Notes

Foreword

Welcome to ICDL 2008!

It is our great pleasure to welcome you to the 7th International Conference on Development and Learning (ICDL 2008) in Monterey, California, USA. ICDL is a unique interdisciplinary conference, with latest research results that cover challenging gaps among computer science, robotics, psychology, and neuroscience to address the challenging subject of computational autonomous development, both natural and artificial.

ICDL 2008 accepted multiple submission formats, including a regular 6-page paper submission track, a short "late-breaking" 1-page abstract submission track, and proposals for special sessions. This was the first year that the "late-breaking" submission track was offered, and we received very positive feedback on providing the opportunity for many to present preliminary or very recent work and to take part in the ICDL meeting.

From the more than 120 different submissions, we selected 24 full papers for oral presentation, 27 full papers for poster presentation, 2 special sessions with a total of 8 presentations, and 21 late-breaking abstracts for poster presentation. The quality of the submissions was universally very high, providing a difficult (and often hotly debated!) choice for our program committee. Our review process enlisted the help of 38 program committee members who coordinated reviews from 127 reviewers.

We were also fortunate to have three outstanding invited speakers. Terry Jernigan will describe structural changes in the brain in response to experience. Richard Aslin will present theoretical and experimental investigations of distributional pattern learning in infants and adults. Andrew Ng will discuss unsupervised learning of structure via neuroscience inspired learning algorithms.

Our thanks to the ICDL General Chairs Jay McClelland and John Weng, Publications Chair Chad Jenkins, Publicity Chair Charlie Kemp, and ICDL Communications Co-Chair Jochen Triesch. We thank our financial sponsor, the IEEE Computational Intelligence Society, and we thank the Cognitive Science Society for their endorsement.

We hope that you will enjoy this exciting program.

Brian Scassellati
Gedeon Deak
ICDL 2008 Program Chairs

ICDL 2008 Organizing Committee

General Chairs:
Jay McClelland, Stanford University
Juyang (John) Weng, Michigan State University

Program Chairs:
Gedeon Deak, University of California, San Diego
Brian Scassellati, Yale University

Publication Chair:
Odest Chadwicke Jenkins, Brown University

Publicity Chair:
Charlie Kemp, Georgia Tech and Emory University

ICDL 2008 Program Committee Members

Minoru Asada	Osaka University
James Bednar	University of Edinburgh
Luc Berthouze	University of Sussex
Andrew Bremner	Goldsmiths, University of London
Kerstin Dautenhahn	University of Hertfordshire
Gedeon Deak	University of California, San Diego
Anne Fernald	Stanford University
Li Hong	Southwest University
Koh Hosoda	Osaka University
Shoji Itakura	Kyoto University
Odest Jenkins	Brown University
Scott Johnson	University of California, Los Angeles
Frederic Kaplan	EPFL
Charlie Kemp	Georgia Tech
Charles Kemp	Carnegie Mellon University
Jeffrey Krichmar	University of California, Irvine
Benjamin Kuipers	University of Texas at Austin
Stephen Levinson	Beckman Institute
Hong Lu	Fudan University
Karl MacDorman	Indiana University
Jay McClelland	Stanford University
Giorgio Metta	University of Genoa
Risto Miikkulainen	University of Texas at Austin
Yukie Nagai	Bielefeld University
Chrystopher Nehaniv	University of Hertfordshire
Sandy Pentland	Massachusetts Institute of Technology
Michael Ramscar	Stanford University
Ayse Pinar Saygin	University College London
Brian Scassellati	Yale University
Matthias Scheutz	Indiana University
Matthew Schlesinger	Southern Illinois University
Gregor Schöner	Ruhr-Universität Bochum
Thomas Shultz	McGill University
Sylvain Sirois	University of Manchester
Nathan Sprague	Kalamazoo College
Jochen Triesch	Frankfurt Institute for Advanced Studies
John Weng	Michigan State University
Chen Yu	Indiana University
Zhengyou Zhang	Microsoft Research

7th IEEE International Conference on Development and Learning – Monterey, California

"Late-Breaking" Submissions

Poster Session #2 – One-page Abstracts

Surprise-Based Learning for Developmental Robotics
Nadeesha Ranasinghe, Wei-Min Shen

Generating Analogies to Stories: Investigating Young Children's Analogical Transfer
Bryan Matlen, David Klahr

A Hybrid Architecture for Function Approximation
Osman Hassab Elgawi

The Influence of Lexical Association on Young Children's Inductive Inference
Bryan Matlen, Anna Fisher

Children's Facial and Motor Emotional Expressions in Emotion Regulation
Twila Tardif, Tal Shafir, Yunyi He, Chit Yuen (Frankie) Yi, Rosa Angulo-Barroso, Alison Miller, Sheryl Olson, Niko Kaciroti

Using Semantic Information for Language Learning
Leonor Becerra-Bonache, Dana Angluin

Studying Facial Expressions Using Manifold Learning and Support Vector Machines
Daniel Messinger, Mohammad Mahoor, Lisa Ibanez, Maria Kimijima, Yang Wang, Steven Cadavid, Jeffrey Cohn

Interaction Based Knowledge Acquisition and Exchange Using Grounded Symbols
Joerg Irran, Brigitte Krenn

Learning From Interactive Videos and Video Games
Cheyenne Cummings, Rebekah Richert

Learning at Home and Learning at the Lab: An Eyetracking Study
Karinna Hurley, Lisa Oakes

Incremental Life-long Learning by Demonstration
Nathan Koenig, Maja Mataric

What to Infer from an Inference
Daniel Sternberg, James McClelland

Ten-month-old infants can localize changes in visual short-term memory: An eye tracking study
Ian Messenger, Lisa Oakes

6-month old infant point-following is related to infant emotions and maternal sensitivity
Jacqueline Acuna, Gedeon Deak, Andrea Chiba, Kaya de Barbaro

