This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

Are We There Yet? Grounding Temporal Concepts in Shared Journeys

Ruth Schulz, Gordon Wyeth, *Member, IEEE* and Janet Wiles

Abstract— **An understanding of time and temporal concepts is critical for interacting with the world and with other agents in the world. What does a robot need to know to refer to the temporal aspects of events – could a robot gain a grounded understanding of "a long journey", or "soon"? Cognitive maps constructed by individual agents from their own journey experiences have been used for grounding spatial concepts in robot languages. In this paper, we test whether a similar methodology can be applied to learning temporal concepts and an associated lexicon to answer the question "*how long*" did it take to complete a journey. Using evolutionary language games for specific and generic journeys, successful communication was established for concepts based on representations of time, distance, and amount of change. The studies demonstrate that a lexicon for journey duration can be grounded using a variety of concepts. Spatial and temporal terms are not identical, but the studies show that both can be learned using similar language evolution methods, and that time, distance, and change can serve as proxies for each other under noisy conditions. Effective concepts and names for duration provide a first step towards a grounded lexicon for temporal interval logic.**

***Index Terms*—cognitive maps, temporal concepts, robots, evolution of language**

I. INTRODUCTION: WHAT NEED A ROBOT KNOW ABOUT TIME?

CONCEPTIONS of time and space are the foundations of an agent's embodied knowledge and grounded language [1]. A naive assumption is that time and space are metric structures of an environment that are intrinsically known by all agents, but this turns out not to be the case. In physics, general relativity theory links time and space in a combined spacetime view [2]. Time and space are not directly perceived and need to be constructed from sequences of perceptions, a fact appreciated by researchers in fields as distinct as gestalt psychology [3], cognitive linguistics [4], geographic information systems [5], cognitive architectures [6], and robotics [7]. Theories of time and its relation to space suggest principled but different ways that temporal terms could be grounded in embodied action. In computational systems two of the most useful and enduring temporal logics have been based on intervals between points in time (called *interval algebra* [8] inspired by representations in linguistics) and states organized into state transition graphs [9] (for a comparison see [10]). In robotics, temporal logics have been proposed for motion planning [11] and spatial reasoning [12]. Note that in the following paragraphs, we do not advocate any single theory, but rather seek to explore the different ways they lead us to think about grounded language learning for temporal terms, in particular ones that can be used in describing a shared journey.

Cultural and linguistic notions of time have a deep literature [13]. Cultures differ in their ways of talking about time [14-16], and psychological theories differ on the relationship between concepts of time and space [17]. Theories include time and space as independent dimensions [18, 19], as inherently combined [2, 20], and views that emphasize both the similarities between time and space and the limits of the similarities [21]. A prominent view from cognitive linguistics is that time is understood in terms of metaphors [22], the predominant one mapping time into space [23-25]. Common mappings from time into space include 'time as a path' and 'life as a journey', with a variety of metaphor types including moving ego and moving time [17], and perspective specific and perspective neutral [26]. However there are limits to the spatialization of time in linguistic metaphors [4, 21]. Galton [21] outlines four key aspects of time, which he calls extent, linearity, directness, and transience, only the first three being properties shared by both time and space. One way of summarizing this view is that time and space are linked by change-based metaphors in which there are both similarities and differences. The duration of journeys enables us to focus on the shared experiences of two agents, rather than the transient qualities of time per se.

A. Biological and robotic representations for time and space

Complementing cultural and cognitive views of time and space are theories about their neural representations. Cognitive maps are represented in the neural circuits in the hippocampus and parahippocampal areas [27]. Less is known about neural circuits for temporal information, although the hippocampus is also implicated in sequence learning [28] and other time codes [29]. In autonomous robotics, such cognitive maps have been used in a bio-inspired navigation algorithm known as

Manuscript received August 20, 2010; revised October 17, 2010; accepted November 26, 2010. This work was supported in part by the Australian Research Council through Discovery Project DP0987078 and the National Science Foundation Temporal Dynamics of Learning Center.

R. Schulz is with the University of Queensland, Brisbane 4072, Australia (phone: +61-7-3365-9765; fax: +61-7-3365-4999; e-mail: ruth@itee.uq.edu.au).

G. Wyeth is with the Queensland University of Technology, Brisbane 4000, Australia (e-mail: gordon.wyeth@qut.edu.au).

J. Wiles is with the University of Queensland, Brisbane 4072, Australia (e-mail: wiles@itee.uq.edu.au).

RatSLAM (rat-inspired Simultaneous Localization and Mapping) [7]. The RatSLAM system uses a graph-based representation of events and transitions between events that would support state-based temporal logic. The time taken to traverse links between locations could be augmented with duration information as the basis for interval-based temporal reasoning. However, as far as we are aware, no temporal logics have yet been applied to language learning with autonomous robots. The technological revolution of the past millennium has provided accurate clocks, so that precise timing can be used in robotics, rather than estimated from experience. However, for robots to interact with humans through a common language for time and space, clock time will not suffice to ground terms like "long" when describing a journey or "soon" when describing an event.

Different types of concepts underpin temporal aspects of language, including duration (a span of time such as an hour, or a day, which can be used to measure the duration of an event or time between two events), points in time (which we call *temponyms*, such as a birth date), and temporal sequencing (including the grammatical notions of tense and aspect). The sets of concepts used to formally represent time form an ontology for the domain of time. Allen's interval algebra [8] has been used extensively in temporal reasoning (for an early review see [30]). Temporal ontologies have been developed in a variety of domains, including the semantic web [31], geographic information systems and database management systems [32, 33], and medical data [34]. Previous studies have developed models for learning temporal knowledge from natural language [35, 36]. However, all the studies just mentioned either hand-code the ontology or learn it from text corpora rather than experiences embedded in time.

A different approach was taken by De Beule [37] who investigated the ability of autonomous agents to develop their own temporal concepts from experience. De Beule studied the formation of a temporal ontology using discriminations required for sequencing of events, based on the evolution of language approach of Steels [38]. To our knowledge, no other studies have been implemented in which agents develop their own concepts for time.

B. Adapting learning of spatial lexicons to durations

While there is existing work on constructing temporal ontologies, as discussed above, and previous studies by our group and others have investigated the evolution of language grounded in real robots using cognitive maps [39-41], an adaptation of existing methodology for grounded concepts of time has not been made to date. Building on the extensive research on spatial concepts, and drawing on similarities between space and time, it is still an open question whether the processes used for spatial learning could be used as a basis for constructing temporal concepts. Note that our position is not that temporal concepts are grounded in space; rather we are looking at how behavior that is shared and named can be used in the formation of temporal concepts. The question is the extent to which the processes and cognitive architectures that have been used for spatial learning can be applied to temporal learning.

Previous studies have indicated that important features of an evolution of language methodology include the grammar of the language game played by the agents, shared attention between the agents about the topic of the language game, and intrinsic representational abilities in the cognitive architecture of the agents (see section II). Previous studies have also highlighted the impact of inherent uncertainty in spatial concepts due to several factors: When both robots are located 'here' they cannot be located in exactly the same position, but are located nearby each other. Robots that build their own maps of the world grounded in their own experience also have intrinsic differences in the maps and in their localization within these maps. The categorization of the continuous spatial world into symbols adds to the uncertainty in the meaning of the spatial concepts, but is necessary for the robots to be able to generalize.

For embodied action, robots need to perform a practical task. To investigate the adaptation of spatial learning to temporal learning, we focus on robots learning concepts and a lexicon to communicate about the duration of a journey, a task that has both temporal and spatial components: When a journey is referred to as a path, spatial aspects are highlighted, when formulated as duration, temporal aspects are highlighted.

To study the formation of duration concepts through a variety of grounding types and underlying representations, we used Lingodroids [42], which are robots with the dual abilities (1) to explore and map their environment using RatSLAM's bio-inspired cognitive mapping system [43], and (2) to evolve a grounded language for spatial terms that include *toponyms* (places in the world) and words for spatial relations between toponyms (directions and distances).

Our overall goal with Lingodroids is to enable a population of robots to explore their world and develop a lexicon that can adequately learn grounded lexicons to refer to all aspects of the world. In this paper, we aim to add to the existing spatial lexicon of the robots with temporal concepts of duration. In the two studies we tested the extent to which the processes and cognitive architectures that have been used for spatial learning could be applied to temporal learning. The studies extend previous work by introducing a new language game, the *how-long* game, in which two robots take a journey together. The agents bootstrap their duration concepts from their experiences and from toponyms grounded in their cognitive maps. In Study 1, a variety of representations for duration concepts was investigated, and in Study 2, the uncertainty inherent in communication was investigated by comparing communicative success for different values of language game factors such as lexicon size and noise.

