attempts to clarify their understanding by repeating what they believe *B* has said as a form of teachback

(Pask, 1975b). The discursive agent *B* then tests to see if what has been taught back coheres with their understanding. If it does not, *B* may modify their explanation until they purport *A*'s explanation of their concept to be analogous to their concept of their concept (and *A* agrees with this; Pask, 1975b, 1976).

In doing this, two discursive agents who have asynchronous understandings as to what each other means by a given topic can test the coherence of each other's conceptions by examining context appropriateness and reconstitute their conceptions until they form a generalised analogy between each other. While concepts normally take the form of a general analogy rather than a strict isomorphism, when *A*'s concept of a topic approaches convergence with *B*'s concept of a topic, it is argued in conversation theory that this is where common understandings as to the significance of a given topic *T* begin to emerge and hold across all conceptions.

Discussion

The paper has attempted to articulate the view that the cybernetic and systems-oriented design framework of conversation theory can be used by researchers working in the context of AAMAS to design conversational multiagent systems, which could more closely mimic what we humans do when we converse than other alternative approaches currently being utilised in AAMAS and related fields. CT provides a potential social avenue for the development of human-like intelligence, which focuses on the interaction, design, and co-evolution of joint knowledge by discursive agents in relation to one another rather than on the analysis of datasets (Duenez-Guzman et al., 2023). While there has been contemporary research dealing with the emergence of natural language through multiagent synchronisation (See. Du et al., 2023), the current model specifically focuses on the dynamic modelling of different perspectives through a series of conversational moves rather than relying on large data sets. It also implies the intentionality of agents by virtue of asking for clarification and not as merely responding to a question posed at it (Gordon et al., 2018; Wijmans et al., 2019; Chen et al., 2023; Hong et al., 2023).

While it is possible to emulate some aspects of conversation theory based on the literature discussed here, there is a requirement that discursive agents become co-participants in knowledge construction—*a la* Negroponte's Architectural Machine—

rather than merely respondents to input, which would mean significant reconfiguration of contemporary models to account for this (See. Negroponte & Pask, 1976). If, for example, chatbot models were designed under such considerations, then a greater focus would be had on clarifying why you have asked or commanded it to do something and specifying how you want it to do it such that it would become a co-participant in knowledge production rather than a mere responder to some input. The potential to create systems which co-design with themselves and us through conversation becomes a foreseeable possibility in future if such an approach is adopted.

Because CT focuses on the coherence of conceptual contents, it also becomes easier for discursive agents to backtrack on any judgments they make if contradictory, such that they are required not only to iteratively solve a problem but also to design an explanation as to how they got to it. This requires them to make explicit their commitments through the means of teach-back, where they attempt to cohere what they have done in solving some problem with what they already know and keep score on each other based on such considerations (Pask, 1975b; Brandom, 1994). This is not something—as far as contemporary literature suggests—yet utilised in other models of conceptual activity (Lake et al., 2015; Mao et al., 2019; Ding et al., 2023), and could potentially provide for such researchers a bountiful line of inquiry for both emulating conceptual activity found in humans and modelling the emergence of natural language in AAMAS research.

Conclusion

This paper has been an attempt to introduce considerations when designing conceptual or discursive agents for AAMAS researchers using conversation theory as a cybernetic and systems-oriented design framework. I have focused on the importance of conceptual coherence, different modes of conversational moves, teaching back what one has been taught, and also the role of analogy in generating a common understanding between two or more participants. The literature on CT and its sister theories is vast and profound, so I have only given a snapshot of some aspects of it. However, what has been produced in this paper is more than sufficient to explicate some of its core considerations and show how it could be a useful design framework for those working in AAMAS to deploy in their own research.

