Artificial Intelligence in Medicine (2008) **44**, 155—170

ARTIFICIAL INTELLIGENCE IN MEDICINE

http://www.intl.elsevierhealth.com/journals/aiim

Drinking from the firehose of experience

Benjamin Kuipers *

Computer Science Department, University of Texas at Austin, Austin, TX 78712, USA

Received 4 February 2008; received in revised form 23 July 2008; accepted 25 July 2008

KEYWORDS
Consciousness;
Sensory trackers;
Information content;
Dynamical systems

Summary

Objective: Computational concepts from robotics and computer vision hold great promise to account for major aspects of the phenomenon of consciousness, including philosophically problematical aspects such as the vividness of qualia, the first-person character of conscious experience, and the property of intentionality.
Methods: We present a dynamical systems model describing human or robotic agents and their interaction with the environment. In order to cope with the enormous information content of the sensory stream, this model includes trackers for selected coherent spatio—temporal portions of the sensory input stream, and a self-constructed plausible coherent narrative describing the recent history of the agent's sensorimotor interaction with the world.
Results: We describe how an agent can autonomously learn its own intentionality by constructing computational models of hypothetical entities in the external world. These models explain regularities in the sensorimotor interaction, and serve as referents for the agent's symbolic knowledge representation. The high information content of the sensory stream allows the agent to continually evaluate these hypothesized models, refuting those that make poor predictions. The high information content of the sensory input stream also accounts for the vividness and uniqueness of subjective experience. We then evaluate our account against 11 features of consciousness "that any philosophical—scientific theory should hope to explain", according to the philosopher and prominent AI critic John Searle.
Conclusion: The essential features of consciousness can, in principle, be implemented on a robot with sufficient computational power and a sufficiently rich sensorimotor system, embodied and embedded in its environment.

1. Introduction

Consciousness is one of the most intriguing and mysterious aspects of the phenomenon of mind. Artificial Intelligence (AI) is a scientific field built around the creation of computational models of mind (using not

* Tel.: +1 512 471 9561; fax: +1 512 471 8885.
E-mail address: kuipers@cs.utexas.edu.

0933-3657/$ — see front matter
doi:10.1016/j.artmed.2008.07.010

only logic-based methods for knowledge representation and inference, but also such methods as probabilistic inference, dynamical systems, neural networks, and genetic algorithms). Computational approaches to understanding the phenomena of mind have been controversial, to say the least, but nowhere more than when applied to the problem of consciousness.

This paper discusses the problem of consciousness from the pragmatic perspective of a researcher in AI and robotics, where the focus is on how an intelligent agent can cope effectively with the overwhelming information content of the "pixel level" sensory input stream. Since humans are agents who cope successfully with overwhelming sensory information, we look to evidence from humans for clues about the more general problem and its possible solutions. Inspired by properties of human cognition, the paper proposes a computational model of consciousness and discusses its implications.

1.1. Recent approaches to consciousness

There have been a number of recent books on the problem of consciousness, many of them from a neurobiological perspective. The more clinically oriented books [3—5] often appeal to pathological cases, where consciousness is incomplete or distorted in various ways, to illuminate the structure of the phenomenon of human consciousness through its natural breaking points. Another approach, taken by Crick and Koch [6,7], examines in detail the brain pathways that contribute to visual attention and visual consciousness in humans and in macaque monkeys. Minsky [8], Baars [9], and Dennett [10] propose architectures whereby consciousness emerges from the interactions among large numbers of simple modules.

Philosophical writings on consciousness are useful where they help define and clarify the different questions to be answered. A particularly important distinction is between the "Easy" and "Hard" problems of consciousness [11]. The "Easy Problem" is relatively congenial to AI/robotics researchers and sympathizers [8—10], since it asks, *What does consciousness do for the agent, and how does it work?* Neuroscientists [6,7,12] ask a more restricted version of this question, *What are the neural correlates of consciousness?* The "Hard Problem" is far less tractable than either of these, since it asks, *Why does subjective experience feel like it does? In fact, how can it feel like anything at all?* This is closely tied to the question of the nature of "qualia" or "raw feels" [10,13]. The core issue behind the famous "Chinese Room" story [14] is the problem of Intentionality, which is, *How can knowledge in the mind of an agent refer to objects in the external world?*

In John Searle's recent book on the philosophy of mind [15], he articulates a position he calls *biological naturalism* that describes the mind, and consciousness in particular, as "entirely caused by lower level neurobiological processes in the brain." Although Searle rejects the idea that the mind's relation to the brain is similar to a program's relation to a computer, he explicitly endorses the notion that the body is a biological machine, and therefore that machines (at least biological ones) can have minds, and can even be conscious. In spite of being nothing beyond physical processes, Searle holds that consciousness is not *reducible* to those physical processes because consciousness "has a first-person ontology" while the description of physical processes occurring in the brain "has a third-person ontology." He lays out 11 central features of consciousness "that any philosophical—scientific theory should hope to explain."