The following section outlines the principles and origins of evolutionary language games as a methodology for grounding language. We describe our Lingodroids robotic environment, and provide an overview of our previous results in grounding spatial terms. Section III explains the methodology for grounding duration terms and concepts, and describes the six conditions of grounding and representations for duration concepts to be investigated in this paper. Section IV presents the results from the first study which compares and contrasts success of communication using the different duration concepts. Section V presents the second study which investigates why some duration concepts are more effectively communicated than others. Section VI discusses the results and concludes with directions for further investigation.

II. Grounding through Language Games

A. Evolutionary Language Games

Evolutionary language games are a methodology developed for investigating models of language in communities of simulated agents [44]. In a typical language game, two agents choose a common topic of conversation, where the topic is chosen by some mechanism of shared attention. In Steels' seminal Talking Heads project [45], the agents shared attention by focusing cameras on the same point in space, albeit from slightly different perspectives. Once attention has been appropriately focused for both agents, one will generate an utterance to describe the subject at hand, while the other listens (and sometimes responds) to that utterance. The process is repeated, with attention focused over a range of concepts, and utterances exchanged by both agents. Over time, through an adaptive process based on the exchange of utterances, both agents form a shared lexicon that describes the range of concepts experienced by the agents.

The methods for achieving shared attention and perspective alignment affect how concepts are grounded in the agents' lexicons. While humans seem to naturally establish shared attention, it can be challenging to establish shared attention and perspective alignment between embodied agents [46, 47]. When working in simulation, direct measurements of proximity can be trivially obtained from the simulation engine [48]. For real robots, shared attention may involve looking at the same scene [49], or by determining proximity using algorithms for agent and object recognition, possibly aided by markers as in [50]. The use of an audible handshake can provide a combination of audible and visual cues [51]. Overall, when determining a common topic of conversation, possible methods and their relative difficulty depend on the complexity of the world (the range of possible targets) and the abilities of agents for synchronizing aspects of their behaviors (shared attention, perspective alignment, other agent recognition).

Using language games, autonomous robots create the semantic categories that are natural to their embodiment and ground these categories in a language that they evolve themselves. The simplest of the grounding problems, that of direct grounding of percepts (or directly perceived concepts), has been declared 'solved' by Steels [52], albeit with limits on the solutions proposed. He described embodied agents that evolve their own categories and languages, and defined grounding in terms of a method that links symbols and objects. If a method exists that can classify the perceptual information and decide if a concept is appropriate, then the symbol associated with the concept is grounded. Beyond the perceptual categories that form the focus of Steels' work, concepts such as time and space cannot be grounded directly in sensorimotor actions, rather requiring the robots to develop internal representations with which they can refer to events, locations and times that are not currently experienced. Lingodroids (see next section) represent such information in experience maps inspired by the cognitive maps of the hippocampus [27, 43].

B. Lingodroids Robotic Environment

The Lingodroid project builds on previous work in which Pioneer 3 DX robots (see Fig. 1) were endowed with the robot equivalent of a hippocampus in a fully functional navigation system called RatSLAM [53]. In addition to their RatSLAM mapping capabilities, we enhanced the robots by equipping them with a microphone and speakers that enable them to communicate using mobile phone tones. Lingodroid tests can be conducted with the real robots and in simulation. The studies in this paper used high fidelity simulations of the robots that have been used extensively over several years, demonstrating the reliability of the simulation world [40, 54]. The simulation world was built to mirror the real world, with images from the real world used in constructing the views of the robot (see Fig. 2). The simulation world enables simulated robots to pass messages to other robots within a set distance of their current locations, allowing the hearing distance for audible communication to be explicitly set.

Fig. 1. In the real world studies, the Lingodroid robots are Pioneer 3 DX robots fitted with an omni-directional camera, sonar range finders, laser range finders, wireless communication, a microphone, and two speakers.

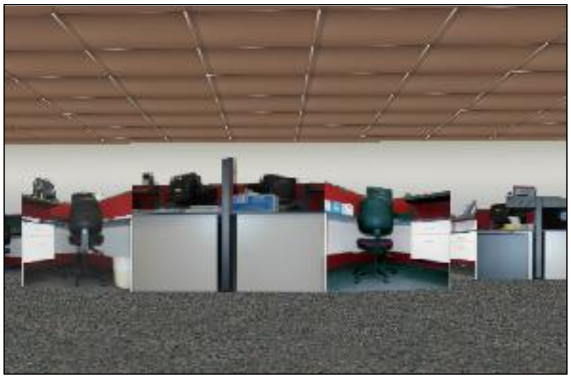

Fig. 2. In the simulation world studies, the agents experience a 3D virtual reality environment that closely mimics the robot's experience. The virtual environment is constructed from digital photographs of the office environment mapped on to a detailed 3D structure.

Lingodroid experiments take place in an open plan office environment featuring an array of desks, chairs, partitions, and breakout tables, as well as the usual personal items associated with a busy working office. The robots move freely around the office environment of our university research group, planning paths and autonomously navigating to locations in corridors and offices, avoiding chairs, people, and other paraphernalia. They are able to determine when their batteries are running low and return autonomously to their charger. When given no other task, they will explore their environment, building and adjusting their internal map of the world [55].

The RatSLAM system in each robot builds a cognitive map representation that is termed an *experience map*. Over time, the robots accumulate knowledge about a set of visual experiences organized in space based on motion information. The map can be formally described as a graph: The nodes of the graph (called *experiences*) hold information about the visual scene (*local views*) and locally consistent pose (*local pose estimate*) at each location; the edges hold information about links between locations derived from motion (*odometry*, *travel time*, and *behavior*). Local views and local pose estimates are organized into a globally coherent map (see Fig. 3).

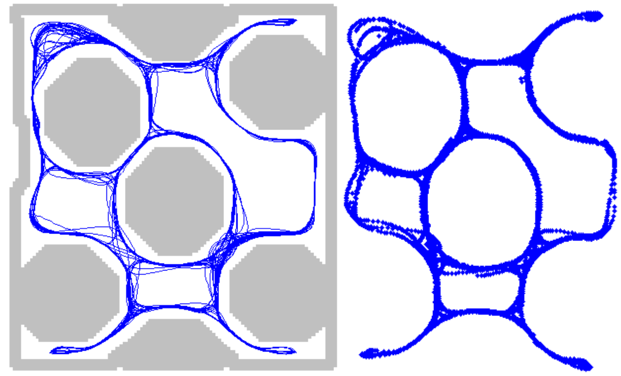

Fig. 3. The office layout consists of walls (grey perimeter) and desks (grey octagons). The left hand plot shows the path of a robot in a typical simulation run (thin lines). The right hand plot shows the resulting experience map with the clusters of experiences (small dots).

C. Spatial Language Games with Lingodroids

Location language games were developed to form and use spatial languages. A *where-are-we* game (see Fig. 4) is played when the agents are near each other, defined as being within hearing distance of each other. The topic of the game is the current location of the agents, with the speaker producing a word for the current location. The hearer updates its lexicon based on the speaker's utterance, and sends an acknowledgement. The speaker then also updates its lexicon. The net effect of a series of *where-are-we* games is a shared toponymic language describing names for places in the world in which the two agents have interacted. High coherence can be achieved for the toponymic languages, but there is inherent uncertainty in the locations corresponding to toponyms due to the hearing distance for deciding when to play games, and the neighborhood size used during word production. Coherence can be tested using *go-to* games (see Fig. 5).

A *go-to* game begins when agents are near each other. The speaker decides on a goal word. The hearer determines whether the goal word is in its lexicon and both speaker and hearer independently determine whether a path can be planned to the location that is the best example of the goal word. If both agents can plan a path to the goal location, both agents then move to the goal location, otherwise, the *go-to* game does not continue. The agents use their own internal maps to navigate to the goal, continuing to play *where-are-we* games along this path when they are within hearing distance. Once the agents reach the goal location, each agent announces its arrival, waits to hear the other agent then repeats the announcement. The *go-to* game provides an indication of the coherence of the languages: If the agents comprehend each utterance similarly, they will meet each other at the goal location specified in each game. Details of our previous work with spatial language games can be found in [40, 54].

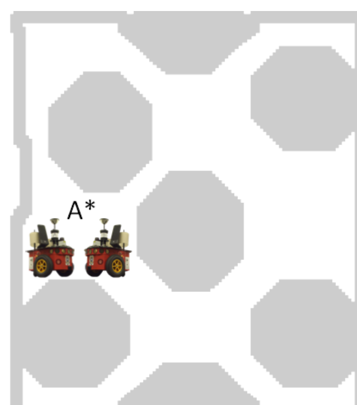

Fig. 4. In the *where-are-we* game the robots establish shared attention through proximity and the speaker determines the best toponym for the current location (A). The star (*) indicates that the speaker may invent a new word, and that both agents will update their lexicon.