ICDL 2008 Reviewers

Tiziano Agostini	University of Trieste
William H Alexander	Indiana University
Aris Alissandrakis	Tokyo Institute of Technology
Mike Anderson	Franklin & Marshall College
Jan Antolik	University of Edinburgh
Tom Armstrong	UMBC
Minoru Asada	Osaka University
Dick Aslin	University of Rochester
Jean-Julien Aucouturier	University of Tokyo
Jennifer Aydelott	Birkbeck University of London
Khaled Bachour	EPFL
Dare Baldwin	University of Oregon
Frank Baughman	Birkbeck College
Michael Beetz	Technische Universität Munich
Tony Belpaeme	University of Plymouth
Nadia Bianchi-Berthouze	University College London
Estela Bicho	Universidade de Minho
Liz Bonawitz	MIT
Arielle Borovsky	University of California, San Diego
Gavin Bremner	Lancaster University
Jennifer Burke	University of Southern Florida
Nicholas Butko	University of California, San Diego
Houston-Price Carmel	University of Reading
Ricardo Chavarriaga	IDIAP Research institute
Emmanuel Chemla	Laboratoire de Sciences Cognitives et Psycholinguistique
Antao Chen	Southwest University
Sonia Chernova	Carnegie Mellon University
Yoonsuck Choe	Texas A&M University
Jeffrey Coldren	Youngstown State University
Christopher Conway	Indiana University
Matthew Cook	ETH Zurich
Stephen Cowley	University of Hertfordshire
Christopher Crick	Yale University
Rick Dale	University of Memphis
Bart de Boer	Universiteit van Amsterdam
Gedeon Deak	University of California, San Diego
Evelina Dineva	University of Iowa
Ian Fasel	University of Texas at Austin
Christian Faubel	Ruhr-Universität Bochum
Jerry Feldman	University of California, Berkeley
Rui Feng	Fudan University
Jason Fleischer	The Neurosciences Institute
Anna Franklin	University of Surrey
Shic Frederick	Yale University
Tamami Fukushi	JST
Lakshmi Gogate	Florida Gulf Coast University
Kevin Gold	Yale University
David Grimes	NeuroVista Corp
Verena Hafner	Humboldt-Universität zu Berlin
Verena Heidrich-Meisner	Ruhr-Universität Bochum
Mattias Heldner	KTH Speech
Stefanie Hoehl	MPI for Human Cognitive and Brain Sciences
Koh Hosoda	Osaka University
Mutsumi Imai	Keio University
Toshio Inui	Kyoto University
Ioannis Iossifidis	Ruhr-Universität Bochum
Shoji Itakura	Kyoto University
Brenda Jansen	University of Amsterdam
Zhengping Ji	Michigan State University
Cheng Jin	Fudan University
Justin Hart	Yale University
Frederic Kaplan	EPFL
Kazuhiko Kawamura	Vanderbilt University
Charlie Kemp	Georgia Tech
Cory Kidd	Intuitive Automata Inc.
Taemie Kim	MIT
Judith Law	University of Edinburgh
Stephen Levinson	Beckman Institute
Martin Loetzsch	Sony CSL Paris
Matthew Luciw	Michigan State University
Gary Lupyan	Cornell University
Ankur Mani	MIT
Danielle Matthews	The University of Manchester
Matthew McClain	21st Century Technologies
Gerald McRoberts	Haskin Lab
Francisco Melo	Carnegie Mellon University
Ali Minai	University of Cincinnati
Naeem Assif Mirza	University of Hertfordshire
Joseph Modayil	University of Rochester
Clayton Morrison	USC/ISI
Jonathan Mugan	University of Texas at Austin
Cota Nabeshima	University of Tokyo
Yukie Nagai	Bielefeld University
Lorenzo Natale	Italian Institute of Technology
Daniel Navarro	University of Adelaide
Chrystopher Nehaniv	University of Hertfordshire
Monica Nicolescu	University of Nevada
Francesco Nori	Italian Institute of Technology
Tuna Oezer	Google
Masaki Ogino	JST ERATO Asada Synergistic Intelligence Project
Daniel Olguin Olguin	MIT
Francesco Orabona	IDIAP Research Istitute
Sarah Parsons	University of Birmingham
Alfredo Pereira	Indiana University
Amy Perfors	MIT
Robert Platt	NASA
Daniel Polani	University of Hertfordshire
Jefferson Provost	University of Pittsburgh
Steven Robertson	Cornell University
Ben Robins	University of Hertfordshire
Katharina Rohlfing	Bielefeld University
Yulia Sandamirskaya	Ruhr-Universität Bochum
Joe Saunders	University of Hertfordshire
Brian Scassellati	Yale University
Matthew Schmill	UMBC
Anil Seth	University of Sussex
Sylvain Sirois	University of Manchester
Mojtaba Solgi	Michigan State University
Michael Spratling	King's College London
Ida Sprinkhuizen-Kuyper	Radboud University Nijmegen
Kevin Squire	Naval Postgraduate School
Daniel Stronger	University of Texas
Dan Swingley	University of Pennsylvania
Botond Szatmary	The Neurosciences Institute
Yasutake Takahashi	Osaka University
Shinya Takamuku	ERATO Asada Project
Andrea Thomaz	GA Tech
Julia Trommershaeuser	Giessen University
Vinod Valsalam	University of Texas
Benjamin Waber	MIT Media Laboratory
Michael Walters	University of Hertfordshire
Yuni Xia	Indiana University
Yuichiro Yoshikawa	Asada Synergistic Intelligence Project
Jiajin Yuan	Southwest University
Shuqing Zeng	GM
Wei Zhang	Fudan University

Conference Program

Schedule of Events

Saturday, August 9th

Time	*Activity*
6:00-7:00	**Dinner**
7:00-7:15	**Welcome** **Jay McClelland & John Weng, General Chairs** **Gedeon Deak & Brian Scassellati, Program Chairs**
7:15-8:15	**Keynote Presentation:** ***Structural imaging reveals neurobiological responses to experience and other environmental factors*** **Terry Jernigan, UCSD**
8:15-10:00	**Welcome Reception (Patio Party)**

Sunday, August 10th

Time	*Activity*
7:30-8:30	**Breakfast**
9:00-10:00	**Paper Session:** **Intention and Causality**
	Parental Action Modification Highlighting the Goal versus the Means Yukie Nagai, Katharina Rohlfing
	Ockham's razor as inductive bias in preschooler's causal explanations Liz Bonawitz, Isabel Chang, Catherine Clark, Tania Lombrozo
	Inferring Narrative and Intention from Playground Games Christopher Crick, Brian Scassellati
10:00-10:30	**Break**
10:30-12:00	**Special Session:** **Bayesian and Connectionist Approaches to Learning**
	Participants: Tom Griffiths (Berkeley), Jay McClelland (Stanford), Alison Gopnik (Berkeley), and Mark Seidenberg (Wisconsin)
12:00-1:00	**Lunch**
1:00-2:00	**Keynote Presentation** ***Unsupervised learning of distributional patterns by adults and infants: Empirical findings and ideal-learner models*** **Richard Aslin, University of Rochester**

explanatory power of each of these theories for the characteristic behaviors of autism, as defined by the DSM-IV-TR diagnostic criteria for Autistic Disorder.

A Reinforcement Learning Model of Social Referencing
Hector Jasso, Jochen Triesch, Gedeon Deak

We present a novel computational model of social referencing. The model replicates a classic social referencing experiment where an infant is presented with a novel object and has the choice of consulting an adult's informative facial expression before reacting to the object. The infant model learns the value of consulting the adult's facial expression using the temporal difference learning algorithm. The model is used to make hypotheses about the reason for a lack of social referencing found in autistic individuals, based on an aversion to faces. Comparisons are made between this reinforcement learning model and a previous model based on mood contagion.

TAMER: Training an Agent Manually via Evaluative Reinforcement
William Knox, Peter Stone

Though computers have surpassed humans at many tasks, especially computationally intensive ones, there are many tasks for which human expertise remains necessary and/or useful. For such tasks, it is desirable for a human to be able to transmit knowledge to a learning agent as quickly and effortlessly as possible, and, ideally, without any knowledge of the details of the agent's learning process. This paper proposes a general framework called Training an Agent Manually via Evaluative Reinforcement (TAMER) that allows a human to train a learning agent to perform a common class of complex tasks simply by giving scalar reward signals in response to the agent's observed actions. Specifically, in sequential decision making tasks, an agent models the human's reward function and chooses actions that it predicts will receive the most reward. Our novel algorithm is fully implemented and tested on the game Tetris. Leveraging the human trainers' feedback, the agent learns to clear an average of over 50 lines by its third game, an order of magnitude faster than the best autonomous learning agents.

A Robot Rehearses Internally and Learns an Affordance Relation
Erdemir Erdem, Carl Frankel, Sean Thornton, Baris Ulutas, Kazuhiko Kawamura

This paper introduces a novel approach to a crucial problem in robotics: Constructing robots that can learn general affordance relations from their experiences. Our approach has two components. (a) The robot models affordances as statistical relations between actual actions, object properties and the experienced effects of actions on objects. (b) To exploit the general-knowledge potential of its actual experiences, the robot, much like people, engages in internal rehearsal, playing-out imagined scenarios grounded in but different from actual experience. To the extent the robot veridically appreciates affordance relations, the robot can autonomously predict the outcomes of its behaviors before executing them. Accurate outcome prediction in turn facilitates planning of a sequence of behaviors, toward executing the robot's given task successfully. In this paper, we report very first steps in this approach to affordance learning, viz., the results of simulations and humanoid-robot-embodied experiments targeted toward having the robot learn one of the simplest of affordance relations, that a space affords traversability vs. impediment to a goal-object in the space.