In the following sections, I discuss the Easy Problem, the Intentionality Problem, and the Hard Problem of consciousness from perspective of the information-processing problems the robot must solve. I conclude after discussing how this approach responds to Searle's 11 features of consciousness.

1.2. Overview

The key ideas here are the following:

1. The sensory data stream presents information to the agent at an extremely high rate (gigabits/s).
2. This information is managed and compressed by selecting, tracking, and describing spatio—temporal portions of the sensory input stream.
3. A plausible coherent narrative is constructed to describe the recent history (500 ms or so) of the agent's sensorimotor interaction with the world.
4. An agent can autonomously learn its own intentionality by constructing computational models of hypothetical entities in the external world. These models explain regularities in the sensorimotor interaction, and serve as referents for the agent's symbolic knowledge representation.
5. The high information content of the sensory stream allows the agent to continually evaluate these hypothesized models, refuting those that make poor predictions.
6. The high information content of the sensory input stream accounts for the vividness and uniqueness of subjective experience.

2. The Easy Problem

The "Easy Problem" is: *What does consciousness do for us, and how does it work*? Only a philosopher could call this problem "Easy", since solving it will likely require decades at least, and dozens or hundreds of doctoral dissertations. What the name means is that scientists applying the methods of various disciplines have been able to formulate useful technical statements of the problem, and they have tools that apply to those problem statements. Progress may be difficult, but we know what it means. (The "Hard Problem" does not enjoy these benefits.)

2.1. Drinking from the firehose of experience

To a researcher in AI and robotics, one of the driving forces behind cognitive architecture is the need to cope with the enormous volume of sensory data, arriving on many asynchronous channels. The sensor stream $\mathbf{z}(t)$ in Fig. 1 is what I call "the firehose of experience"—the extremely high bandwidth stream of sensor data that the agent must cope with, continually. (A stereo pair of color cameras alone generates data at over 440 megabits per s.[1]) For a biological agent such as a human, the sense vector $\mathbf{z}$ contains millions of components representing the individual receptors in the two retinas, the cochleal cells in the two ears, and the many touch and pain receptors over the entire skin, not to mention taste, smell, balance, proprioception, and other senses.[2] With such a high data rate, any processing applied to the entire sensor stream must be simple, local, and parallel. In the human brain, arriving sensory information is stored in multiple short-term memory buffers, remains available for a short time, and then is replaced by newly arriving information [7].

In a biological agent, the motor vector $\mathbf{u}$ includes control signals for hundreds or thousands of individual muscles. An artificial robot could have dozens to hundreds of motors (though a simple mobile robot will have just two).

Following Harnad's extension [17] to Searle's Chinese Room metaphor [14], in addition to comparatively infrequent slips of paper providing symbolic input and output, we can visualize the room as receiving a huge torrent of sensory information that rushes in through one wall, flows rapidly through the room, and out the other wall, never to be recovered. Inside the room, John can examine the stream as it flows past, and can perhaps record fragments of the stream or make new symbolic notations based on his examination, in accordance with the rules specified in the room.