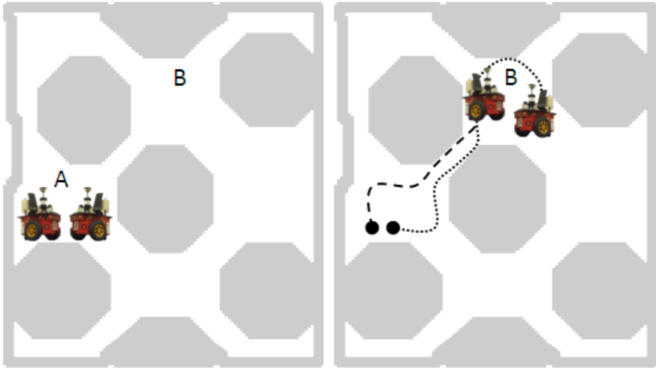

Fig. 5. In the *go-to* game the robots establish shared attention through proximity at the current location, A, and the speaker determines the best toponym for a second, goal location (B). The robots independently navigate to meet at the goal. On reaching the goal, each robot announces its arrival, waits for a similar announcement from the other robot, and then repeats its own message, thereby ensuring that both robots know of the success of the game. The lexicon tables are not updated in a *go-to* game.

III. Methodology for Grounding Duration Concepts in Shared Journeys

A. Language Games for Space and Time

A new game was developed for studies of journey durations, called *how-long* in which agents name the relationship between the start and end locations of a journey, based on time, space or amount of change experienced. In order to play *how-long* games, *where-are-we*, and *go-to* games were required to specify the journey. *Where-are-we* games were played as in our previous studies to enable the agents to develop a shared toponymic lexicon [40]. Following the *where-are-we* games, *how-long* games were played together with *go-to* games to form a duration lexicon. The *go-to* games were played as in our previous studies. Once at the goal location, the agents announced that they were at the goal, waited to hear the announcement of the other agent, and then re-announced that they were at the goal. At this point, the *how-long* game was played, with the agent that specified the goal location naming the duration relationship between the two toponyms specified in the *go-to* game. Together, the three games allowed the agents to develop fully grounded words for toponyms and durations.

B. Grounding and representation conditions

"How long did it take?" is a question with many answers in natural languages and also when framing a robot language game. Two factors were of particular practical interest for the robots: firstly, a journey between two locations can be experienced as a specific journey (an *instance*) or generalized from memory (a *prototype*); and secondly, the length of a journey can be specified in different ways, relating to the *time* taken, the *distance* traveled, and the amount of motion (represented by *change* in the visual input). Concepts that represent the duration of a journey can be constructed from any combination of these factors, resulting in six possible conditions for a *how-long* game (see Fig. 6). The six conditions were named as follows:

1. Prototype – Distance
2. Prototype – Time
3. Prototype – Change
4. Instance – Distance
5. Instance – Time
6. Instance – Change

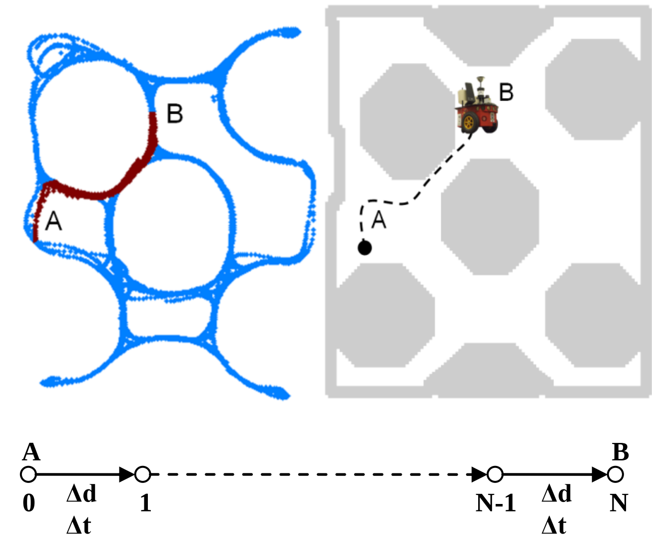

Fig. 6. In the *how-long* game the robots ground their duration concepts in six different ways corresponding to two different grounding types (prototype and instance) and three different representation types (distance, time, and change). In the prototype grounding conditions (top left) the robots use their experience maps built over many journeys through the world. In the instance grounding conditions (top right) the robots use the actual path taken in a *go-to* game. In both grounding types the robots calculate the representations of distance, time, and change from the sequence of experiences and links between experiences in the path (bottom): Distance and time are summed over the corresponding measurements (Δd and Δt respectively) between experiences; and change is calculated as the number of experiences on the path.

To calculate the concept types grounded in the prototype (1-3), the agents use their cognitive maps to consider the shortest path that can be planned between the two points specified by the toponyms. For the shortest path found, the agents calculate the distance of the path (1), the time taken to traverse the path (2), and the number of experiences that make up the path from the start to the goal location (3). The similarity of the representations between the robots depends on the similarity of their experiences when they were exploring the world and when they formed their toponym lexicon table.

To calculate the concept types grounded in an instance (or specific journey) (4-6), the agents record the details of a journey as a game is being played. From these details they track the distance travelled between the start and goal locations (4), the actual clock time taken from the start of the game until the agents meet at the goal location (5), and the number of experience transitions made between the start location and the goal location (6). These measures are from a direct experience recently shared by both robots. They are similar to the prototype estimates (1-3) in that they are based on the travel time along a path, but differ in that they reflect a shared experience. The time value is measured by how long it takes the robots to meet at the goal location, and as such is expected to be less noisy than all other grounding and representation types.

In both grounding types, the route between the two locations may be different, as each robot may find a different shortest route depending on the exact locations of the start and end of the journey. When the routes are different, the values for distance, time, and change may be substantially different.

Note that in order to study all six conditions for the different duration concept types, six simulations were conducted in parallel using the same training and test data, with each simulation involving a single type of duration concept.

C. Concept Elements

In order to play *how-long* games, for the studies presented in this paper, the Lingodroids were augmented with an ability to create *duration elements* to represent durations between pairs of experiences. Each time they experienced a duration in a *how-long* game, they compared that duration to previously experienced durations. If the current duration did not match any stored durations, then a new duration element was created. In this manner, duration concepts elements were created as they were experienced, and the resulting elements covered the space of experienced durations.

D. Distributed lexicon table

In Lingodroids, words are associated with concept elements using a many-to-many mapping system called a *distributed lexicon table*. The data structures of a distributed lexicon table include the lexicon of words, the set of concept elements, and the associations between the words and the concept elements, stored in the lexicon table. In the toponym and duration lexicons, the primitive concept elements were experiences and duration elements. Experiences were created and added to the experience networks as the agents explored the world. Duration elements were created as needed during *how-long* games.

The associations between primitive concept elements and words in distributed lexicon tables were strengthened whenever the concept elements and words were used together. In the *where-are-we* game, both agents incremented the association between the current experience and the word used by the speaker. In the *how-long* game, both agents determined the duration between the two toponyms specified by the closest duration element, D, and calculated the new association, a_{ij}^{D}', between the duration element, i, and the duration word, j, used by the speaker, as follows:

$$a_{ij}^{D}{}' = a_{ij}^{D} + 1 \quad (1)$$

where a_{ij}^{D} is the old association between the duration element, i, and the duration word, j.

For word production, all of the primitive concept elements within the neighborhood of the current element were considered. A confidence value was calculated, indicating the confidence that the chosen word referred to the current situation. For the *where-are-we* game, the current situation was defined as the current location of the agents and the word chosen was a toponym. For the *how-long* game, the current situation was defined as the duration between the two toponyms specified, and the word chosen was a duration word. The confidence value, h_{ij}, for word, j, at concept element, i, was calculated as follows:

$$h_{ij} = \frac{\sum_{m=1}^{Y} \frac{a_{mj}^{D}(D^{D} - dist_{mi}^{D})}{D^{D}}}{\sum_{n=1}^{N} a_{nj}^{D}} \quad (2)$$

where D^D is the neighborhood size, Y is the number of concept elements in the neighborhood of element i, N is the total number of concept elements, and a_{ij}^{D} is the association between duration element.

For word comprehension, all of the primitive concept elements that had been associated with the current word were considered. In the studies presented in this paper, word comprehension was used to determine the current situation described in the *how-long* games, with the location determined for each of the specified toponyms. The location calculated for a toponym was a location in the experience map of the agent, using the global estimates of location for the experiences.

Words were invented probabilistically, using an annealing framework. The probability of word invention depended on the temperature and the confidence value of the concept element – word combination as follows:

$$p = k \exp\left(\frac{-h_{ij}}{(1-h_{ij})T}\right) \quad (3)$$

where k = 1, h_{ij} was the confidence value of the concept element-word combination, and T was the temperature. The temperature influenced the word invention rate, and was decreased linearly from 0.1 to 0.0 over the course of the interactions. By the end of a trial, agents only invented words if there was no existing word for the situation.