How Thinking in Pictures Can Explain Many Behavioral Characteristics of Autism
Maithilee Kunda, Ashok Goel

We give a brief overview of Thinking in Pictures as a cognitive account of autism, inspired by the book of the same name by Temple Grandin. Then, we examine Thinking in Pictures alongside other prominent cognitive theories of autism (Mindblindness, Weak Central Coherence, and Executive Dysfunction) as a functional, rather than just descriptive, cognitive theory. We focus on the

2:00-3:00	**Paper Session: Learning Linguistic and Perceptual Categories**
	Acquisition of Semantics through Unsupervised Discovery of Associations between Perceptual Symbols Tuna Oezer
	Modeling Unsupervised Perceptual Category Learning Brenden Lake, Gautam Vallabha, Jay McClelland
	A self-referential childlike model to acquire phones, syllables and words from acoustic speech Holger Brandl, Britta Wrede, Frank Joublin, Christian Goerick
3:00-3:30	**Break**
3:30-5:00	**Poster Teasers** ***(25 x 3min each)***
5:00-5:30	**Break**
5:30-7:00	**BBQ Dinner**
7:30-10:00	**Poster Session #1** ***(Full Papers)***

Monday, August 11th

Time	***Activity***
7:30-8:30	**Breakfast**
9:00-10:00	**Paper Session: Language Learning Principles**
	Input Affects Uptake: How Language Experience Influences Processing Efficiency and Vocabulary Learning Anne Fernald, Virginia Marchman, Nereyda Hurtado
	Acquiring Linguistic Argument Structure from Multimodal Input using Attentive Focus G. Satish, Amitabha Mukerjee
	Caregiver's Sensorimotor Magnets Lead Infant's Vowel Acquisition through Auto Mirroring Hisashi Ishihara, Yuichiro Yoshikawa, Katsushi Miura, Minoru Asada
10:00-10:30	**Break**
10:30-12:00	**Special Session: Visual Attention and Recognition**
	What Roles can Attention Play in Recognition? John Tsotsos
	Where-What Network 1: "Where" and "What" Assist Each Other Through Top-down Connections Zhengping Ji, John Weng, Prokhorov Danil
	The Effects of Neuromodulation on Attention and Action-Selection Jeffrey Krichmar
	Motor Aspect of Understanding, Development, and Recognition in Vision Yoonsuck Choe, Huei-Fang Yang, Navendu Misra
12:00-1:00	**Lunch**

1:00-2:00	**Keynote Presentation** ***Unsupervised discovery of structure via neuroscience inspired learning algorithms*** **Andrew Ng, Stanford University**
2:00-3:30	**Paper Session:** **Autism**
	Autism, Eye-Tracking, Entropy Frederick Shic, Katarzyna Chawarska, Jessie Bradshaw, Brian Scassellati
	Modeling the Development of Overselectivity in Autism Trent Kriete, David Noelle
	Paper Session: **Spatial Learning and Manipulation**
	Adaptive Temporal Difference Learning of Spatial Memory in the Water Maze Task Erik Stone, Marjorie Skubic, James Keller
	Detecting the Functional Similarities Between Tools Using a Hierarchical Representation of Outcomes Jivko Sinapov, Alexander Stoytchev
3:30-4:00	**Break**
4:00-5:30	**Paper Session:** **Architectures and Principles**
	Embodied Solution: The World from a Toddler's Point of View Chen Yu, Linda Smith, Alfredo Pereira
	From Pixels to Policies: A Bootstrapping Agent Jeremy Stober, Benjamin Kuipers
	Internal State Predictability as an Evolutionary Precursor of Self-Awareness and Agency Jaerock Kwon, Yoonsuck Choe
	Motor Initiated Expectation through Top-Down Connections as Abstract Context in a Physical World Matt Luciw, John Weng, Shuqing Zeng
5:30-6:00	**Break**
6:00-7:30	**Dinner**
7:30-10:00	**Poster Session #2** ***(1-page submissions)***

Tuesday, August 12

Time	***Activity***
7:30-8:30	**Breakfast**
9:00-10:00	**Paper Session:** **Dynamical Systems and Neural Fields**
	Homeostatic Development of Dynamic Neural Fields Claudius Gläser, Frank Joublin, Christian Goerick
	A Dynamic Systems Approach to Usage-based Model: From the Results of Robotic Learning Experiments Yuuya Sugita, Jun Tani
	Dynamic Field Theory of Sequential Action: A Model and its Implementation on an Embodied Agent Yulia Sandamirskaya, Gregor Schöner

Body-part categories of early-learned verbs: Different granularities at different points in development
Josita Maouene, Shohei Hidaka, Linda Smith

This paper builds on our previous finding that early verbs are strongly related to body parts. One evidence for this relation is the strong word associations among adults between common verbs and body parts. Although many common verbs are related to body parts, the prior evidence suggests that some verbs are strongly related to highly specific body regions (e.g., fingers) and others to larger or more diffuse regions (e.g., hand and arm). Here we ask whether this granularity or specificity in associations is related to age of acquisition. We examine the structure of adult associations of common verbs to body parts as a function of age of acquisition for a 100 verbs normatively acquired 16 to 30 months. We generated verb clusters based on body parts features, and analysed how these body-parts-based clusters account for variance of age of acquisition (AoA) of verbs. By applying this analysis from the 50 earliest learned verbs to the 50 latest learned ones, we found several clusters relevant to AoA in different granularity of body parts. The results fit with growing behavioural and neuro-imaging results on the role of the body ñ and sensory-motor interactions in the world ñ in verb processing.

What Prosody Tells Infants To Believe
Elizabeth Kim, Kevin Gold, Brian Scassellati

We examined whether evidence for prosodic signals about shared belief can be quantitatively found within the acoustic signal of infant-directed speech. Two transcripts of infant-directed speech for infants aged 1;4 and 1;6 were coded with acoustic predictions based on Pierrehumbert and Hirschberg's theory of pitch accents [26]. Predictions were made within a simple single-tone model that reflected only whether the speaker intended to add a word's meaning to the discourse (high tone, H*) or not (low tone, L*), as well as a more complex five category model that added categories reflecting whether the associated word was one of several possible alternatives (L*+H), a contrasting alternative (L+H*), or something about which the listener should make an inference (H*+L). The acoustic signal was then manually segmented and classified based solely on whether the pitches at the beginning, end, and peak intensity points of a stressed syllable were closer to the utterance's pitch minimum or maximum on a log scale. Evidence supporting the pitch predictions was found for the L*, H*, and L*+H accents, but not for L+H* or H*+L. No evidence was found to support the hypothesis that infant-directed speech simplifies two-tone pitch accents into single-tone pitch accents.

Sound versus meaning: What matters most in early word learning?
Sarah Sahni, Timothy Rogers

Previous work suggests that phonological neighborhood density is a key factor in shaping early lexical acquisition. Such studies have, however, have not considered how semantic neighborhoods may influence word-learning. We studied how phonological and semantic densities affect both comprehension and production of nouns from the Macarthur-Bates Communicative Development Inventory (MCDI). New measures of semantic and phonological densities, along with child-directed word frequency counts were used to predict the percentage of children who know each word at different ages (8 - 30 months) as indicated in MCDI lexical norms. Production was predicted by frequency and phonological density at all time points, replicating previous research. Semantic density predicted production only at 30 months. Comprehension norms were predicted by frequency and semantic density, and never by phonological density. Two- and three-way interactions reveal that semantic density may moderate effects in production, while sound density may moderate effects in comprehension.