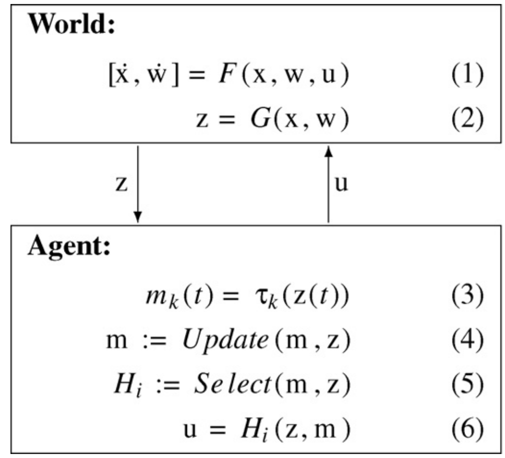

Figure 1 Dynamical model of a cognitive agent. A cognitive agent can be modeled at a high level as a dynamical system interacting with its environment. For the cognitive agent, its physical body is part of its environment. At any time t, the agent receives a sense vector $\mathbf{z}(t)$ and sends a motor vector $\mathbf{u}(t)$ to the world. The robot's body has state vector $\mathbf{x}(t)$ whose derivative is denoted $\dot{\mathbf{x}}(t)$, and the state vector of the rest of the world is described by $\mathbf{w}(t)$ and $\dot{\mathbf{w}}(t)$. The functions F and G represent the physics of the world and the sensor model, and neither is known to the agent. The agent acts by selecting a control law H_i based on the current sensor input z and the symbolic state $\mathbf{m}$ of its internal computational processes. Given this control law, Eqs. (1), (2) and (6) define a dynamical system, describing how the robot-environment system evolves until a new control law is selected. Meanwhile, trackers τ_k (Eq. (3)) contribute dynamic descriptions to the computational state $\mathbf{m}$. The robot's behavior alternates between (a) following a trajectory determined by a particular dynamical system until reaching a termination condition, and then (b) selecting a new control law H_j that transforms the coupled robot-environment system into a different dynamical system, with different trajectories to follow [36,1].

2.2. Trackers bridge the gap

The "firehose of experience" provides information at a rate much greater than the available symbolic inference and storage mechanisms can handle. The best we can hope for is to provide pointers into the ongoing stream, so that relevant portions can be retrieved when needed.

The primitive elements in this architecture are called *trackers*. Each tracker (e.g., τ_k in Eq. (3) in Fig. 1) can be thought of as having two ends. One end consists of a number of pointers into the firehose of experience, designating a spatio—temporal region in the input stream $\mathbf{z}(t)$, tracking that region

[1] 2 cameras $\times$ 640 $\times$ 480 pixels/frame/camera $\times$ 24 bits/pixel $\times$30 frames/s = 442 megabits/s.

[2] Because of differences in coding and processing, the comparison between camera and retina is not straight-forward. Koch, et al [16] estimate that a retina transmits approximately 10 megabits/s to the brain.

as its natural boundaries evolve in real time. The other end consists of a dynamic symbolic representation $m_k(t)$ of that particular portion of sensory experience, supporting inference about its static properties, its current state, and its history. The symbolic representation is "dynamic" in the sense that the values of certain attributes are automatically updated in real time by processes operating on the tracked elements of the sensory stream. For example, one tracker might describe the changing location and shape of a pedestrian walking through the agent's field of view. Another might dynamically describe the agent's pose within the frame of reference of the enclosing room as the agent moves through it.

Quine [18] describes human knowledge as a symbolic "web of belief" anchored at the periphery in sensory experience. Trackers are the anchoring devices for symbols. We say that a tracker is *bound to* a spatio—temporal segment of the sensor stream when that portion of ongoing experience satisfies the criteria of the tracker's defining concept, and when tracking is successful in real time. The tracker mediates between signal processing and symbol manipulation. At the signal processing end, the tracker implements a dynamical system keeping its pointers corresponding as closely as possible with the relevant portion of the sensor stream. At the symbol manipulation end, the tracker serves as a logical constant, with time-dependent predicates (*fluents* [19]) representing the attributes of the tracked object.

The location of the tracker within the sensor stream is regulated and updated by control laws [20], including dynamic state-estimation methods such as Kalman filters [21], responding to the image properties expected for the tracked concept and their contrast with the background.[3] Image processing strategies such as dynamical "snakes" [23] represent changing boundaries between figure and ground. With adequate contrast, the fine temporal granularity of the sensor stream means that updating the state of the tracker is not difficult. With increasing knowledge about the dynamics of the tracked object and its sensor image, the tracker can maintain a good expectation about the relevant portion of the sensor stream even in the presence of occlusion, poor contrast, and other perceptual problems.

The idea of trackers is not new. Versions of the sensorimotor tracker concept include Minsky's "vision frames" [24], Marr and Nishihara's "spatial models" [25], Ullman's "visual routines" [26], Agre and Chapman's "indexical references" [27], Pylyshyn's "FINSTs" [28], Kahneman and Triesman's "object files" [29], Ballard, et al., "deictic codes" [30], and Coradeschi and Saffiotti's "perceptual anchoring" [31].