E. Communicative Success

The communicative success for a duration concept considered how far away the interpretation of the concept was from the actual concept that the agent was trying to convey. Communicative success, s, was calculated for each interaction, with the actual duration experienced by the agent, d_1, compared to the duration interpreted by the word used for the interaction, d_2, as follows:

$$s = 1 - \frac{abs(d_1 - d_2)}{d_1 + d_2} \quad (4)$$

The communicative success was averaged over all the interactions to obtain the communicative success for the whole lexicon. The success was calculated for both the training set over which the lexicon was created (called the *familiar* environment), and the test set, which was a separate set of *how-long* games not used for the construction of the lexicon (called the *unfamiliar* environment). Communicative success was determined for each agent as listener, based on how well the agents understood their own use of the lexicon (called *self*), and how well the agents understood each other's use of the lexicon (called *other*).

IV. Study 1. Grounding duration terms in *how-long* games

A. Aims

The aim of Study 1 was to demonstrate the grounding of duration terms for a journey using the Lingodroid robots, and to compare and contrast duration types that differed in aspects of grounding and representation.

The study design comprised six conditions (see Fig. 6), one for each duration type {2 grounding types × 3 representation types}. We predicted that all six duration types would enable the development of useful lexicons, with time for a specific journey (#5 Instance – Time) likely to result in the highest communicative success due to the precision afforded by the use of clock time combined with a specific instance of a journey. The analysis was designed to enable comparison of the relative successes of the different conditions.

B. Methods

The study consisted of 10 trials per condition. A unique set of games was used for each trial in the Training Phase, and a common set of games in the Testing Phase.

Training Phase (familiar environment): The agents first developed a toponymic lexicon by playing 500 *where-are-we* games, and then developed a set of duration lexicons by playing 250 *go-to* and *how-long* games. In each *how-long* game, six conditions were tested, corresponding to the six types of duration concept elements used for grounding. The neighborhood sizes used for word production for each type of duration concept element were set to achieve lexicons with an average of six to eight words. The pairs of toponyms used for the start and finish of each of these games were recorded (termed the "training data"). At the end of the training phase, communicative success was calculated for all the durations in the training data.

Testing Phase (unfamiliar environment): To test the duration words in a novel environment, the agents retained their duration lexicons but were moved to a novel environment where they played an additional 500 *where-are-we* games to learn a new set of toponyms. Learning was then disabled and communicative success was measured on 250 new *go-to* and *how-long* games (termed the "testing data"). All ten trials were evaluated on the same test data.

Theoretical limit on communicative success: Since the agents developed individual experience maps, duration concept elements and lexicons, the communicative success was bounded by the degree of similarity between the agents' representations, and could be calculated directly from the agents' representations. The theoretical limit is a measure of the uncertainty of the representation, calculated by directly comparing the durations used by each agent for each game of the data set. This theoretical limit was calculated for the test data and used to determine the relative performance of each concept type (calculated as the ratio of the empirical communicative success divided by the theoretical limit).

C. Results and Discussion

1) How well did the different duration concepts support communication between the agents?

The ten trials created ten distinct lexicons per duration condition. The lexicons within each condition were similar in terms of number of concept elements, with a range of 25-35 concept elements per duration type. The number of duration words varied from 3-10 in each condition.

Communicative success varied across the different conditions in the study, ranging from 0.53 (#3) to 0.88 (#5) (see Fig. 7). These differences reflect different sources of variability in the robots' concept maps and individual experiences in the simulation world. Some of these differences are inherent in the world itself, and are reflected in the theoretical limits for each condition (see Fig. 7 Limits). Others are due to the interaction of the robots with the world and with each other through the shared attention and language evolution process.

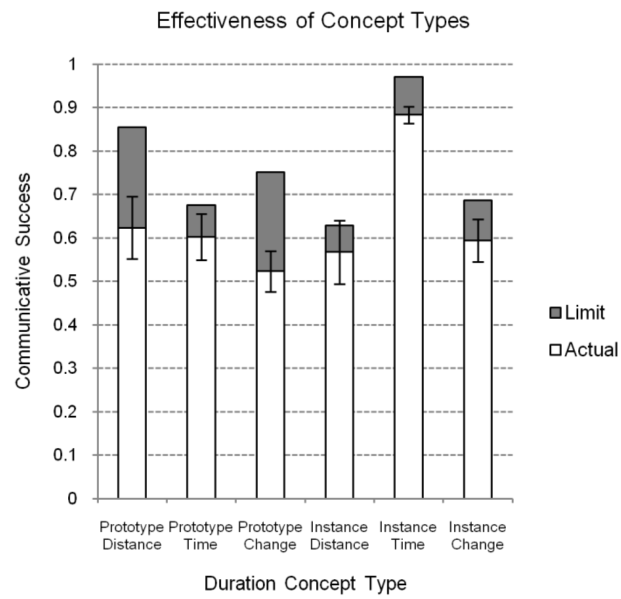

Fig. 7. Effectiveness of the six types of duration concepts when tested in the unfamiliar environment. Averages are over the ten trials, with error bars showing standard deviations.

The relative ranking of performances based on the theoretical limit and actual values is shown in Table 1. The similarity in these rankings suggests that much of the variation was due to factors related to the underlying duration concept elements. As expected, the highest communicative success in both measures – theoretical limit and empirical outcome – were recorded in condition #5 Instance – Time.

The remaining five conditions all resulted in lexicons with some communicative utility but much lower average communicative successes (0.53-0.62). With respect to their theoretical limits, all the types of duration concepts resulted in languages that achieved communicative success in the range 70-90% of their theoretical limit. These types of duration concepts are based on the experiences and paths in the experience map or directly traversed during the *go-to* games. Unlike the estimates of metric time (#5), paths are subject to a variety of sources of noise including cases where agents travelled along different routes during the games, or failed to initially recognize a target toponym and created an extra loop along a path. Actual paths are highly reliant on the agents' experiences in the world.

TABLE 1
RELATIVE RANKING OF COMMUNICATIVE SUCCESS FOR DURATION CONCEPTS

Rank	Theoretical limit	Actual (averaged over 10 trials)
1	6 Instance – Time	6 Instance – Time
2	2 Prototype – Distance	2 Prototype – Distance
3	4 Prototype – Change	3 Prototype – Time
4	7 Instance – Change	7 Instance – Change
5	3 Prototype – Time	5 Instance – Distance
6	5 Instance – Distance	4 Prototype – Change

2) Was the relative order reflected in the structure of the lexicons for each duration concept type?

The structure of a representative duration lexicon was examined to see how the different duration words spanned the space of possible durations, and how closely the structure was mirrored between the two agents. Fig. 8 Rows 1-6 show the lexicons for each of the different concept types. In each graph, the *x*-axis shows the duration value to be expressed (from minimum to maximum journey durations experienced), and the *y*-axis shows the relative production score for each word. The word with the highest relative score was selected when that agent was speaking. The lexicons shown comprise 5-9 words, and are typical of the different types of duration concepts.

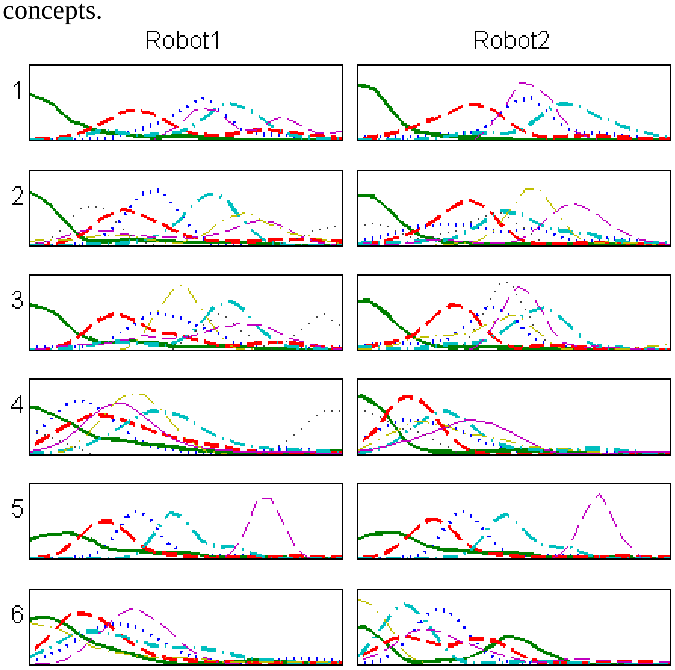

Fig. 8. Example lexicons for a representative trial. Each column shows word profiles of the lexicon for one agent. Words in each column are matched for line style.