Autonomous Segmentation of Human Action for Behaviour Analysis
Jonathan Hunter, Mitch Wilkes, Daniel Levin, Caroline Heaton, Megan Saylor

To correctly understand human actions, it is necessary to segment a continuous series of movements into units that can be associated with meaningful goals and subgoals. Recent research in cognitive science and machine vision has explored the perceptual and conceptual factors that a) determine the segment boundaries that human observers place in a range of actions, and b) allow successful discrimination among different action-types. In this project we investigated the degree to which specific movements effectively predict key sub-events in a broad range of actions in which a human model interacts with objects. In addition, we aimed to create an accessible tool to track human actions for use in a wide range of machine vision and cognitive science applications. Results from our analysis suggest that a set of basic movement cues can successfully predict key sub-events such as hand-to-object contact, across a wide range of specific tasks, and we specify parameters under which this prediction might be maximized.

Development of Joint Attention Related Actions Based on Reproducing Interaction Contingency
Sumioka Hidenobu, Yuichiro Yoshikawa, Minoru Asada

Understanding the developmental process of joint attention related actions, such as gaze following and alternation, is one of essential issues for the emergence of communication. Previous synthetic studies have proposed learning methods for gaze following without any explicit instructions as the first step to understand the development of these actions. However, a robot was given a priori knowledge about which pair of sensory information and action should be associated. This paper addresses the development of social actions without such knowledge with a learning mechanism that iteratively acquires social actions by finding and reproducing the contingency inherent in the interaction with a caregiver. The measurement of contingency based on transfer entropy is used to find appropriate pairs of variables for acquiring social actions from possible candidates. The reproduction of found contingency promotes a change of causal structure in the subsequent actions of a caregiver and a robot. In computer simulations of human-robot interaction, we examine what kinds of actions related to joint attention can be acquired in which order by controlling the behavior of caregiver agents. The result shows that a robot acquires joint attention related actions in an order that resembles an infant's development of joint attention.

Realizing being imitated: vowel mapping with clearer articulation
Miura Katsushi, Yuichiro Yoshikawa, Minoru Asada

It is difficult for a robot to transfer our behavior into its own action program directly because robot's system is different from humankind's physical structure. A study about the vowel sound acquisition of the robot has addressed such a physical correspondence problem. This study shows that a robot can acquire several vowel sounds from the caregiver by repeating an imitation process. However, in that study it was already supposed that the robot has the ability to segment caregiver's utterance and to collect answers with proper imitation. If we don't assume this premise, it becomes difficult for the robot to find out the imitation sound in the caregiver's utterance. Therefore in this study, we suppose the situation that we cannot always expect correct imitation sounds from the caregiver. We then propose a method to find out the imitation sound by caregiver's utterance in the situation that it is necessary for the robot to discover the imitation sound of the caregiver, that is identifying the phoneme from the continuation sounds, which might not include exactly corresponding the imitation sound. Furthermore, we show that the robot can find correspondences with the vowel sounds of the caregiver and utter natural vowel sounds by learning with weakly- and self-supervised learning.

10:00-10:30	**Break**
10:30-12:00	**Paper Session:** **Sensorimotor Control and Representation**
	I-POMDP: An Infomax Model of Eye Movement Nicholas Butko, Javier Movellan
	VIP neuron model: Head-centered cross-modal representation of the peri-personal space around the face Sawa Fuke, Masaki Ogino, Minoru Asada
	Sensorimotor Abstraction Selection for Efficient, Autonomous Robot Skill Acquisition George Konidaris, Andrew Barto
	Visual attention by saliency leads cross-modal body representation Mai Hikita, Sawa Fuke, Masaki Ogino, Takashi Minato, Minoru Asada
12:00-1:00	**Lunch**

Paper Abstracts I

Oral Presentations

Sunday, August 10

Parental Action Modification Highlighting the Goal versus the Means
Yukie Nagai, Katharina Rohlfing

Parents significantly alter their infant-directed action compared to adult-directed one, which is assumed to assist the infants' processing of the action. This paper discusses the differences in parental action modification depending on whether the goal or the means is more crucial. When demonstrating a task to infants, parents try to emphasize the important aspects of the task by suppressing or adding their movement. Our hypothesis is that in a goal-crucial task, the initial and final state of the task should be highlighted by parental action, whereas in a means-crucial task the movement is underlined. Our analysis using a saliency-based attention model partially verified it: When focusing on the goal, parents tended to emphasize the initial and final state of objects used in the task by taking a long pause before/after they started/fulfilled the task. When focusing on the means, parents added movement to highlight the object, which consequently made its state invisible. We discuss the results regarding the uniqueness and commonality of parental action modification. We also describe our contribution to the development of robots capable of imitating human actions.

Ockham's razor as inductive bias in preschooler's causal explanations
Liz Bonawitz, Isabel Chang, Catherine Clark, Tania Lombrozo

A growing literature suggests that generating and evaluating explanations is a key mechanism for learning and development, but little is known about how children evaluate explanations, especially in the absence of probability information or robust prior beliefs. Previous findings demonstrate that adults balance several explanatory virtues in evaluating competing explanation, including simplicity and probability. Specifically, adults treat simplicity as a probabilistic cue that trades-off with frequency information. However, no work has investigated whether children are similarly sensitive to simplicity and probability. We report an experiment investigating how preschoolers evaluate causal explanations, and in particular whether they employ a principle of parsimony like Ockham's razor as an inductive constraint. Results suggest that even preschoolers are sensitive to the simplicity of explanations, and require disproportionate probabilistic evidence before a complex explanation will be favored over a simpler alternative.

Inferring Narrative and Intention from Playground Games
Christopher Crick, Brian Scassellati

We present a system which observes humans participating in various playground games and infers their goals and intentions through detecting and analyzing their spatiotemporal activity in relation to one another, and then builds a coherent narrative out of the succession of these intentional states. We show that these narratives capture a great deal of essential information about the observed social roles, types of activity and game rules by demonstrating the system's ability to correctly recognize and group together different runs of the same game, while differentiating them from other games.

complementary to each other, that is the {*shape mutual exclusivity selection principle*} (μX principle). Coping with an example task of learning multimodal joint attention with a caregiver based on observing both another's gaze direction and utterances, the μX principle is synthetically modeled in both intra- and inter-module levels of output so that the system selects more mutually exclusive output and consequently updates mappings to obtain much more mutually exclusive output. In a computer simulation series, we analyze the effects of the μX principle on the mutual facilitation in learning multi-functions and robustness against errors in segmentation of observation. Finally, through a simulation with a caregiver who behaves more carefully, we argue the correspondence of the synthesized development to infant and discuss future work.

Early Interactive Emotional Development
Daniel Messinger, Mohammad Mahoor, Steven Cadavid, Sy-Miin Chow, Jeffrey Cohn

Early infant emotional development concerns the interactive emergence of emotional states that motivate approach and withdrawal. These are indexed by different patterns of infant facial expressions, vocalization, and gazing that emerge within parent-infant interactions in the first 10 months of life. Specifically, the interface of a limited number of interactive parameters creates complex real-time patterns which change over developmental time. These phenomena are described below using techniques from our laboratory such as statistical simulations, continuous ratings, and computer vision modeling.

A robotic model of the development of gaze following
Hyundo Kim, Hector Jasso, Gedeon Deak, Jochen Triesch

For humanoid robots, the skill of gaze following is a foundational component in social interaction and imitation learning. We present a robotic system capable of learning the gaze following behavior in a real-world environment. First, the system learns to detect salient objects and distinguish a caregiver's head poses in a semi-autonomous manner. Once the robot has learned about the objects and head poses, we present multiple scenes containing different combinations of objects and head poses to the robot head. The system learns to associate the detected head pose with correct spatial location of where potentially "rewarding" objects would be using biologically plausible reinforcement learning.

Detection and Categorization of Facial Image through the Interaction with Caregiver
Masaki Ogino, Ayako Watanabe, Minoru Asada

This paper models the process of Applied Behavior Analysis (ABA) therapy of autistic children for eye contact as the learning of the categorization and the preference through the interaction with a caregiver. The proposed model consists of the learning module and visual attention module. The learning module learns the visual features of higher order local autocorrelation (HLAC) that are important to discriminate the visual image before and after the reward is given. The visual attention module determines the attention point by the bottom-up process based on the saliency map and the top-down process based on the learned visual feature. The experiment with a virtual robot shows that the robot successfully learns visual features corresponding to the face firstly and the eyes afterwards through the interaction with a caregiver. After the learning, the robot can attend to the caregiver's face and eyes as autistic children do in the actual ABA therapy.