Trackers make it possible to "use the world as its own model", directing attention to a particular aspect of the world to answer a query, rather than attempting to retrieve or infer an answer from stored knowledge [32,33]. Trackers may have a hierarchical structure, allowing the sensory image of a person, for example, to be tracked at varying levels of detail: entire body; head-torso-arms-legs; upper-arm-forearm-hand; palm-fingers; etc [25].

2.3. Creating a plausible coherent narrative

The cognitive architecture must be organized to make use of the information provided by the trackers. There appears to be a growing consensus that the mind includes a collection of processes that interact to create a *plausible coherent narrative* from its multiple parallel asynchronous sensory input streams [34,7]. This narrative is the agent's explanation to itself of what is going on around it. Just as an individual tracker provides an index into the sensory stream corresponding to a symbolic description, this plausible coherent narrative provides organizing structure on a temporal window into the agent's recent experience.

There is much work to be done to determine the precise structure of this architecture, but a consensus appears to be emerging on some sort of shared memory architecture such as Baars' Global Workspace Theory [9], drawing on Minsky's "Society of Mind" [8], Dennett's "multiple drafts" [10], and others. Neuroscientists in search of the "neural correlates of consciousness" [7,35] are identifying neural processes in the brain that appear to be participating in a similar architecture.

Natural selection ensures that, on average, these processes must do a pretty good job of describing the relevant objects and events in the external world. Some of the processes are innate to the individual, wired into the brain, but were "learned" by the species over evolutionary time. Other processes are learned by the individual from its own experience.

2.4. Defining consciousness

Addressing the Easy Problem, we claim that Fig. 1 provides a high-level description of what consciousness does for the agent and how it works functionally. This includes:

[3] This technology for tracking objects in the sensor stream has its roots in radar signal interpretation from the 1940s [22,21].

1. A high-volume sensor stream $\mathbf{z}(t)$ and a motor stream $\mathbf{u}(t)$ that are coupled, through the world and the robot's body, as described by Eqs. (1) and (2);
2. A non-trivial collection of trackers $m_k(t) = \tau_k(\mathbf{z}(t))$ grounded in the sensor stream (Eq. (3)) capable of providing dynamically updated symbolic descriptions for the agent's knowledge representation system, with top—down and bottom—up activation methods;
3. A non-trivial collection of control laws $\mathbf{u}(t) = H_i(\mathbf{z}(t), \mathbf{m}(t))$ (Eqs. (5) and (6)) from which the agent can select to implement reasonably reliable actions in the world;
4. A process for creating a plausible coherent narrative (which is part of $\mathbf{m}(t)$, the agent's internal model of itself and its environment) to explain the history of events in the agent's recent experience;
5. A sufficiently good correspondence between the actual properties of action and perception in the physical world (1) and (2), and the agent's symbolic theory of the world ($\mathbf{m}(t)$ (Eq. (4)) including symbols grounded via trackers (3) and actions implemented by control laws (5) and (6), so that the agent can interact effectively with its world. (This correspondence is the subject of the Intentionality Problem.)

The creation of the plausible coherent narrative is important. The narrative describes the events and activities that occupy the focus of consciousness. According to Global Workspace Theory [9], processes operating on the plausible coherent narrative determine which trackers fall within the "spotlight of consciousness", making them globally visible to the "audience" of active processes,[4] and which may simply participate in low-level control loops without conscious experience. Sometimes the shifts in the focus of consciousness from one thing to another have explanations in terms of events represented in the narrative ("Suddenly there was an explosion!" or "She walked into the room."), but other times, the cause of the shift may not be a represented event, so it appears to "just happen", like flipping the Necker cube.

Strictly speaking, this management of attention may not require a subjective, first-person view of the world (Sections 5.1 and 5.2). However, McDermott [37] and others argue that vivid conscious experience — the experience of qualia — is a retrospective phenomenon, using the plausible coherent narrative to access portions of the sensory input stream from the recent past. "Recent" in this case means within the last 50—500 ms, so a great deal of sensory information remains available in short-term sensory memory.

[4] This language simply describes access to a shared-memory communication channel, not a homunculus.

