Construal Level Theory as a Means for Anticipating Human-Robot Interactions

Alan R. Wagner[1], Sarah Rajtmajer[2], and Christopher McClurg[3]

***Abstract*— This paper describes our effort to develop methods for human-robot interaction based on Construal Level Theory (CLT). CLT is a process which connects abstraction to physical and temporal distance. With respect to interactions with people, CLT may govern how to interact with different categories of people. We present preliminary results showing that we can create superordinate, high-level construals. Upcoming work will demonstrate the use of this approach by a robot to plan for interactions with people.**

I. INTRODUCTION

Construal level theory (CLT) hypothesizes that people create construals in order to make predictions and guide behavior with respect to psychological distance [1]–[3]. A *construal* is the way an individual mentally represents a situation, ranging from high-level (abstract) to low-level (detailed) terms. *Psychological distance* is the the subjective feeling of proximity (temporal, spatial, or social) to an individual. CLT states that as psychological distance is decreased, construals evolve from high-level to low-level with mental representations becoming decorated with details (Figure 1). Moreover, CLT suggests a process by which an individual uses prototypical experiences from memory to generate a rough estimate of a future situation, then uses an evolving mental representation to generate increasingly specific counterfactual alternatives to aid in decision making. For example, agreeing to speak at a conference months in the future, an individual will use a memory from some prior talk to produce rough expectations regarding the task. As the time to give the talk gets closer, the person begins to mentalize about specific aspects of the task in order to prepare, perhaps considering how the size of the room will impact how loud they must speak, how a recent cold may hamper their ability to speak, and so on.

The construal process is advantageous because it offers a very flexible and, in some regards, efficient method for planning and using prior knowledge to influence future decision making [3]. The process is flexible in that it allows the person to hold off adding details to their evolving plan until those details are needed, yet also affords opportunities for contingency planning. A disadvantage of the process is that the person's prior experiences may strongly bias them towards some favored course of action without detailed consideration of all options [2]. To an observer with a different psychological distance or perspective, a person or robot's behavior may appear irrational.

Researchers from the planning and reinforcement learning communities have looked at ways to use abstraction as part of the planning and policy generation process [4]. Construal level theory goes beyond these efforts by connecting the construal process not only to planning and action selection, but also to a person's perception and evaluation of objects, events, and other people [2], [5], [6]. We believe that by developing a computational construal process we will gain insight not only into how to develop robots that plan better, but also toward creating systems that plan with respect to abstractions over objects, events, and actors. As such, this paper describes our recent work which will complement current efforts in hierarchical planning and reinforcement learning for human-robot interaction.

A. Why Construal Level Theory?

CLT offers a new perspective on *experience-based reasoning*. CLT describes a process for abstracting away the details surrounding an agent or robot's experience as psychological distance increases. Abstract experiences may serve as initial solutions to many new problems that the robot may encounter. We believe that this approach holds significant promise over state-of-the-art methods in cognitive modeling or planning. Specifically, when the robot is faced with a problem for which it must make some distant future decision, the future problem will often be defined in only abstract terms, excluding much of the detail needed to make the best possible decision. Nevertheless, the robot may use its abstract construals of related experiences to begin to limit the space of its search for solutions. The robot may also be able to use this process to manage multiple, competing goals at the same time, including strategic reasoning about the other robots.

Further, construal level theory is supported by a great deal of psychological evidence [3], [7]–[9]. In tasks ranging from visual perception, to categorization, to action identification, the basic tenets of CLT have been reaffirmed multiple times. This evidence shows that humans employ a construal type process which allows them to reason over and communicate about different levels of abstraction. Moreover, CLT relates abstraction to planning by suggesting that people perform a type of least-commitment planning in which one's most distant goals are only broadly construed in the most abstract

This work was supported by Air Force Office of Scientific Research (AFOSR) grant number FA9550-21-1-0197.

[1]Alan R. Wagner is an assistant professor in the Department of Aerospace Engineering at The Pennsylvania State University alan.r.wagner@psu.edu

[2]Sarah Rajtmajer is an assistant professor in the Department of Information Science and Technology at The Pennsylvania State University smr48@psu.edu

[3]Christopher McClurg is a graduate student in the Department of Aerospace Engineering at The Pennsylvania State University cam7498@psu.edu