Clear structure can be seen in the lexicons for the highest performing type (#5 Instance – Time). The words are well separated, they span the space of possible concepts effectively, and both agents have very similar lexicon structures (see Fig. 8 row 5). By contrast, the lower performing duration types have much less well organized lexicons. The ranges overlap with both broad and focused terms, the two agents have lexicons organized in different ways and the individual words have higher variance in meaning.

The results as a whole show that the ability to evolve a coherent well organized lexicon depends not just on the interaction between the agents, but also the underlying types of duration concepts. It may be that the lexicons in the lower performing conditions are simply too large to enable a coherent set of concepts and lexicon to bootstrap in conditions provided. The second study explored the issue of whether lexicon size could affect the communicative success of the evolved languages by considering the different factors that interact with lexicon size.

V. STUDY 2: FACTORS AFFECTING THE EFFICACY OF DURATION CONCEPTS

1) Could lexicon size affect the communicative success of the evolved languages?

The example lexicons developed in the first study showed clear differences in the distribution of word profiles. Condition #5 Instance – Time with the highest communicative success had words with meanings evenly distributed across the space of possible durations, and both agents learned similar word profiles. Conditions with lower communicative success frequently had two or more words with similar word profiles, some overlapping profiles, and clear differences between the meanings of the robots' lexicons.

The size of a lexicon and meaning of each word depends on aspects of word choice and word invention during the production of a word in the *how-long* games. The number of words in the lexicon can be altered by varying the effective window size of each word, which determines the neighborhood of concept elements surrounding a word.

Increasing the specificity of each word has the potential to increase the communicative efficacy of a language. However, additional precision in word meaning is only useful if shared attention can be similarly precise, otherwise larger lexicons are created with divergent meanings for the words and communication fails. The different types of duration concepts tested in Study 1 varied in the noise inherent in shared attention and individual maps of embodied agents, which would be expected to impact on the ability of the agents to converge on shared meanings. Hence, interactions could be expected between the precision of word meanings, size of the lexicons and the communicative efficacy of each condition.

2) Factors that interact with lexicon size

The noise inherent in each duration concept condition is a property of all the facets of the language game and cannot be directly manipulated. However, there are measures that can be taken during the simulations that can be used as proxies for different sources of noise in the study.

Study 1 reported communicative success *between the agents* using the duration lexicon in an *unfamiliar environment*. Both these factors can be tested independently and together:

By calculating the communicative success an individual agent would have with understanding their own utterances, the role of differences between the private representations of the agents can be assessed. This measure is indicative of (albeit not identical to) the performance that could be expected if the robots had unlimited opportunity to develop highly coherent maps and duration concepts.

The choice of unseen test data (in this case an *unfamiliar* environment) is used as a gold standard for evaluation in machine learning studies. However, it is rare for evolution of language studies to use such a stringent generalization test. An alternative is to test the communicative success in the environment in which the agents had developed their lexicons (termed the *familiar* environment).

A. Aims

The aim of the second study was to investigate whether communicative success could be improved by varying factors that affected the level of noise in the underlying representations. We conjectured that influences on the communication efficacy could include lexicon size, the degree of shared attention, and the familiarity of the environment. In particular the following questions were formulated:

- Would a larger (or smaller) lexicon be more useful?
- How much better would a robot understand its own utterances compared to those of the other robot?
- What impact does familiarity with the environment have on the utility of a duration lexicon?

The specific prediction was that communicative efficacy would improve as lexicons increased in size for conditions with low noise, and would not improve or would reduce in conditions with higher noise or inability to precisely share attention.

B. Methods

The second study was based on the same methods as Study 1. In addition to the medium sized lexicon of Study 1, two additional sets of simulations were conducted (giving a total of three different lexicons sizes): small (1-5 words), medium (3-10 words), and large (6-19 words) lexicons. The neighborhood size used for word production was adjusted to achieve these lexicon sizes. The same set of training and test games from Study 1 was employed. Thus, the study comprised 2 grounding types {prototype, instance} × 3 representation types {distance, time, change} × 3 lexicon sizes {small, medium, large}, with 10 trials per condition, resulting in a 2 × 3 × 3 design (18 conditions). Measures were taken for 4 test cases per condition for Listener {*other agent, self*} × Environment {*unfamiliar, familiar*}.

C. Results and Discussion

1) Effects of listener and familiarity with the environment

The overall effects for the familiarity of the environment (*unfamiliar* or *familiar*) and for the type of listener (*other agent* or *self*) was assessed by collating measures of communicative success over all 18 conditions. In the most stringent case (listening to the other agent in the unfamiliar environment) the average communicative success was 0.63. Communicative success was also measured in the environment in which the words were first learned (termed the *familiar environment*), when a robot needed only to understand its own lexicon (termed *self*) and in the combined case of self-understanding in the familiar environment. All these changes result in higher ratings: Communicative success was 9.7% higher in a familiar environment, 17% higher when the agent was listening to itself, and combining the two conditions, it was 25.2% higher (0.63 to 0.79), see Table 2.

TABLE 2
IMPACT OF ENVIRONMENT AND LISTENER ON COMMUNICATIVE SUCCESS AVERAGED OVER ALL TYPES OF DURATION CONCEPTS (#1-6) AND LEXICON SIZES

		Listener: Other agent	Listener: Self	Gain
Environment	**Unfamiliar (test)**	0.6324	0.7402	17.0%**
Environment	**Familiar (train)**	0.6936	0.7920	14.2%**
	Gain	9.7%*	7.0%*	25.2%***

* Gain = (Train – Test) / (Test)
** Gain = (Self – Other) / (Other)
*** Gain = (Train Self – Test Other) / (Test Other)

2) Trends within each case: When is a larger lexicon more useful?

Within the four evaluation cases {*Case 1 Unfamiliar-other, Case 2 Unfamiliar-self, Case 3 Familiar-other, Case 4 Familiar-self*} the individual trends for each type of duration concept were grouped according to lexicon size (see Fig. 9-Fig. 12). In all four cases, condition *#5 Instance – time* showed a clear difference between the smallest and the larger two lexicon sizes. The remaining conditions did not show obvious differences, but trends from the smaller to larger lexicon sizes can be observed in cases 2 and 4, both cases in which an agent was communicating with itself. These cases address aspects of the three questions set out in the section above.

3) Would a larger (or smaller) lexicon be more useful?

Trends towards larger lexicons being more useful were observed in 17/24 of the conditions and cases recorded (2 grounding types × 3 representation types × 4 test cases), however, the occurrences were highly skewed towards when the agent was listening to itself.

Case 1. Unfamiliar-other – Communication between the agents in an unfamiliar environment (see Fig. 9): In the standard test environment, there was no consistent trend. The two conditions that directly used clock time (#2 Prototype – Time and #5 Instance – Time) showed a trend towards improved performance for larger lexicon sizes, but the other concept types show no improvement or a decrease in communicative success as the lexicon size increased.

Case 2. Unfamiliar-self – Communication by an agent with itself in an unfamiliar environment (see Fig. 10): When interpreting its own lexicon, 5/6 concept types showed a trend towards improved communication with larger lexicons (all but #1 Prototype – Distance).

Case 3. Familiar-other – Communication between the agents in a familiar environment (see Fig. 11): When communicating with each other in the familiar environment, 4/6 concepts types showed a trend towards improved communication with larger lexicons (all but #4 Instance – Distance, #6 Instance – Change).

Case 4. Familiar-self – Communication by an agent with itself in a familiar environment (see Fig. 12): All 6 measures improved with larger lexicons when the agents were interpreting their own lexicons in the familiar environment.

In subsequent simulations (data not shown), single word lexicons formed in a *how-long* game were tested for their communicative efficacy. Surprisingly, even a single word lexicon for duration was useful for robots communicating in both unfamiliar and familiar environments. We conjecture that a single word lexicon could be acting as a robotic equivalent of “soon” when the underlying concepts have an inherently high variance (such as a standard answer to the generic question from children on a long journey “Are we there yet?”).

4) How much better would an agent understand its own utterances compared to those of the other robot?

This question was answered by testing how well the agents understood their own lexicon. There was a clear effect of higher performance when the agents were listening to themselves rather than the other agent.

This result is interesting as it may point to a difference between the value of larger lexicons in communicative success between experts, who could be expected to have similar lexicons and an ability to reach precise shared attention, compared to novices who have more general concepts with only approximate shared attention.

5) What impact does familiarity with the environment have on the utility of a duration lexicon?

All types of duration concept resulted in higher communicative success in the familiar environment in which the duration words were first learned compared to the unfamiliar test environment.