3. The Intentionality Problem

The "Intentionality Problem" is: *How can symbols in an internal cognitive knowledge representation refer to objects and events in the external world*? Or equivalently, *Where does meaning come from*? The core of Searle's "Chinese room" argument [14] is that the mind necessarily *has* intentionality (the ability to refer to objects in the world), while computation (the manipulation of formal symbols according to syntactic rules) necessarily *lacks* intentionality. Therefore (claims Searle), the mind must be more than a computation.

The same problem comes up in a much more pragmatic form in point 5 of the definition of consciousness presented previously, which specified that the internal knowledge representation, grounded by trackers in the sensory input stream, must correspond sufficiently well with the properties of the external world for useful predictions and actions to be possible. Since the agent has no direct access to the state of the world, and its only indirect access is through its own sensory and motor streams, the problem of establishing and maintaining such a correspondence is critical.

Note that Searle, and I, and everyone else, are born locked inside our own skulls, receiving coded information along nerve fibers, and sending coded responses along other nerve fibers. We humans face the same intentionality problem as the Chinese Room. The source of the meaning by which knowledge representations refer to the external world is as much a mystery for biological humans as it is for computational systems.

3.1. Learning from the pixel level

We have taken significant steps toward learning a hierarchy of models that explain "pixel level" input in the sensory stream in terms of hypothesized entities such as places, paths, objects, and so on. The Spatial Semantic Hierarchy [38,36] maps an unknown environment by identifying *locally distinctive states* and linking them into a topological map. The ability of a symbol to refer to a distinctive state in the physical environment depends on the behaviors of the dynamical systems defined by the control laws, not on any pre-existing intentionality in the set of symbols. Pierce and Kuipers [39] showed that these control laws could be learned from the

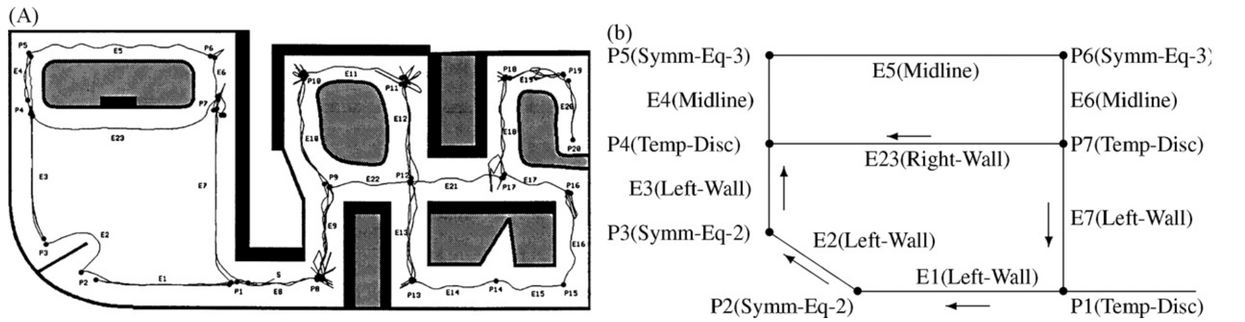

Figure 2 (a) A simulated robot applies the SSH exploration and mapping strategy. (b) A portion of the topological graph of distinctive places and connecting path segments identified from the behaviors of control laws in the environment. From Kuipers and Byun [38].

dynamical regularities in the robot's own experience with its uninterpreted sensors and effectors, constrained by their causal connections with the environment.

In the Spatial Semantic Hierarchy (SSH) [36], a robot explores and maps an unknown environment by identifying neighborhoods within which hill-climbing control laws can bring the robot reliably to isolated *locally distinctive states*. A trajectory-following control law carries the robot reliably from one distinctive state to the neighborhood of another, where hill-climbing brings it to the next distinctive state and prevents the accumulation of position error. The robot can thus abstract the continuous environment to a discrete topological map, with symbols representing places and paths, as well as distinctive states and the actions linking them. Fig. 2 shows the behavior trace of a robot exploring a simulated environment, hill-climbing to distinctive states equidistant from multiple obstacles, and following trajectories defined by midline- or wall-following control laws.

For the robot to learn these symbols *for itself*, it must learn its own collection of hill-climbing and trajectory-following control laws, starting with an uninterpreted set of sensors and effectors. Pierce and Kuipers [39] accomplished this task for a simulated robot with unknown sensors and effectors, which required learning the structure of a ring of sonar-like range sensors and learning an abstract model of turn and travel motor commands in terms of the velocities of the right and the left wheels.