Buliding a More Effective Teaching Robot Using Apprenticeship Learning
Paul Ruvolo, Jacob Whitehill, Marjo Virnes, Javier Movellan

What defines good teaching? While attributes such as timing, responsiveness to social cues, and pacing of material clearly play a role, it is difficult to create a comprehensive specification of what it means to be a good teacher. On the other hand, it is relatively easy to obtain examples of expert teaching behavior by observing a real teacher. With this inspiration as our guide, we investigated methods that use data recorded from expert teachers as a means of improving the teaching abilities of RUBI, a social robot immersed in a classroom of 18-24 month old children. While this approach has achieved considerable success in mechanical control, such as automated helicopter flight, until now there has been little work on applying it to the field of social robotics. This paper explores two particular approaches to apprenticeship learning, and analyzes the models of teaching that each approach learns from the data of the human teacher. Empirical results indicate that the apprenticeship learning paradigm, though still nascent in its use in the social robotics field, holds promise, and that our proposed methods can already extract meaningful teaching models from demonstrations of a human expert.

What does Shaping Mean for Computational Reinforcement Learning?
Tom Erez, William Smart

This paper proposes a new formulation of shaping as a homotopy-continuation method applied to reinforcement learning. By considering reinforcement learning (RL) tasks as elements in an abstracted task space, we formulate shaping as a homotopy trajectory in that task space, leading from simple tasks to harder ones. We present an exhaustive list of the different ways RL tasks may be modified, and present evidence from RL literature (most of which was originally presented outside the context of shaping) on the usefulness of these modifications in specific domains. We contrast our proposed view with previous work on computational shaping, and argue against the often-held view that equates shaping with a rich reward scheme. We conclude by discussing a proposed research agenda for the computational study of shaping.

Implicit Learning of Arithmetic Principles
Richard Prather, Martha Alibali

This paper addresses the acquisition of arithmetic principles through implicit learning. A large amount of research has investigated children's arithmetic principle knowledge at various points in development. Although we have some idea what typical learners at various ages know, little is known about the learning mechanisms involved. This research takes the position that learning comes through experience with certain types of arithmetic equations, specifically equations that violate the principle to be learned. We hypothesize that this experience leads to modifications in the learner's equation encoding, which in turn facilitates acquisition of principle knowledge. Experimental results suggest that certain types of experience do lead to increased principle knowledge and that equation encoding is a key part of acquiring principle knowledge.

Multimodal joint attention through cross facilitative learning based on *μX* principle
Yuichiro Yoshikawa, Tsukasa Nakano, Minoru Asada, Hiroshi Ishiguro

Learning multi-functions is one of the fundamental issues not only for intelligent robots but also for cognitive development. Namely, in multimodal joint attention, how does learning one function affect another one: is there a mutual effect of learning acceleration, is it neutral, or is there deceleration? In this paper, we enhance the well-known mutual exclusivity bias in child language development as a general bias to simultaneously learn mappings for different functions that are expected to be

Furthermore, the system can use the narratives it constructs to learn and theorize about novel observations, allowing it to guess at the rules governing the games it watches. For example, after watching several different games, the system figures out on its own that Tag-like games require close physical proximity in order for the role of ``it'' to swap from one person to another. Thus a rich and layered trove of social, intentional and cultural information can be drawn out of extremely impoverished and low-context trajectory data.

Acquisition of Semantics through Unsupervised Discovery of Associations between Perceptual Symbols
Tuna Oezer

This paper introduces an unsupervised method to acquire the lexical semantics of action verbs. The eventual goal of the presented method is allowing a robot to acquire language under realistic conditions. The method acquires lexical semantics by forming association sets that contain general perceptual symbols associated with a certain concept as well as perceptual symbols of the utterances of the name of a concept. The lexical semantics is learned with the help of a narrator who comments on what the robot sees. The technique works even if the narrator only occasionally comments on what the robot sees. The paper presents experimental results that show that the method can acquire the lexical semantics of action verbs while the robot is watching a human who performs actions and hearing a narration that only occasionally actually describes what the robot is currently seeing. A comparison with supervised learning algorithms shows that the method discussed in this paper outperforms other techniques.

Modeling Unsupervised Perceptual Category Learning
Brenden Lake, Gautam Vallabha, Jay McClelland

During the learning of speech sounds and other perceptual categories, category labels are not provided, the number of categories is unknown, and the stimuli are encountered sequentially. These constraints provide a challenge for models, but they have been recently addressed in the Online Mixture Estimation model of unsupervised vowel category learning [Vallabha et al., PNAS, 2007, 104:13273-13278]. The model treats categories as Gaussian distributions, proposing both the number and parameters of the categories. While the model has been shown to successfully learn vowel categories, it has not been evaluated as a model of the learning process. We account for three results regarding the learning process: infants' discrimination of speech sounds is better after exposure to a bimodal rather than unimodal distribution [Maye et al., Cognition, 2002, 82:B101-B111], infants' discrimination of vowels is affected by acoustic distance [Sabourin et al., Developmental Science, under revision], and subjects place category centers near frequent stimuli in an unsupervised visual classification task [Rosenthal et al., PNAS, 2001, 98:4265-4270].

A self-referential childlike model to acquire phones, syllables and words from acoustic speech
Holger Brandl, Britta Wrede, Frank Joublin, Christian Goerick

Speech understanding requires the ability to parse spoken utterances into words. But this ability is not innate and needs to be developed by infants within the first years of their life. So far almost all computational speech processing systems neglected this bootstrapping process. Here we propose a model for early infant word learning embedded into a layered architecture comprising phone, phonotactics and syllable learning. Our model uses raw acoustic speech as input and aims to learn the structure of speech unsupervised on different levels of granularity. We present first experiments which evaluate our model on speech corpora that have some of the properties of infant-directed speech. To further motivate our approach we outline how the proposed model integrates into an embodied multi-modal learning and interaction framework running on Honda's Asimo robot.

Monday, August 11

Input Affects Uptake: How Language Experience Influences Processing Efficiency and Vocabulary Learning
Anne Fernald, Virginia Marchman, Nereyda Hurtado

Two studies explore how early vocabulary learning is influenced both by maternal speech to the child and by the child's developing skill in real-time comprehension. Study 1 shows that amount and quality of mothers' speech predict language growth in Spanish-learning children, providing the first evidence that language input shapes speech processing efficiency as well as lexical development. Study 2 demonstrates that early efficiency in speech processing is beneficial for vocabulary growth, showing how fluency in online comprehension facilitates learning.

Acquiring Linguistic Argument Structure from Multimodal Input using Attentive Focus
G Satish, Amitabha Mukerjee

This work is premised on three assumptions: that the semantics of actions can be learned independently of language, that objects in attentive focus indicate the arguments participating in that action, and that knowing the arguments helps align linguistic attention on the relevant predicate (verb). Using a computational model of dynamic attention, we present an algorithm that clusters ongoing events into action classes in an unsupervised manner using the Merge Neural Gas algorithm. With few clusters, the model correlates to coarse concepts such as 'come-closer', but with a finer granularity, it reveals hierarchical substructure such as come-closer-one-object-static and come-closer-both-moving. That the argument ordering is non-commutative is discovered for actions such as chase or come-closer-one-object-static. Knowing the arguments, and given the noun-referent mappings that are easily learned, we can constrain the language learning by considering only linguistic expressions and actions that refer to the objects in perceptual focus. We learn action schemas for linguistic units like moving towards or chase and validate our results by producing output commentaries for 3D video.