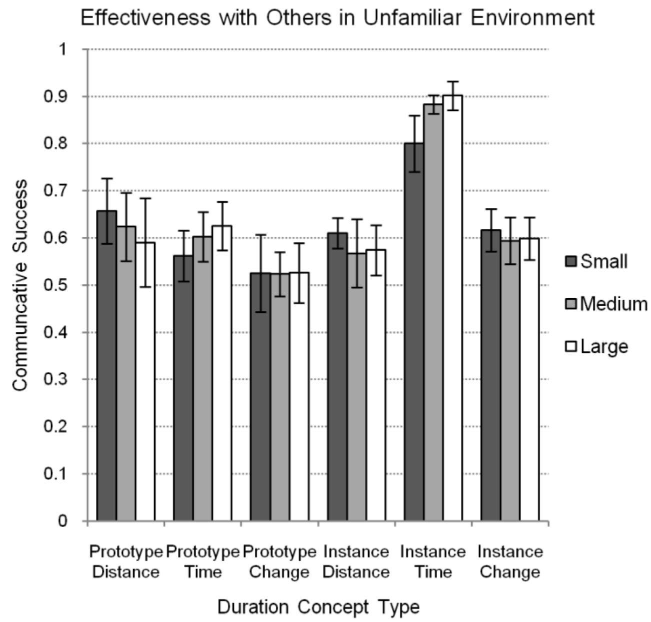

Fig. 9. Case 1: Communicative success between the agents in an unfamiliar environment. The six groups correspond to the different types of durations (as in Fig. 7). The three bars per group correspond to the lexicon sizes: small, medium, and large. The communicative success for #5 increased with lexicon size, whereas other concept types show no improvement or a decrease in communicative success as the lexicon size increases.

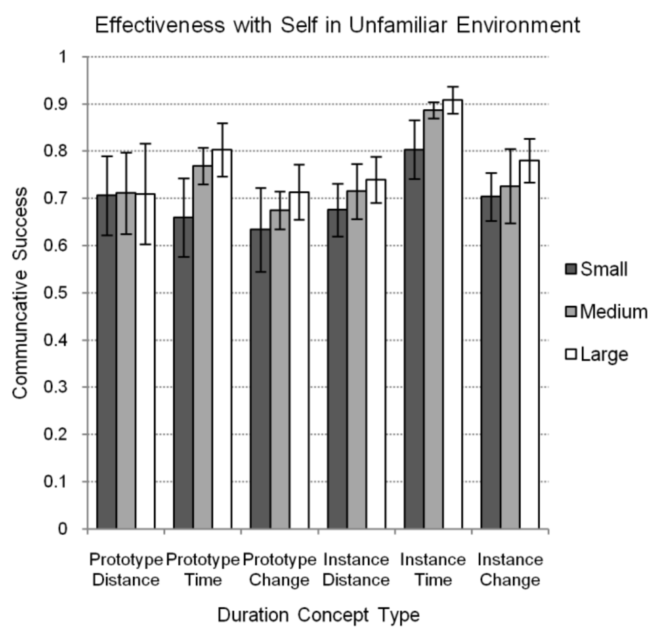

Fig. 10. Case 2: Communicative success by an agent with itself in an unfamiliar environment.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

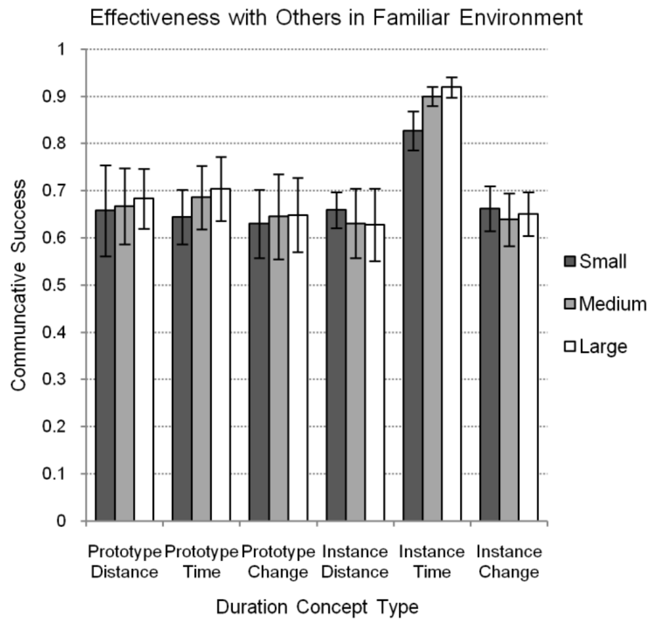

Fig. 11. Case 3: Communicative success between the agents in a familiar environment.

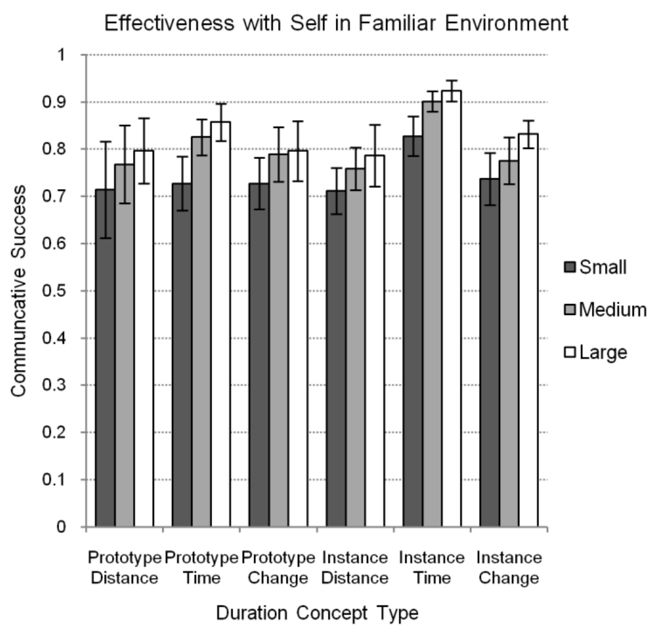

Fig. 12. Case 4: Communicative success by an agent with itself in a familiar environment. Note the high performance for all concept types.

In summary, smaller lexicons had the same or higher communicative success than larger ones when the robots were communicating with each other in unfamiliar environments for all types of duration concept except when clock time was used. The trends in lexicon sizes suggest that larger lexicons were most useful when the robot was using the duration words in an environment in which they were formed, and when they had direct access to their own grounded terms. For all types of duration concept, communicative efficacy improved from the unfamiliar environment of the test data to the familiar environment of the training data, and from the standard communication between agents to the agent understanding itself. It should be noted that the results presented in this paper show trends in the data and larger studies would be required to present measures of statistical significance.

VI. Discussion

"*Quid enim est tempus?*" (What then is time?)
St. Augustine

Time and space are foundational concepts in any language, and evolved robot languages are no different. The Lingodroid languages were originally developed using concepts for space. The studies in this paper have added the first temporal terms to the robots' lexicons.

St Augustine's question was formulated before the invention of reliable clocks, and may seem to have a scientific answer in terms of atomically-precise clock time. However, clocks provide reliable references only for points in time. To understand and refer to events and the motion of objects, languages need to refer to extended periods of time. The value of durations as the first terms in the Lingodroids' temporal lexicon is highlighted in the widespread adoption of temporal interval algebra [8] in logic and reasoning.

Study 1 demonstrated two important points: Firstly, the grounding methodology developed for learning spatial lexicons can be extended to learning temporal lexicons; and secondly, the *how-long* game enables the grounding of a range of different concepts for the duration of a journey. A journey inherently contains aspects of time, distance, and motion. Because the robots have embodied actions and can autonomously move between locations, the *how-long* game enables them to accrue measures of the time taken, the distance traveled, and the amount of sensory change as they travel. A term can refer to a specific instance of a journey, or a prototypical journey, and duration concepts were supported for both types of duration. All the duration concepts tested in Study 1 enabled effective communication, however differences were observed. The duration concept based on time for a specific instance of a journey had the highest communicative success. While the robots in the studies constructed their representations for space using experience maps, they measured elapsed time using clocks. However, the robots cannot have identical concepts for duration. Each robot's perception of the duration of a journey is inherently noisy. The start and end times can differ slightly for each agent, agents can travel by different paths reaching the goal at different times, and they have different internal maps based on their personal experiences.

Study 2 demonstrated that communicative success was affected by several factors: the size of the duration lexicons interacted with familiarity with the environment, and whether a robot was listening to its own, or the other robot's utterances. Large lexicons were useful in familiar environments and when a robot was interpreting its own utterances. Smaller lexicons were more useful when the robots were communicating with each other in unfamiliar environments. These factors all impacted on the precision with which the shared lexicon was associated with the individual robots' concepts. The results highlight the importance of shared experiences, not just shared occupancy of a world. The critical aspect of grounded language learning highlighted in the Lingodroids project (as in all other fully embodied language learning studies) is the necessity for shared experience and shared attention. Precision is key to their effective use in communication. Given that time, space, and change duration concepts can all be used to develop a grounded duration lexicon, an interesting question to ask in future studies is what would happen if the robots are required to make their own selection of which concept to use in each game. To what extent can communication succeed when shared terms are grounded in disparate concepts? Some communicative success is likely, but based on the results in Studies 1 and 2, effective communication is also likely to be affected by the size of the lexicon, expertise in the concepts and familiarity with the environment. Terms in small lexicons that represent concepts such as "soon" may be effectively used in similar situations to "near". It is an open question what would happen for larger lexicons.