References

1. Bergen, B., & Feldman, J. (2008). Embodied concept learning. In *Handbook of cognitive science* (pp. 313-331). Elsevier. https://doi.org/10.1016/B978-0-08-046616-3.00016-5
2. Brandom, R. (1994). *Making it explicit*. Cambridge, MA: Harvard University
3. Press.
4. Brandom, R. (2000). *Articulating reason*. Cambridge, MA: Harvard University Press.
5. Bruner, J. (1983). *Children's talk: Learning to use language*. New York: W. W. Norton and Company.
6. Chen, Z., Zhou, Q., Shen, Y., Hong, Y., Zhang, H., & Gan, C. (2023). See, think, confirm: Interactive prompting between vision and language models for knowledge-based visual reasoning. arXiv preprint arXiv:2301.05226. https://doi.org/10.48550/arXiv.2301.05226
7. Ding, M., Xu, Y., Chen, Z., Cox, D. D., Luo, P., Tenenbaum, J. B., & Gan, C. (2023). Embodied concept learner: Self-supervised learning of concepts and mapping through instruction following. In *Conference on robot learning* (pp. 1743-1754). PMLR. https://doi.org/10.48550/arXiv.2304.03767
8. Du, Y., Li, S., Torralba, A., Tenenbaum, J. B., & Mordatch, I. (2023). Improving factuality and reasoning in language models through multiagent debate. arXiv preprint arXiv:2305.14325. https://doi.org/10.48550/arXiv.2305.14325
9. Dubberly, H., & Pangaro, P. (2009). What is conversation? How can we design for effective conversation. *Interactions Magazine*, *16*(4), 22-28.
10. Duenez-Guzman, E. A., Sadedin, S., Wang, J. X., McKee, K. R., & Leibo, J. Z. (2023). A social path to human-like artificial intelligence. *Nature Machine Intelligence*, *5*(11), 1181-1188. https://doi.org/10.1038/s42256-023-00754-x
11. Glanville, R. (2008). Conversation and design. In R. Luppicini (Ed.), *Handbook of conversation design for instructional applications* (pp. 59-79). IGI Global.
12. Gordon, D., Kembhavi, A., Rastegari, M., Redmon, J., Fox, D., & Farhadi, A. (2018). Iqa: Visual question answering in interactive environments. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 4089 4098). https://doi.org/10.48550/arXiv.1712.03316

13. Hong, Y., Lin, C., Du, Y., Chen, Z., Tenenbaum, J. B., & Gan, C. (2023). 3d concept learning and reasoning from multi-view images. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 9202-9212). https://doi.org/10.48550/arXiv.2303.11327
14. Hsu, J., Mao, J., Tenenbaum, J., & Wu, J. (2024). What's left? concept grounding with logic-enhanced foundation models. *Advances in Neural Information Processing Systems*, *36*.
15. Laing, R. D., Philipson, H., & Lee, A. R. (1964*). Interpersonal perception: A theory and a method of research* (Vol. 6). Abingdon, Oxfordshire: Routledge.
16. Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350*(6266), 1332-1338. https://doi.org/10.1126/science.aab3050
17. Lazaridou, A., & Baroni, M. (2020). Emergent multi-agent communication in the deep learning era. arXiv preprint arXiv:2006.02419. https://doi.org/10.48550/arXiv.2006.02419
18. Lazaridou, A., Peysakhovich, A., & Baroni, M. (2016). Multi-agent cooperation and the emergence of (natural) language. arXiv preprint arXiv:1612.07182. https://doi.org/10.48550/arXiv.1612.07182
19. Manning, T. (2023a). Paskian algerbra: A discursive approach to conversational multi-agent systems. *Cybernetics & Human Knowing*, *30*(1-2), 67-81.
20. Manning, T. (2023b). What is conversation theory. *Cybernetics & Human Knowing*, *30*(1-2), 45-63.
21. Mao, J., Gan, C., Kohli, P., Tenenbaum, J. B., & Wu, J. (2019). The neurosymbolic concept learner: Interpreting scenes, words, and sentences from natural supervision. arXiv preprint arXiv:1904.12584. https://doi.org/10.48550/arXiv.1904.12584
22. Negroponte, N., & Pask, G. (1976). Aspects of Machine Intelligence. In *Soft architecture machines* (pp. 7-51). Cambridge, MA: The MIT Press. https://doi.org/10.7551/mitpress/6317.001.0001
23. Pangaro, P. (2003). The past-future of cybernetics: Conversations, von foerster, and the blc. Retrieved from https://pangaro.com (Unpublished Monograph)
24. Pask, G. (1972). Anti-hodmanship: A report on the state and prospect of cai. *Programmed Learning and Educational Technology*, *9*(5), 235-244. https://doi.org/10.1080/1355800720090502