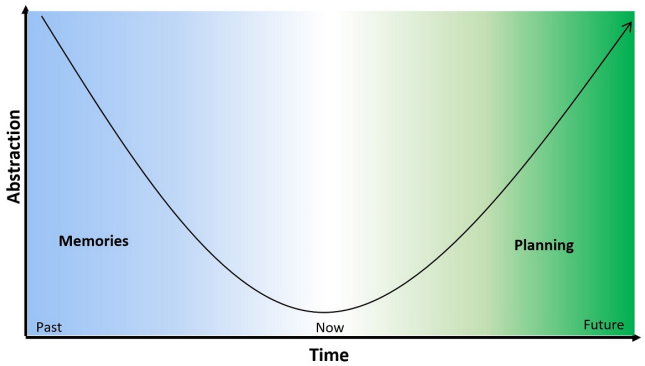

Fig. 1: A conceptual description of how construals influence planning. The graph depicts construal abstraction with respect to time. As the future becomes now, the construals the robot uses to plan become less abstract. As experience becomes the past, memories become more abstract.

terms. As goals become more proximate they become decorated with details until, at the moment of decision making, actions are selected reactively, based on a limited number of remaining choices. This contrasts with both traditional cognitive modeling and classic game theoretic approaches which tend to focus on deliberative context-sensitive appraisals of situations [10]. While deliberation is unquestionably important, ultimately robots must deliberate over one or, at most, a small set of goals. Alternatively, we suggest that CLT may allow robots to manage a number of different, conflicting goals by utilizing temporal or physical distance as a moderator of those goals. Moreover, because CLT connects many different psychological phenomena ranging from emotion to perceptual reasoning, our approach may provide a new perspective on experience-based reasoning. For instance, our approach could be used to simulate higher order social phenomena, such as ingroup/outgroup categorization and reasoning. It has long been shown that egocentric social distance to outgroups is greater than one's social distance to ingroups [11]. CLT posits that outgroups are cognitively represented in terms of high-level construals whereas ingroups are represented using low-level construals. Not only does the cognitive representation explain phenomena such as ingroup/outgroup bias, it also shapes normative behavior, norm development and application, interactive responses among individuals and so on.

More significantly, CLT may lend insight into the reward/utility shaping process itself [3], [12]. CLT does not suggest what an agent will find rewarding or how or when rewards are obtained. But it does note, for example, that some actions will be seen as more positive in the distant future than in the near term. These actions will be represented as high-level construals and the attributes of these will be positive overall [1]. As the action becomes more concrete it may be decorated with negative features which come to dominate the evaluation. For example, the action "helping someone" is viewed as more positive than the concrete action "giving $20 to a homeless person on a subway in Los Angeles at midnight." As the construal becomes decorated negative attributes override the positive high-level construal. All of this is to say that, in addition to human-robot interaction, construal level theory may also provide valuable insights for the creation of reinforcement learning or game-theoretic reward functions.

II. RELATED WORK

Our effort touches on and relates to research from several other fields. We could not, however, locate any prior research directly attempting to develop a computational process based on CLT implemented in an autonomous agent or robot.

A. Hierarchical Planning

The use of multi-level abstractions for planning and behavior control has been thoroughly documented in psychology and neuroscience [13], [14]. Multi-level abstraction hierarchies have also played an important role in the development of planners capable of solving complex problems [15]–[18]. Hierarchical planning ultimately requires both elements of search and task decomposition [19], [20]. From the planning perceptive, our research ties task decomposition to psychological distance and uses psychological distance as a reactive, environmental cue to decompose tasks (create low-level construals) that are immediately relevant.

B. Reinforcement Learning

Reinforcement learning (RL) models are widely used within cognitive science research as a basis for explaining human behavior [21], [22]. Reinforcement learning, especially deep reinforcement learning, is an active area of research [23]. Yet, RL suffers from a number of well-known limitations, most notably the need for extremely large data or training sets, the lack of ability to transfer tasks learned in one context to other contexts, and the general difficulty of designing or learning reward functions [24]. Methods for learning a reward function, such as inverse reinforcement learning [25] are often brittle and require a great deal of data. We believe that CLT can help simplify reward function learning by limiting the space of solutions with abstract construals of related experiences. Finally, CLT applies not only to action decomposition and selection, but is also used to represent knowledge about items, events, and people. Our research will use the techniques from CLT to represent knowledge about items, events, and people to guide behavior.