The current studies can be extended in several directions: The robots need to have ways of directing each other's attention to aspects that are of interest for naming games (a point highlighted recently in the use of prenominal and postnominal sentences and children's learning of color words [56]). The Lingodroid Project is concerned with how robots can have the intrinsic abilities to learn spatial and temporal concepts from any natural language, not just the concepts that are used in English. A variety of temporal concepts are used in natural languages, with cultural differences in how they are used in embodied communication. All of the different concept types are potentially relevant to the Lingodroid agents. In Newtonian physics, each point in time can be viewed as a position along a single dimension. Languages differ in how this dimension is represented: In languages like English, time is conceptualized as a horizontal line and there are two ways of viewing time passing: either the viewer is moving with respect to time "we are nearing the end of semester" or events are moving "the holidays are coming" [17]. In Aymara (the language spoken by the Aymara people of the Andes), the arrow of time is reversed and the future is considered to be behind the speaker (a sensible notion from the perspective of knowledge, since the future is obscured, but the past is known and hence visible [15]). Other languages such as Mandarin use a vertical dimension for time [57]. Using tense and sequencing, points on the time line can either be specified relative to or independently from the current time [58, 59].

Although the conceptualization of time as a single linear dimension is common [21], it is not a universal view. Even in English temporal terms are not easily grounded in any precise metric. Cyclic terms are abundant, dawn and dusk, spring and autumn, days and weeks. Some terms have their origins in cycles in the physical world (e.g. diurnal or seasonal), others originated in cycles in human culture (such as the division of a day into 24 uniform segments).

As mentioned in the introduction, in some disciplines, space is treated as a primary metaphor for understanding time; in others, the issue is not yet resolved, and the debate is between symmetric, dependent and asymmetric relationships between time and space [17, 23-25, 60]. Journeys inherently contain both temporal and spatial aspects and it is possible to consider these issues as they apply to embodied tasks. Lingodroids construct their experience maps from their changing sensory input as they explore their environment. Space, time and change are inter-related aspects of navigation tasks and terms for all three types of concept can be grounded directly in experience. In the *how-long* games they were treated symmetrically and relatively similar concepts were formed. However, a simple thought experiment could separate their contributions by introducing variable speed: Shared duration terms could be precisely formulated, but the variable speed would cause the distance and time concepts to diverge.

VII. Conclusions and Future Work

In conclusion, a variety of concepts can be used to ground a lexicon for journey duration using similar language evolution methods. Representations based on time, distance, and amount of change were found to be useful for duration concept formation. A systematic understanding of how robots can ground temporal concepts will require concepts for duration (the battery charger is ten minutes away), tense (when *did* the robots last meet, when *will* they next meet), proper names for historical and personal temporal events (a robot's birth date, the first time two robots met), and names for cyclic temporal events (morning, time when the robot wakes up from the charger each day, notion of a day, "sleeps"). These core concepts could be extended to allow agents to obtain knowledge about and use a measurement system for durations and temporal events. Future directions for the Lingodroids platform will involve bootstrapping increasingly powerful lexicons for spatial and temporal concepts using games that enable shared attention to specific aspects of embodied experience.

Acknowledgment

We thank colleagues in the Thinking Systems project at UQ and QUT and in the Temporal Dynamics of Learning Center at UCSD for discussions over many years about time, space and how they interact. We also thank two anonymous reviewers for their insightful comments.

References

[1] I. Kant, *Critique of pure reason.* Translated and edited by Paul Guyer and Allen W. Wood. New York: Cambridge University Press, 1998.
[2] R. P. Feynman, "Space-time approach to quantum electrodynamics," *Physical Review*, vol. 76, pp. 769-789, 1949.
[3] J. J. Gibson, "Events are perceivable but time is not," in *The Study of Time II: Proceedings of the Second Conference of the International Society for the Study of Time, Lake Yamanaka-Japan*, J. T. Fraser and N. Lawrence, Eds. Berlin: Springer-Verlag, 1975.
[4] E. Engberg-Pedersen, "Space and Time," in *Cognitive Semantics: Meaning and Cognition*, J. Allwood and P. Gardenfors, Eds. Amsterdan: John Benjamins, 1998.
[5] A. U. Frank, "Qualitative spatial reasoning about distances and directions in geographic space," *Journal of Visual Languages & Computing*, vol. 3, pp. 343-371, 1992.

[6] B. Kuipers, "An Intellectual History of the Spatial Semantic Hierarchy," in *Robotics and Cognitive Approaches to Spatial Mapping*, M. Jefferies and W.-K. Yeap, Eds. Berlin: Springer, 2008, pp. 243-264.

[7] M. Milford and G. Wyeth, "Spatial mapping and map exploitation: A bio-inspired engineering perspective," in *Spatial Information Theory* Berlin: Springer, 2007, pp. 203-221.

[8] J. F. Allen, "Maintaining knowledge about temporal intervals," *Commun. ACM,* vol. 26, pp. 832-843, 1983.

[9] E. Clarke and E. Emerson, "Design and synthesis of synchronization skeletons using branching time temporal logic," in *Logics of Programs*, D. Kozen, Ed. Berlin: Springer, 1982, pp. 52-71.

[10] S. A. Miller and L. K. Schubert, "Time revisited," *Computational Intelligence,* vol. 6, pp. 108-118, 1990.

[11] G. E. Fainekos, H. Kress-Gazit, and G. J. Pappas, "Temporal Logic Motion Planning for Mobile Robots," in *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, 2005, pp. 2020-2025.

[12] H. W. Güsgen, "Spatial reasoning based on Allen's temporal logic," International Computer Science Institute, Berkeley Technical report TR-89-049, 1989.

[13] V. Evans, *The structure of time: Language, meaning and temporal cognition*. Amsterdam: John Benjamins, 2004.

[14] S. C. Levinson, "Language and Space," *Annual Review of Anthropology,* vol. 25, pp. 353-382, 1996.

[15] R. E. Núñez and E. Sweetser, "With the Future Behind Them: Convergent Evidence From Aymara Language and Gesture in the Crosslinguistic Comparison of Spatial Construals of Time," *Cognitive Science* vol. 30, pp. 401–450, 2006.

[16] N. Evans, *Dying words : endangered languages and what they have to tell us*. Chichester, UK: Wiley-Blackwell, 2010.

[17] D. Gentner, M. Imai, and L. Boroditsky, "As time goes by: Evidence for two systems in processing space-time metaphors," *Language and cognitive processes,* vol. 27 (5), pp. 537-565, 2002.

[18] V. Walsh, "A theory of magnitude: Common cortical metrics of time, space and quantity.," *Trends in Cognitive Science,* vol. 7, pp. 483–488, 2003.

[19] D. Kemmerer, "The spatial and temporal meanings of English prepositions can be independently impaired," *Neuropsychologia,* vol. 43, pp. 797-806, 2005.

[20] D. Casasanto and L. Boroditsky, "Time in the mind: Using space to think about time," *Cognition,* vol. 106, pp. 579-593, 2008.

[21] A. Galton, "Time flies but space does not: Limits to the spatialisation of time," *Journal of Pragmatics,* vol. In Press, Corrected Proof.

[22] G. Lakoff and M. Johnson, *Metaphors we live by*. Chicago: The University of Chicago Press, 1980.

[23] D. Gentner, "Spatial metaphors in temporal reasoning," in *Spatial Schemas and abstract thought*, M. Gattis, Ed. Cambridge, MA: MIT Press, 2001, pp. 203-222.

[24] H. H. Clark, "Space, Time, Semantics, and the Child," in *Cognitive Development and the Acquisition of Language*, T. E. Moore, Ed. New York: Academic Press, 1973.

[25] L. Boroditsky, "Metaphoric structuring: understanding time through spatial metaphors," *Cognition,* vol. 75, pp. 1-28, 2000.

[26] K. E. Moore, "Space-to-time mappings and temporal concepts," *Cognitive Linguistics,* vol. 17, pp. 199-244, 2006.

[27] J. O'Keefe and L. Nadel, *The hippocampus as a cognitive map*. New York: Oxford University Press Inc, 1978.

[28] W. E. Skaggs, B. L. McNaughton, M. A. Wilson, and C. A. Barnes, "Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences," *Hippocampus,* vol. 6, pp. 149-172, 1996.

[29] J. B. Aimone, J. Wiles, and F. H. Gage, "Computational Influence of Adult Neurogenesis on Memory Encoding," *Neuron,* vol. 61, pp. 187-202, 2009.