25. Pask, G. (1975a). *The cybernetics of human learning and performance: A guide to theory and practice*. London: Hutchinson Educational.
26. Pask, G. (1975b). *Conversation, cognition and learning: A cybernetic theory and methodology*. Amsterdam: Elsevier Scientific Publishing Company.
27. Pask, G. (1976). *Conversation theory: Applications in education and epistemology*. Amsterdam: Elsevier Scientific Publishing Company.
28. Pask, G. (1981). Organizational closure of potentially conscious systems. *Autopoiesis: A theory of living organization*, 265-308.
29. Rescher, N. (1966). *The logic of commands*. Oxford: Routledge.
30. Rescher, N. (1973). *The coherence theory of truth*. Oxford: Oxford University Press.
31. Scott, B (2021). *Cybernetics for the social sciences*. Brill: Leiden, The Netherlands. https://doi.org/10.1163/25900587-12340002
32. Sevaldson, B. (2013). Systems Oriented Design: The emergence and development of a designerly approach to address complexity. In J. B. Reitan, P. Lloyd, E. Bohemia, L. M. Nielsen, I. Digranes & E. Lutnæs (Eds.), *DRS // Cumulus: Design Learning for Tomorrow* (pp. 1765-1786). Oslo, Norway.
33. Spelke, E. (2023). Human Intelligence: Insights from Infants. In Workshop on AI and the Barrier of Meaning 2. Santa Fe Institute, Santa Fe, NM, USA. https://www.youtube.com/watch?v=NGMXUzr8MTM&list=PLZlVBTf7N6GpOCwMH-TlhwrUwV8YJMANM&index=8
34. Sweeting, B. (2019). Why design cybernetics? In T. Fischer & C. M. Herr (eds.), *Design cybernetics: Navigating the new* (pp. 185-194). Cham: Springer International Publishing.
35. Tenenbaum, J., Wong, L., Grand, G., Lew, A., Zhang, C., Collins, K., ... others (2023). Scaling ai the human way: Building machines that understand the world as we do, and that we can understand. In Workshop on AI and the Barrier of Meaning 2. Santa Fe Institute, Santa Fe, NM, USA. https://www.youtube.com/watch?v=gzaIrD3jki8&list=PLZlVBTf7N6GpOCwMH-TlhwrUwV8YJMANM&index=2
36. Tilak, S., Doroudi, S., Manning, T., Pangaro, P., Glassman, M., Wen, Z., ... Scott, B. (2023a). The potential of second-order cybernetics in the college classroom. In *Relating systems thinking and design (rsd12)*. Carnegie Mellon University, PA, USA. https://rsdsymposium.org/cybernetics-in-the-college-classroom
37. Tilak, S., Kadir, L., Wen, Ziye., Pangaro, P., & Glassman, M. (2023b). Designing sociohistorically sensitive information search: Experimental analyses of

essays written using thoughtshuffler and google. *Cybernetics & Human Knowing*, *30*(1-2), 133-151.
38. Vygotsky, L. S. (1934/1962). *Thought and language* (E. Hanfmann & G. Vakar, Eds.). Cambridge, MA: The MIT Press.
39. Vygotsky, L. S. (1978). M*ind and society: The development of higher psychological processes* (M. Cole, V. John-Stiner, S. Scribner, & E. Souberman, Eds.). Cambridge, MA: Harvard University Press.
40. Wijmans, E., Datta, S., Maksymets, O., Das, A., Gkioxari, G., Lee, S., ... Batra, D. (2019). Embodied question answering in photorealistic environments with point cloud perception. *Proceedings of the ieee/CVF conference on computer vision and pattern recognition* (pp. 6659-6668).

Author

Thomas Manning, Mr., Institute of Education, UCL, stnvtom@ucl.ac.uk

Acknowledgements

I am grateful to Dr. Scott, Prof. Pangaro, Prof. Laurillard, and Dr. Tilak for their various contributions to my academic development as a scholar utilising the framework of conversation theory. I would likewise also express gratitude to Prof. Derry and Dr. Surendran for getting me to appreciate the Soviet School of Psychology and the Pittsburgh School of Philosophy (it's made connecting a few dots easier).