C. Interpersonal Interaction

Interpersonal interaction often involves planning. Humans may plan and rehearse conversations, generate expectations about upcoming interactions, and select actions, such as white lies, to improve future interactions. CLT posits that as the time for an interpersonal interaction draws near, each person will develop increasingly specific plans meant to shape the interaction ([3], see [26] for a review). High level construals related to interpersonal interaction center on the use of categories to abstractly reason about a future

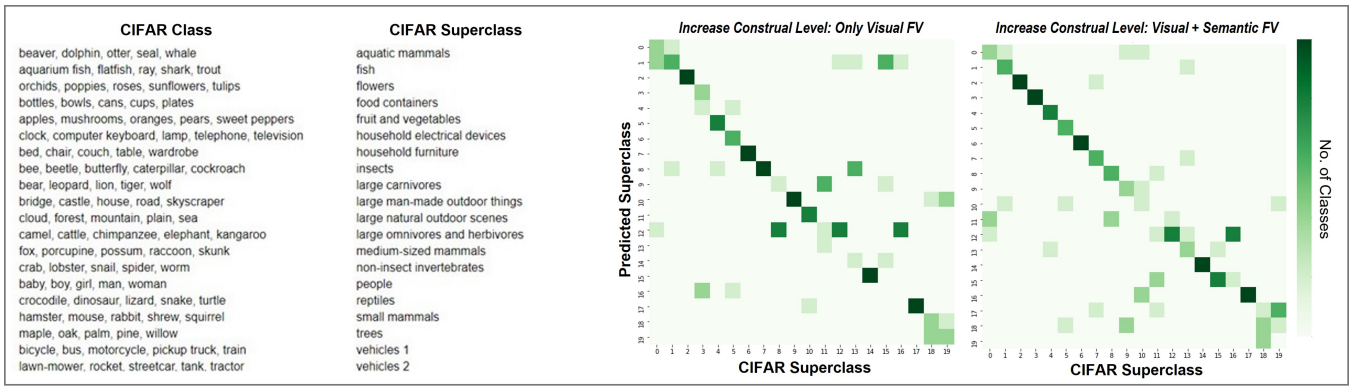

Fig. 2: The lists (left) give CIFAR-100 standard classes and superordinate classes. The confusion matrices depict the CBCL's grouping of classes based on visual features only (middle) and a combination of visual features and semantic embeddings from ConceptNet (right). The horizontal axis depicts the "ground truth" CIFAR-100 superclass and the vertical axis depicts the predicted superclass. The grouping of classes forms superclasses, which represent a higher level of construal.

interaction [27]. Categorical models of people influence a person's evaluations, impressions, and recollections of other individuals [28]. Categorical thinking applied to interpersonal expectations is commonly referred to as stereotyping in the psychological literature [27]. In spite of its negative connotation, categorical or stereotyped models play an important role in deciding who, how, and when to cooperate with others.

III. INITIAL WORK

As a first step towards using construal level theory for human-robot interaction, we are focused on developing methods that will allow a robot to create higher and lower level construals of objects, places and people. Towards this goal, our prior work has investigated methods that use hierarchical clustering of the features generated by a pre-trained Convolutional Neural Network to create clusters representing different categories of objects or places [29], [30].

Inspired by the EpCon model for concept learning in the hippocampus [31], our clustering method initially creates one cluster from the first image of an object. As the system receives additional examples of objects, feature vectors are generated and compared using the Euclidean distance to all of the existing clusters. If the distance of the feature vectors to the closest cluster is below a pre-defined distance threshold, the vector is added to the cluster, otherwise it is used to create a new cluster and centroid representing a new category of object. When presented with a new object, the label for the object is determined by calculating the distance to the known centroid. The label of the closest centroid is used to label the unknown item. We have used this approach to create a state-of-the-art method, termed Centroid-Based Concept Learning (CBCL), for continual learning of objects [30], scenes [29], and have demonstrated that this method can be used by a robot to identify unknown objects [32].

Our prior work has also previously sought to develop methods allowing a robot to create categorical models representing different emergency personnel roles (EMT, police officer, firefighter, etc.) [33], [34]. This prior work

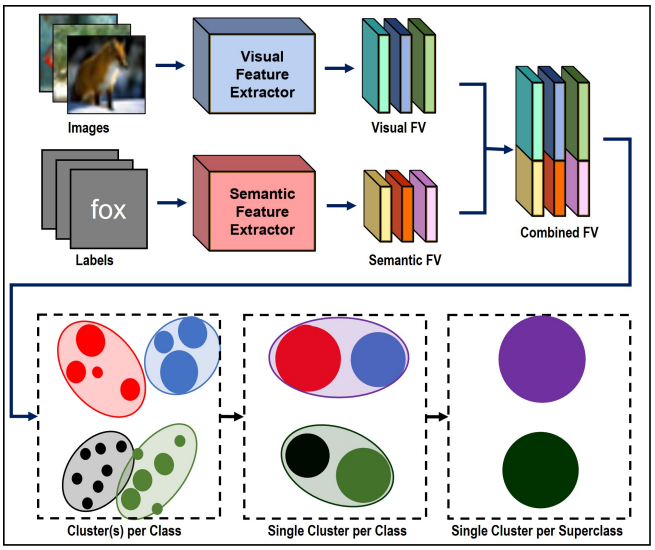

Fig. 3: In the initial framework, CBCL extracts feature vectors (FV), which are then clustered with Agg-Var clustering. Higher-level and lower-level construals can be generated by merging and splitting clusters, respectively.