Caregiver's Sensorimotor Magnets Lead Infant's Vowel Acquisition through Auto Mirroring
Ishihara Hisashi, Yuichiro Yoshikawa, Katsushi Miura, Minoru Asada

Mother-infant vocal communication is a sort of mystery of human cognitive development since they can communicate although their body structures and therefore their utterable areas are different. This paper proposes a method that aids unconscious guidance in mutual imitation for infant development based on a biasing element with two different kinds of modules. The first is based on the normal magnet effect in perceiving heard vocal sounds as the listener's own vowels (perceptual magnet) and also includes another magnet effect for imitating vocal sounds that resemble the imitator's vowels (articulatory magnet). The second is based on what we call "auto mirroring bias," by which the heard vowel is much closer to the expected vowel because the other's utterance is an imitation of the listener's own utterance. Computer simulation results of mother-infant interaction show the validity of the proposed bias. Finally future issues are discussed.

What Roles can Attention Play in Recognition?
John Tsotsos

Does attention have relevance for visual recognition? If so, under what circumstances? Is there a particular role (or roles) for attentive processes? These are not so simple to answer. Attention, if used at all in computer vision, has traditionally played one or both of the following roles: where to look

Analyzing Composability in a Sparse Encoding Model of Memorization and Association
Jacob Beal, Thomas Knight

A key question in neuroscience is how memorization and association are supported by the mammalian cortex. One possible model, proposed by Valiant, uses sparse encodings in a sparse random graph, but the composability of operations in this model (e.g. an association triggering another association) has not previously been evaluated. We evaluate composability by measuring the size of ``items'' produced by memorization and the propagation of signals through the ``circuits'' created by memorization and association. While the association operation is sound, the memorization operation produces ``items'' with unstable size and produces circuits that are extremely sensitive to noise. We therefore amend the model, introducing an association stage into memorization. The amended model preserves and strengthens the sparse encoding hypothesis and invites further characterization of properties such as capacity and interference.

Hierarchical Voting Experts: An Unsupervised Algorithm for Hierarchical Sequence Segmentation
Matthew Miller, Alexander Stoytchev

This paper extends the Voting Experts (VE) algorithm for unsupervised segmentation of sequences to create the Hierarchical Voting Experts (HVE) algorithm for unsupervised segmentation of hierarchically structured sequences. The paper evaluates the strengths and weaknesses of the HVE algorithm to identify its proper domain of application. The paper also shows how higher order models of the sequence data can be used to improve lower level segmentation accuracy.

Dynamic systems modeling of cortisol response in pre-schoolers using emotion and activity inputs
Twila Tardif, Dawn Tilbury, Matthew Plunkett, Barbara Felt, Niko Kaciroti, Rosa Angulo-Barroso, Tal Shafir, Li Wang

This paper presents preliminary results for using dynamic systems models to describe cortisol responses to stressful events. Linear discrete-time models are used to approximate the nonlinear model around an operating condition. Two different inputs are considered: a stressful task, and activity. Choices made for input modeling are discussed in some detail. Results are presented that indicate an impulse model for the stressful input gives a better fit than no input, and that including simple motor activity levels provide model fits that are as good as emotional impulse models. Combining both impulse and motor activity further improves the parsimony of model fits. In addition, the paper discusses how the resulting dynamic systems models can be used for statistical analysis, as well as for integrating across multiple levels of stress responses.

Automatic Cry Detection in Early Childhood Education Settings
Paul Ruvolo, Javier Movellan

We present results on applying a novel machine learning approach for learning auditory moods in natural environments to the problem of detecting crying episodes in preschool classrooms. The resulting system achieved levels of performance approaching that of human coders and also significantly outperformed previous approaches to this problem.

Paper Abstracts II

Poster Session #1 – Full-length Papers

Trait or Situaiton? Cultural Differences in Judgments of Emotion
Megumi Kuwabara, Ji Son, Linda Smith

Traditional research in cognition assumes that fundamental processes such as memory and attention are universal. However, a growing number of studies suggest cultural differences in the attention and evaluation of information (Masuda & Nisbet 2001; Maass, et al 2006; Markus & Kitayama 1991; Heddenn, et al 2008). One cultural comparison, between Westerners, such as Americans and Easterners such as Japanese suggest that whereas Westerners typically focus on a central single object in a scene Easterners often integrate their judgment of the focal object with surrounding contextual cues. The research reported here considers this cultural difference in the context of children's developing understanding of emotions. The results demonstrate cultural differences in children as young as 3 and 4 years of age. In particular, Japanese children judge emotions based more on contextual information than facial expressions whereas the opposite is true for American children. The addition of language (labeling the emotions) increases the cultural differences.

Neuromodulation and Time-Dependent Plasticity in a Model of Foraging Behavior
Jeffrey Krichmar

In foraging behavior, where an animal searches for food caches, it is imperative for the animal to remember the locations and routes to these caches. An important consideration is the means by which the organism takes the appropriate actions to lead it to a goal that satisfies a particular need. We introduce a time-dependent plasticity rule that biases movement in a particular direction by developing asymmetric neuronal receptive fields through experience. The model contains hippocampal areas that respond differentially to locations in space, frontal cortex areas that respond to different salient cues from the environment, and neuromodulators that respond to rewards and costs. This model suggests a means by which neuromodulated time-dependent plasticity in the frontal cortex can facilitate action selection. It also suggests how these neuronal responses may lead to successful performance in a foraging task.

Modeling the development of causality and occlusion perception in infants
Arthur Franz, Jochen Triesch

Developmental researchers investigate many pieces of infants' physical knowledge, e.g. the perception of causality, occlusion or object permanence, but a theoretical framework that would unify all these pieces, account for the most basic phenomena and make testable predictions has not been provided yet. Here we make an attempt to unify and explain the emergence of causality and occlusion perception and its development in infancy using a simple artificial neural network that derived its representations from simplified motion detector and disparity cells as they can be found in the primary visual cortex. The network accounts simultaneously for two experiments on causality and occlusion perception and develops a representation of object permanence during training. It also makes detailed testable predictions for the course of development and provides us with an account of how change occurs. We conclude that many aspects of physical knowledge can probably be learned from the statistical regularities of our environment while only few assumptions are needed.

next (or selection of region of interest), or top-down task influence on visual computation. In this paper, I argue that these are only two of the possible roles. Attention is also closely linked to binding and it is the triad of attention, binding and recognition that go hand in hand for non-trivial visual recognition tasks. This paper describes a set of four novel binding processes that employ a variety of attentive mechanisms to achieve recognition beyond the first feed-forward pass. The description is at a conceptual level with many pointers to papers where details may be found

Where-What Network 1: "Where" and "What" Assist Each Other Through Top-down Connections
Zhengping Ji, John Weng, Prokhorov Danil

This paper gives a design of a single network that integrates both object location ('where') and object type ('what'), from images of a learned object in natural complex backgrounds. In-place learning algorithm is used to develop the synaptic weights of each neuron in the network, such that every neuron is responsible for the learning of its own signal processing characteristics within its connected network environment. There is no "global", or multi-cell, goal to the learning. Local analysis is achieved through multi-scale receptive fields, with increasing sizes of perception from earlier to later layers. In the experiment, it shows how one type of information assists the network to suppress irrelevant information of background, so as to give the required missing information ('where' or 'what').