Fig. 3 shows a lattice of learning methods that analyze data from several different experiments, building a progressively richer description of the sensory and motor systems, eventually supporting the creation of hill-climbing and trajectory-following control laws. Fig. 4 shows exploration traces corresponding to three different levels of competence during learning.

The steps of the learning process are [39,40]:

(1) Gather observations during random sequences of actions. First, coarsely cluster the sensors according to the qualitative properties of a histogram of values returned by each sensor. Then, within appropriate clusters, compute pairwise correlations among sensor values and interpret them as similarity measures.
(2) Assign the sensors in a cluster to positions in a high-dimensional mathematical space reflecting their pairwise similarities. Project to a low-dimensional subspace (2D in our examples) that best accounts for most of the variance in the cluster. Once sensor values have a spatial as well as temporal dependence, we can calculate spatial as well as temporal derivatives, and thus define motion fields.
(3) The motion fields corresponding to different motor signals are analyzed using principal component analysis to determine the most significant motion effects and the motor signals that correspond to them. These signals are used as the natural primitives for the motor space.
(4) Higher-level sensory features are proposed, based on the spatial and temporal attributes of the field of primitive sensory values. These include features such as discontinuities, local minimum and local maximum, with magnitude, position, and scope. Proposed features are evaluated according to stability, predictive power, and extensibility.
(5) Evidence is collected of the effects of primitive motor commands on higher-level features, searching for motor commands that change features in predictable ways. "Local state variables" are defined for particular neighborhoods in the environment. Trajectory-following and hill-climbing control laws are defined according to which local state variables correspond to features that are both observable and controllable.
(6) Open-loop control laws are defined by identifying commands that reliably change one feature