demonstrated a crude, yet effective, method by which a robot's experiences with different categories of individuals are created by clustering with respect to a person's visible features. The visible features for an emergency worker, for example, tend to be related to their uniform and the resulting clusters were role specific. Moreover, we also captured the individual's behavior in an cooperative game theoretic game. Once enough data about the individuals had been collected, we showed that we could use our process to classify and predict the behavior of newly encountered but unknown individuals [34].

We are currently adapting CBCL to create clusters representing interactions with different people (Figure 3). These clusters should capture categories of people with whom the robot has had interaction. Clusters of models of people

will allow the robot to anticipate and predict the nature of future interactions. Unlike standard hierarchical clustering, the clustering is a dynamic process, increasing/decreasing construal level as a robot traverses through time. As a continual learning algorithm, CBCL enables the robot to incorporate real-time observations into the cluster space.

As mentioned above, the value of CLT is the possibility of using high level construals to anticipate the color and flavor of future interactions. In order to create higher and lower level construals, we are currently testing methods of combining clusters into superordinate classes and of dividing classes into subordinate classes. We have recently tested these methods on the CIFAR-100 dataset which, although object focused, includes one level of superordinate class descriptions (Figure 2). Figure 2 (right) depicts the results from our method in creating increased level of construals (superordinate classes) given visual features only (middle) or visual and semantic features of objects (right). The horizontal axis depicts the "ground truth" CIFAR-100 superclass and the vertical axis depicts the predicted superclass. In both cases, the confusion matrices show reasonably close predictions to the actual superclass groups in CIFAR-100 (Figure 2 left). Note that although CIFAR-100 is an image dataset of objects, clustering over people is fundamentally similar. The merging of classes (types of people) averages distinct class centroids to form stronger superclass centroids, which reflect the shared characteristics (stereotypes) of the new superclass (group of people).

IV. CONCLUSIONS

Although still in the early phases, we hope to use CLT to allow robots to anticipate and predict how to interact with unknown people in new situations. CLT describes a process tying abstraction to distance in the future. We believe that this process can be used to categorize interactions in terms of the participants and with respect to the context. We hope to ultimately connect this process with reinforcement learning and game theory to develop robots capable of interacting with a wide variety of people in a wide variety of situations.

REFERENCES

[1] Y. Trope and N. Liberman, "Temporal construal." *Psychological review*, vol. 110, no. 3, p. 403, 2003.

[2] Y. Trope, N. Liberman, and C. Wakslak, "Construal levels and psychological distance: Effects on representation, prediction, evaluation, and behavior," *Journal of consumer psychology*, vol. 17, no. 2, pp. 83–95, 2007.

[3] Y. Trope and N. Liberman, "Construal-level theory of psychological distance." *Psychological review*, vol. 117, no. 2, p. 440, 2010.

[4] M. B. Chang, S. Levine, A. Gupta, and T. L. Griffiths, "Automatically composing representation transformations as a means for generalization," in *7th International Conference on Learning Representations, ICLR 2019*, 2019, p. 23.

[5] K. Fujita, M. D. Henderson, J. Eng, Y. Trope, and N. Liberman, "Spatial distance and mental construal of social events," *Psychological Science*, vol. 17, no. 4, pp. 278–282, 2006.

[6] C. J. Wakslak, S. Nussbaum, N. Liberman, and Y. Trope, "Representations of the self in the near and distant future." *Journal of personality and social psychology*, vol. 95, no. 4, p. 757, 2008.

[7] B. S. Moore, D. R. Sherrod, T. J. Liu, and B. Underwood, "The dispositional shift in attribution over time," *Journal of Experimental Social Psychology*, vol. 15, no. 6, pp. 553–569, 1979.

[8] N. Liberman and Y. Trope, "The role of feasibility and desirability considerations in near and distant future decisions: A test of temporal construal theory." *Journal of personality and social psychology*, vol. 75, no. 1, p. 5, 1998.