The Effects of Neuromodulation on Attention and Action-Selection
Jeffrey Krichmar

Exploit Environmental Events and Explore New Behaviors: The Effects of Neuromodulation on Attention and Action-Selection Biological organisms have the ability to respond quickly to an ever-changing world. Because this adaptability is so critical for survival, all vertebrates have sub-cortical structures, which comprise the neuromodulatory systems, to regulate fundamental behavior by setting the organism's internal and behavioral states. In the vertebrate, the neuromodulatory systems include the cholinergic, dopaminergic, noradrenergic and serotonergic systems. Despite the different origination and chemical signatures of these neuromodulatory systems, there are several commonalities among them: (i) the origination is sub-cortical and consists of small pools of neurons (on the order of thousands in the rodent and tens of thousands in the human), (ii) each of these neuromodulatory systems is the locus of a particular chemical transmitter that is projected to broad areas of the brainstem, thalamus, and cortex, (iii) these neuromodulatory systems are reciprocally connected with the frontal cortex and parts of the limbic system, and (iv) the effect of these neuromodulatory systems on downstream targets is similar. Phasic neuromodulation (i.e. rapid increases in neuromodulatory activity) increases the signal to noise ratio of downstream neuronal targets such that neurons respond to stimuli and suppress their responses to distractions. That is, increases in neuromodulatory activity drive an organism to be more attentive when environmental conditions call for such actions. However, tonic low-level neuromodulatory activity results in downstream neurons responding more to distracters and noise. This response may cause an organism to be more exploratory when there are no pressing events. The main difference between these neuromodulators is that each system is triggered by different environmental stimuli. For example, the serotonergic system appears to be driven by stress or threats, the cholinergic system appears to be driven by attentional effort, the dopaminergic system appears to be driven by reward anticipation, and the noradrenergic system appears to be driven by novelty and saliency. The neuromodulatory system could have a profound and global effect on attention, learning, and action selection in the organism. In this talk, I will show, through simulations and neurorobotic experiments, how principles of the neuromodulatory system could provide a framework for

controlling behavior such that agents attend to the appropriate contextual cues and are decisive when necessary, but indecisive when it is advantageous.

Motor System's Role in Grounding, Receptive Field Development, and Shape Recognition
Yoonsuck Choe, Huei-Fang Yang, Navendu Misra

Vision is basically a sensory modality, so it is no surprise that the investigation into the brain's visual functions has been focused on its sensory aspect. Thus, questions like (1) how can external geometric properties represented in internal states of the visual system be grounded, (2) how do the visual cortical receptive fields (RFs) form, and (3) how can visual shapes be recognized have all been addressed within the framework of sensory information processing. However, this view is being challenged on multiple fronts, with an increasing emphasis on the motor aspect of visual function. In this paper, we will review works that implicate the important role of motor function in vision, and discuss our latest results touching upon the issues of grounding, RF development, and shape recognition. Our main findings are that (1) motor primitives play a fundamental role in grounding, (2) RF learning can be biased and enhanced by the motor system, and (3) shape recognition is easier with motor-based representations than with sensor-based representations. The insights we gained here will help us better understand visual cortical function. Also, we expect the motor-oriented view of visual cortical function to be generalizable to other sensory cortices such as somatosensory and auditory cortices.

Autism, Eye-Tracking, Entropy
Shic Frederick, Katarzyna Chawarska, Jessie Bradshaw, Brian Scassellati

Using eye-tracking, we examine the scanning patterns of 2-year old and 4-year old toddlers with and without autism spectrum disorder as they view static images of faces. We use several measures, such as the entropy of scanning patterns, in order to characterize the differences in attention towards faces by these children. We find a differential pattern of both gross attention (towards the experimental paradigm) and fine attention (towards specific regions of the face) which seem to suggest different developmental trajectories for the two groups of children. We discuss the implications of these trends and, additionally, discuss current methods in eye-tracking and the development of simple, effective, and robust measures and methodology for evaluating scanning patterns.

Modeling the Development of Overselectivity in Autism
Trent Kriete, David Noelle

People with autism consistently demonstrate a lack of sensitivity to the full range of important aspects of everyday situations. Often, an overly restricted subset of the information available in a given situation gains control over their behavior. This can result in problems generalizing learned behaviors to novel situations. This phenomenon has been called overselectivity. Indeed, many behavioral intervention techniques seek to mitigate overselectivity effects in this population. In this paper, we offer an account of overselectivity as arising from an inability to flexibly adjust the attentional influences of the prefrontal cortex on behavior. We posit that dysfunctional dopamine interactions with the prefrontal cortex result in overly perseverative attention in people with autism. Limiting attention to only a few of the features of a situation hinders the learning of associations between the full range of relevant environmental properties and appropriate behavior. Thus, a restricted subset of features gain control over responding. A simple neurocomputational model of the attentional effects of prefrontal cortex on learning is presented, demonstrating how weak dopamine modulation of frontal areas can lead to overselectivity.

Visual attention by saliency leads cross-modal body representation
Sawa Fuke, Mai Hikita, Masaki Ogino, Takashi Minato, Minoru Asada

Body representation is one of the most fundamental issues for physical agents (humans, primates, and robots) in performing various kinds of tasks. Especially during tool-use by monkeys, neurophysiological evidence shows that the representation can be dynamically reconstructed by spatio-temporal integration of different sensor modalities so that it can be adaptive to environmental changes. However, to construct such a representation, an issue to be resolved is how to associate which information among various sensory data. This paper presents a method that constructs cross-modal body representation from vision, touch, and proprioception. Tactile sensation, when the robot touches something, triggers the construction process of the visual receptive field for body parts that can be found by visual attention based on a saliency map and consequently regarded as the end effector. Simultaneously, proprioceptive information is associated with this visual receptive field to achieve the crossmodal body representation. The proposed model is applied to a real robot and results comparable to the activities of parietal neurons observed in monkeys are shown.

I-POMDP: An Infomax Model of Eye Movement
Nicholas Butko, Javier Movellan

Modeling eye-movements during search is important for building intelligent robotic vision systems, and for understanding how humans select relevant information and structure behavior in real time. Previous models of visual search (VS) rely on the idea of ``saliency maps which indicate likely locations for targets of interest. In these models the eyes move to locations with maximum saliency. This approach has several drawbacks: (1) It assumes that occulomotor control is a greedy process, i.e., every eye movement is planned as if no further eye movements would be possible after it. (2) It does not account for temporal dynamics and how information is integrated as over time. To address these limitations, we reformulate the problem of VS as an Information-gathering Partially Observable Markov Decision Process (I-POMDP). We find that contrary to looking to the most likely target locations, as assumed by most current models, an optimal controller should avoid looking directly at the most likely targets. We also find that the optimal control law depends heavily on the Foveal-Peripheral Operating Characteristic (FPOC) of the visual system. This argues against previous ``one-size-fits-all approaches that have tried to make robotic cameras that move their eyes like humans. We show how optimal visual search behavior for standard robotic cameras differs from optimal behavior of a human eye.

VIP neuron model:Head-centered cross-modal representation of the peri-personal space around the face
Sawa Fuke, Masaki Ogino, Minoru Asada

Body representation is one of the most fundamental issues for physical agents (humans, primates, and also robots) to perform various kinds of tasks adaptively, and some learning methods to make robots acquire their body representation have attempted. However, what kind of reference frame (such as head-centered reference frame or camera-centered reference frame) should be constructed has not been dealt. This paper presents a model that enables a robot to acquire the cross modal representation of face parts based on VIP neurons of which function, found in neuroscience, is to connect visual and tactile sensation. The preliminary simulation results are shown and future issues are discussed.

Sensorimotor Abstraction Selection for Efficient, Autonomous Robot Skill Acquisition
George Konidaris, Andrew Barto

To achieve truly autonomous robot skill acquisition, a robot can use neither a single large general state space (because learning is not feasible), nor a small problem-specific state space (because it is not general). We propose that instead a robot should have a set of sensorimotor abstractions that can be considered small candidate state spaces, and select one that is appropriate for learning a skill when it decides to do so. We introduce an incremental algorithm that selects a state space in which to learn a skill from among a set of potential spaces given a successful sample trajectory. The algorithm returns a policy fitting that trajectory in the new state space so that learning does not have to begin from scratch. We demonstrate that the algorithm selects an appropriate space for a sequence of demonstration skills on a physically realistic simulated mobile robot, and that the resulting initial policies closely match the sample trajectory.