[30] L. Vila, "A survey on temporal reasoning in artificial intelligence," *AI Communications,* vol. 7, pp. 4-28, 1994.

[31] J. R. Hobbs and F. Pan, "An ontology of time for the semantic web," *ACM Transactions on Asian Language Information Processing (TALIP),* vol. 3, pp. 66-85, 2004.

[32] D. J. Peuquet, "Making Space for Time: Issues in Space-Time Data Representation," *GeoInformatica,* vol. 5, pp. 11-32, 2001.

[33] A. Galton, "Fields and Objects in Space, Time, and Space-time," *Spatial Cognition & Computation: An Interdisciplinary Journal,* vol. 4, pp. 39 - 68, 2004.

[34] L. Zhou and G. Hripcsak, "Temporal reasoning with medical data--A review with emphasis on medical natural language processing," *Journal of Biomedical Informatics,* vol. 40, pp. 183-202, 2007.

[35] I. Mani, M. Verhagen, B. Wellner, C. M. Lee, and J. Pustejovsky, "Machine learning of temporal relations," in *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics* Sydney, Australia: Association for Computational Linguistics, 2006.

[36] R. J. Passonneau, "A computational model of the semantics of tense and aspect," *Computational Linguistics,* vol. 14, pp. 44-60, 1988.

[37] J. De Beule, "Simulating the syntax and semantics of linguistic constructions about time," in *Evolutionary Epistemology, Language and Culture*, 2006, pp. 407-428.

[38] L. Steels, "The emergence and evolution of linguistic structure: from lexical to grammatical communication systems," *Connection Science,* vol. 17, pp. 213-230, 2005.

[39] M. Levit and D. Roy, "Interpretation of Spatial Language in a Map Navigation Task," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,* vol. 37, pp. 667-679, 2007.

[40] R. Schulz, D. Prasser, P. Stockwell, G. Wyeth, and J. Wiles, "The formation, generative power, and evolution of toponyms: Grounding a spatial vocabulary in a cognitive map," in *The Evolution of Language: Proceedings of the 7th International Conference (EVOLANG7)*, A. D. M. Smith, K. Smith, and R. Ferrer i Cancho, Eds. Singapore: World Scientific Publishing, 2008, pp. 267-274.

[41] R. Moratz and T. Tenbrink, "Spatial Reference in Linguistic Human-Robot Interaction: Iterative, Empirically Supported Development of a Model of Projective Relations," *Spatial Cognition & Computation,* vol. 6, pp. 63-107, 2006.

[42] R. Schulz, A. Glover, G. Wyeth, and J. Wiles, "Robots, communication, and language: An overview of the Lingodroid project," in *Proceedings of the 2010 Australasian Conference on Robotics and Automation*, G. Wyeth and B. Upcroft, Eds. Brisbane, Australia, 2010.

[43] M. Milford, G. Wyeth, and D. Prasser, "RatSLAM: a hippocampal model for simultaneous localization and mapping," in *IEEE International Conference on Robotics and Automation, ICRA 2004, April 26 - May 1, 2004* New Orleans, LA, USA: IEEE Press, 2004.

[44] L. Steels, "Language games for autonomous robots," *IEEE Intelligent systems,* vol. 16, pp. 16-22, 2001.

[45] L. Steels, *The Talking Heads Experiment* vol. I. Words and Meanings. Brussels: Best of Publishing, 1999.

[46] A. D. M. Smith, "Establishing communication systems without explicit meaning transmission," in *ECAL01*, J. Kelemen and P. Sosik, Eds. Prague: Springer, 2001, pp. 381-390.

[47] A. Cangelosi, G. Metta, G. Sagerer, S. Nolfi, C. Nehaniv, K. Fischer, J. Tani, T. Belpaeme, G. Sandini, F. Nori, L. Fadiga, B. Wrede, K. Rohlfing, E. Tuci, K. Dautenhahn, J. Saunders, and A. Zeschel, "Integration of Action and Language Knowledge: A Roadmap for Developmental Robotics," *IEEE Transactions on Autonomous Mental Development,* vol. 2, pp. 167-195, 2010.

[48] P. Bodík and M. Takáč, "Formation of a common spatial lexicon and its change in a community of moving agents," in *Frontiers in Artificial Intelligence and Applications: Eighth Scandinavian Conference on Artificial Intelligence SCAI'03,* B. Tessem, P. Ala-Siuru, P. Doherty, and B. Mayoh, Eds. Amsterdam: IOS Press Inc, 2003, pp. 37-46.

[49] L. Steels and J.-C. Baillie, "Shared grounding of event descriptions by autonomous robots," *Robotics and Autonomous Systems,* vol. 43, pp. 163-173, 2003.

[50] P. Vogt, "Bootstrapping grounded symbols by minimal autonomous robots," *Evolution of Communication,* vol. 4, pp. 87-116, 2000.

[51] L. Steels and M. Loetzsch, "Perspective alignment in spatial language," in *Spatial Language and Dialogue*, K. R. Coventry, T. Tenbrink, and J. A. Bateman, Eds. Oxford, UK: Oxford University Press, 2007.

[52] L. Steels, "The symbol grounding problem has been solved. So what's next?," in *Symbols and Embodiment: Debates on meaning and cognition*, M. De Vega, A. M. Glenberg, and A. C. Graesser, Eds. New York: Oxford University Press, 2008, pp. 223-244.

[53] M. Milford, R. Schulz, D. Prasser, G. Wyeth, and J. Wiles, "Learning spatial concepts from RatSLAM representations," *Robotics and Autonomous Systems - From Sensors to Human Spatial Concepts,* vol. 55, pp. 403-410, 2007/5/31 2007.

[54] R. Schulz, P. Stockwell, M. Wakabayashi, and J. Wiles, "Towards a spatial language for mobile robots," in *The Evolution of Language: Proceedings of the 6th International Conference (EVOLANG6)*, A.

Cangelosi, A. D. M. Smith, and K. Smith, Eds. Singapore: World Scientific Publishing, 2006, pp. 291-298.
[55] M. Milford and G. Wyeth, "Persistent Navigation and Mapping using a Biologically Inspired SLAM system," *The International Journal of Robotics Research,* 2009.
[56] M. Dye, "Why Johnny Can't Name His Colors," *Scienctific American,* vol. July, 2010.
[57] L. Boroditsky, "Does Language Shape Thought?: Mandarin and English Speakers' Conceptions of Time," *Cognitive Psychology,* vol. 43, pp. 1-11, 2001.
[58] E. C. Traugott, "On the expression of spatio-temperal relations in language," in *Universals of human language: Word structure*. vol. 3, J. H. Greenberg, Ed. Stanford, CA: Stanford University Press, 1978, pp. 369-400.
[59] T. Tenbrink, *Space, time, and the use of language: An investigation of relationships*. Berlin: Mouton de Gruyter, 2007.
[60] D. Casasanto, O. Fotakopoulou, and L. Boroditsky, "Space and Time in the Child's Mind: Evidence for a Cross-Dimensional Asymmetry," *Cognitive Science,* vol. 34, pp. 387–405, 2010.

Ruth Schulz received a dual degree in Engineering (Electrical) and Science (Computer Science) in 2003 and a Ph.D. degree in computer science in 2008 from the University of Queensland, Brisbane, Australia. She is currently a Postdoctoral Researcher at the University of Queensland in the Complex and Intelligent Systems group. Her research interests include the evolution of language, spatial language, and computational models of language for mobile robots.

Gordon Wyeth (M'03) received the B.E. and Ph.D. degrees in computer systems engineering from the University of Queensland, Brisbane, Australia, in 1989 and 1997, respectively. He is currently a Professor of Robotics at the School of Engineering Systems, Queensland University of Technology. He is the Co-Director of the CyPhy Lab. His current research interests include biologically inspired robot systems, developmental robotics, and robot education.

Professor Wyeth is a Chief Investigator on the Thinking Systems Project, a joint Australian Research Council / National Health and Medical Research Council (ARC/NHMRC) funded initiative, and has served as the Chief Investigator on several Australian Research Council and industry funded projects. He was the President of the Australian Robotics and Automation Association (2004–2006) and has three times chaired the Australasian Conference on Robotics and Automation (1999, 2003, 2010). He was three times runner-up in the RoboCup small-size league and has served on various RoboCup committees. He developed RoboCup Junior in Australia and has twice chaired the Australian RoboCup Junior championships.

Janet Wiles holds a Ph.D. from the University of Sydney, and is Professor of Complex and Intelligent Systems at the University of Queensland. She is the project leader of the Thinking Systems Project, supervising a cross-disciplinary team studying fundamental issues in how information is

transmitted, received, processed and understood in biological and artificial systems. Her research interests include complex systems biology, computational neuroscience, computational modeling methods, artificial intelligence and artificial life, language and cognition.