[9] N. Liberman, M. D. Sagristano, and Y. Trope, "The effect of temporal distance on level of mental construal," *Journal of experimental social psychology*, vol. 38, no. 6, pp. 523–534, 2002.

[10] J. R. Busemeyer and A. Diederich, *Cognitive modeling*. Sage, 2010.

[11] J. C. Turner, M. A. Hogg, P. J. Oakes, S. D. Reicher, and M. S. Wetherell, *Rediscovering the social group: A self-categorization theory.* Basil Blackwell, 1987.

[12] A. Fishbach and T. Eyal, "The motivation-emotion matching hypothesis," *ACR North American Advances*, 2010.

[13] K. Christoff, K. Keramatian, A. M. Gordon, R. Smith, and B. Mädler, "Prefrontal organization of cognitive control according to levels of abstraction," *Brain research*, vol. 1286, pp. 94–105, 2009.

[14] J. J. Ribas-Fernandes, A. Solway, C. Diuk, J. T. McGuire, A. G. Barto, Y. Niv, and M. M. Botvinick, "A neural signature of hierarchical reinforcement learning," *Neuron*, vol. 71, no. 2, pp. 370–379, 2011.

[15] E. D. Sacerdoti, "Planning in a hierarchy of abstraction spaces," *Artificial intelligence*, vol. 5, no. 2, pp. 115–135, 1974.

[16] D. E. Wilkins, "Domain-independent planning representation and plan generation," *Artificial Intelligence*, vol. 22, no. 3, pp. 269–301, 1984.

[17] M. Kloetzer and C. Belta, "Temporal logic planning and control of robotic swarms by hierarchical abstractions," *IEEE Transactions on Robotics*, vol. 23, no. 2, pp. 320–330, 2007.

[18] L. Johannsmeier and S. Haddadin, "A hierarchical human-robot interaction-planning framework for task allocation in collaborative industrial assembly processes," *IEEE Robotics and Automation Letters*, vol. 2, no. 1, pp. 41–48, 2016.

[19] C. G. Correa, M. K. Ho, F. Callaway, and T. Griffiths, "Resource-rational task decomposition to minimize planning costs," in *CogSci*, 2020.

[20] L. P. Kaelbling and T. Lozano-Pérez, "Hierarchical planning in the now," in *Workshops at the Twenty-Fourth AAAI Conference on Artificial Intelligence*, 2010.

[21] N. D. Daw and D. Shohamy, "The cognitive neuroscience of motivation and learning," *Social Cognition*, vol. 26, no. 5, pp. 593–620, 2008.

[22] N. D. Daw, S. J. Gershman, B. Seymour, P. Dayan, and R. J. Dolan, "Model-based influences on humans' choices and striatal prediction errors," *Neuron*, vol. 69, no. 6, pp. 1204–1215, 2011.

[23] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 26–38, 2017.

[24] A. Irpan, "Deep reinforcement learning doesn't work yet," https://www.alexirpan.com/2018/02/14/rl-hard.html, 2018.

[25] J. Fu, K. Luo, and S. Levine, "Learning robust rewards with adverserial inverse reinforcement learning," in *International Conference on Learning Representations*, 2018.

[26] D. T. Gilbert, "Ordinary personology," *The handbook of social psychology*, vol. 2, pp. 89–150, 1998.

[27] D. J. Schneider, *The psychology of stereotyping*. Guilford Press, 2005.

[28] C. N. Macrae and G. V. Bodenhausen, "Social cognition: Thinking categorically about others," *Annual review of psychology*, vol. 51, no. 1, pp. 93–120, 2000.

[29] A. Ayub and A. R. Wagner, "Centroid-based concept learning for RGB-D indoor scene classification," in *British Machine Vision Conference (BMVC)*, 2020.

[30] ——, "Cognitively-inspired model for incremental learning using a few examples," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 222–223.

[31] M. L. Mack, B. C. Love, and A. R. Preston, "Building concepts one episode at a time: The hippocampus and concept formation," *Neuroscience Letters*, vol. 680, pp. 31–38, 2018.

[32] A. Ayub and A. R. Wagner, "Online learning of objects through curiosity-driven active learning," in *IEEE RoMan (Workshop on Lifelong Learning for Long-term Human-Robot Interaction)*, 2020.

[33] A. R. Wagner, "Robots that stereotype: creating and using categories of people for human-robot interaction," *Journal of Human-Robot Interaction*, vol. 4, no. 2, pp. 97–124, 2015.

[34] ——, "Using cluster-based stereotyping to foster human-robot cooperation," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012, pp. 1615–1622.