Adaptive Temporal Difference Learning of Spatial Memory in the Water Maze Task
Erik Stone, Marjorie Skubic, James Keller

The Morris water maze task is a spatial memory task in which an association between cues from the environment and position must be learned in order to locate a hidden platform. This paper details the results of using a temporal difference (TD) learning approach to learn associations between perceptual states, which are discretized using a Self Organizing Map (SOM), and actions necessary for a robot to successfully locate the hidden platform in a “dry” version of the water maze task. Additionally, the adaptability of the temporal difference learning approach in non-stationary environments is explored.

Detecting the Functional Similarities Between Tools Using a Hierarchical Representation of Outcomes
Jivko Sinapov, Alexander Stoytchev

The ability to reason about multiple tools and their functional similarities is a prerequisite for intelligent tool use. This paper presents a model which allows a robot to detect the similarity between tools based on the environmental outcomes observed with each tool. To do this, the robot incrementally learns an adaptive hierarchical representation (i.e., a taxonomy) for the types of environmental changes that it can induce and detect with each tool. Using the learned taxonomies, the robot can infer the similarity between different tools based on the types of outcomes they produce. The results show that the robot is able to learn accurate outcome models for six different tools. In addition, the robot was able to detect the similarity between tools using the learned outcome models.

Embodied Solution: The World from a Toddler’s Point of View
Chen Yu, Linda Smith, Alfredo Pereira

An important goal in studying both human intelligence and artificial intelligence is an understanding of how a natural or artificial learning system deals with the uncertainty and ambiguity in the real world. We suggest that the relevant aspects in a learning environment for the learner are only those that make contact with the learner’s sensory system. Moreover, in a real-world interaction, what the learner perceives in his sensory system critically depends on both his own and his social partner’s actions, and his interactions with the world. In this way, the perception-action loops both within a learner and between the learner and his social partners may provide an embodied solution that significantly simplifies the social and physical learning environment, and filters irrelevant information for a current learning task which ultimately leads to successful learning. In light of this, we report new findings using a novel method that seeks to describe the visual learning environment from a young child’s point of view. The method consists of a multi-camera sensing environment consisting of two head-mounted mini cameras that are placed on both the child’s and the parent’s foreheads respectively. The main results are that (1) the adult’s and child’s views are fundamentally different when they interact in the same environment; (2) what the child perceives most often depends on his own actions and his social partner’s actions; (3) The actions generated by both social partners provide more constrained and clean input to facilitate learning. These findings have broad implications for how one studies and thinks about human and artificial learning systems.

From Pixels to Policies: A Bootstrapping Agent
Jeremy Stober, Benjamin Kuipers

An embodied agent senses the world at the pixel level through a large number of sense elements. In order to function intelligently, an agent needs high-level concepts, grounded in the pixel level. For human designers to program these concepts and their grounding explicitly is almost certainly intractable, so the agent must learn these foundational concepts autonomously. We describe an approach by which an autonomous learning agent can bootstrap its way from pixel-level interaction with the world, to individuating and tracking objects in the environment, to learning an effective policy for its behavior. We use methods drawn from computational scientific discovery to identify derived variables that support simplified models of the dynamics of the environment. These derived variables are abstracted to discrete qualitative variables, which serve as features for temporal difference learning. Our method bridges the gap between the continuous tracking of objects and the discrete state representation necessary for efficient and effective learning. We demonstrate and evaluate this approach with an agent experiencing a simple simulated world, through a sensory interface consisting of 60,000 time-varying binary variables in a 200 x 300 array, plus a three-valued motor signal and a real-valued reward signal.

Internal State Predictability as an Evolutionary Precursor of Self-Awareness and Agency
Jaerock Kwon, Yoonsuck Choe

What is the evolutionary value of self-awareness and agency in intelligent agents? One way to make this problem tractable is to think about the necessary conditions that lay the foundation for the emergence of agency, and assess their evolutionary origin. We postulate that one such requirement is the predictability of the internal state trajectory. A distinct property of one's own actions compared to someone else's is that one's own is highly predictable, and this gives the sense of ``authorship''. In order to investigate if internal state predictability has any evolutionary value, we evolved sensorimotor control agents driven by a recurrent neural network in a 2D pole-balancing task. The hidden layer activity of the network was viewed as the internal state of an agent, and the predictability of its trajectory was measured. We took agents exhibiting equal levels of performance during evolutionary trials, and grouped them into those with high or low internal state predictability (ISP). The high-ISP group showed better performance than the low-ISP group in novel tasks with substantially harder initial conditions. These results indicate that regularity or predictability of neural activity in internal dynamics of agents can have a positive impact on fitness, and, in turn, can help us better understand the evolutionary role of self-awareness and agency.

Motor Initiated Expectation through Top-Down Connections as Abstract Context in a Physical World
Matt Luciw, John Weng, Shuqing Zeng

Recently, it has been shown that top-down connections improve recognition in supervised learning. In the work presented here, we show how top-down connections represent temporal context as expectation and how such expectation assists perception in a continuously changing physical world, with which an agent interacts during its developmental learning. In experiments in object recognition and vehicle recognition using two types of networks (which derive either global or local features), it is shown how expectation greatly improves performance, to nearly 100% after the transition periods. We also analyze why expectation will improve performance in such real world contexts.

Tuesday, August 12

Homeostatic Development of Dynamic Neural Fields
Claudius Gläser, Frank Joublin, Christian Goerick

Dynamic neural field theory has become a popular technique for modeling the spatio-temporal evolution of activity within the cortex. When using neural fields the right balance between excitation and inhibition within the field is crucial for a stable operation. Finding this balance is a severe problem, particularly in face of experience-driven changes of synaptic strengths. Homeostatic plasticity where the objective function for each unit is to reach some target firing rate seems to counteract this problem. Here we present a recurrent neural network model composed of excitatory and inhibitory units which can self-organize via a learning regime incorporating Hebbian plasticity, homeostatic synaptic scaling, and self-regulatory changes in the intrinsic excitability of neurons. Furthermore, we do not define a neural field topology by a fixed lateral connectivity, rather lateral connections are learned as well.

A Dynamic Systems Approach to Usage-based Model: From the Results of Robotic Learning Experiments
Yuuya Sugita, Jun Tani

We propose a sub-symbolic connectionist model in which a functionally compositional system self-organizes by learning a provided set of goal-directed actions. This approach is compatible with an idea taken from usage-based accounts of the developmental learning of language, especially one theory of infants' acquisition process of symbols. The presented model explains a possible continuous process underlying the transitions from rote knowledge to systematized knowledge by drawing an analogy to the formation process of a regular geometric arrangement of points. An experiment was performed using a simulated mobile robot reaching or turning toward a colored target. By using an identical learning model, three different types of combinatorial generalization are observed depending on the sparseness of the provided examples. Based on the experimental results, the dynamical interpretation of conventional usage-based models is discussed.

Dynamic Field Theory of Sequential Action: A Model and its Implementation on an Embodied Agent
Yulia Sandamirskaya, Gregor Schöner

How sequences of actions are learned, remembered, and generated is a core problem of cognition. Sequence generation is also a challenge for Dynamical Systems approaches to cognition, because the attractor concept seems to be in conflict with the need to destabilize a state in order to switch to the next one. Here we examine how Dynamic Field Theory (DFT), a neuronally grounded dynamical systems approach to embodied cognition, may address sequence learning, sequence memory, and sequential action. To demonstrate that the approach solves the key problem of stabilizing sequences against variable timing of individual actions and noisy sensory feedback about the state of the action system, we implement the approach on a simple autonomous robot. We demonstrate how the robot acquires sequences from experiencing the associated sensory information in serial order and how the robot generates sequences based on visual information from its environment using low-level visual features.