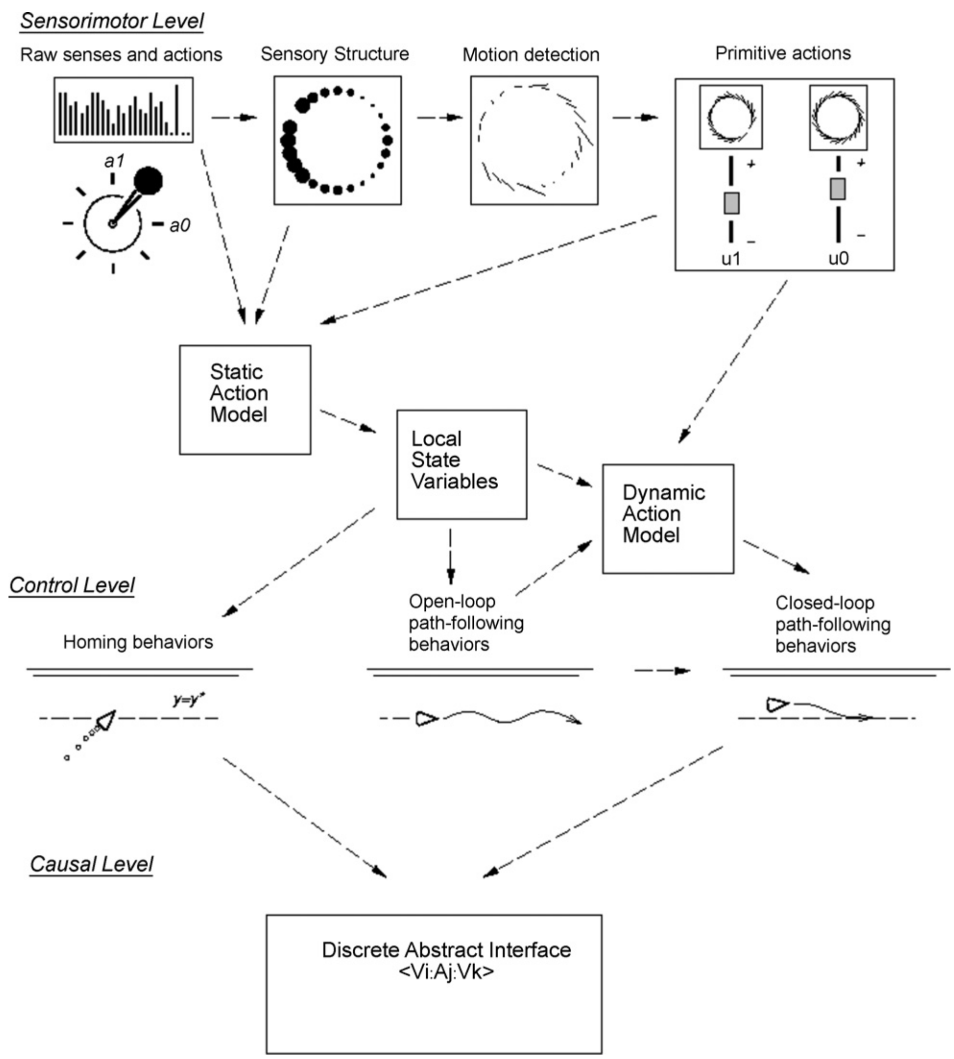

Figure 3 The lattice of learning methods and their results, from Pierce and Kuipers [39].

while keeping another one relatively constant. An open-loop control law terminates when the deviation of the "relatively constant" features passes a threshold. Closed-loop control laws are defined by searching for and identifying commands that can reduce deviations in the relatively constant feature, actively keeping it close to a desired setpoint. (Think of moving along a wall, turning slightly to maintain a desired distance from it. Compare Fig. 4(b) and (c).)

Higher-level sensory and motor features are learned without drawing on prior knowledge of the robot's environment. They are learned by identifying statistical and dynamical regularities in the experiences the robot receives after sending motor commands. In these experiments, the concepts whose intentionality is learned by the robot itself are a set of specific distinctive states (position and orientation), and the places and paths that make up a topological map of the environment.

Modayil and Kuipers [41] have developed these into methods for learning to individuate, track, and describe coherent objects from the "blooming, buzzing confusion" of sensory input, and then to learn meaningful actions to perform on them.

3.2. Learned models must be reliable, or else!

The learning agent constructs a model of the world. Each tracker posits the existence of external entities whose projections onto the sensors account for its portion of the sensory stream. The properties of those entities are inferred from, and account for, the information observed in the sensory stream. This is a version of Quine's "web of belief" [18].

The meanings to which the symbols in the agent's knowledge representation refer are the entities in

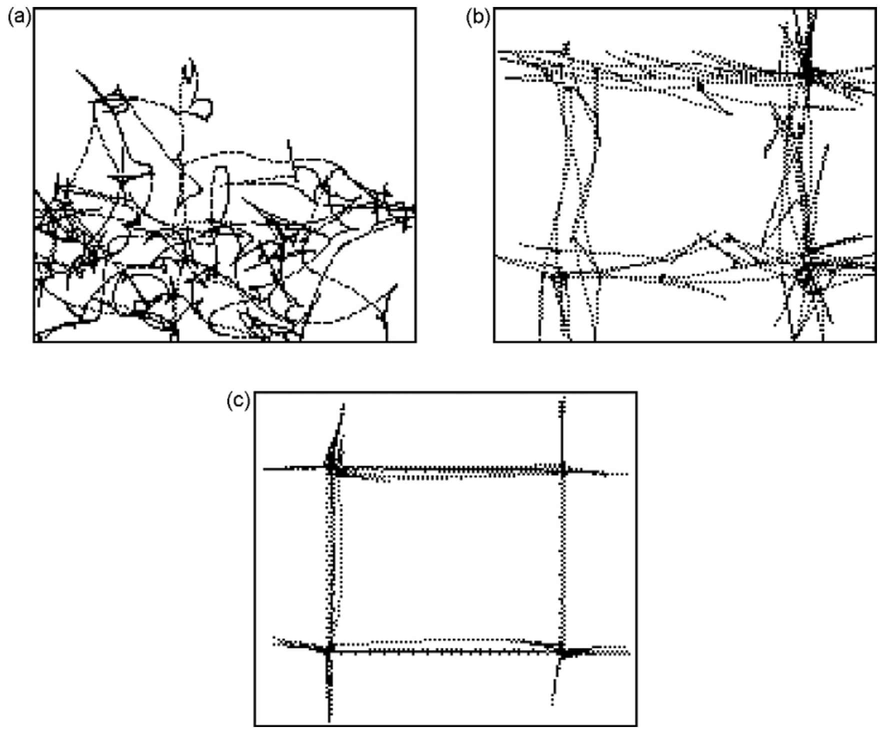

Figure 4 Exploring a simple world at three levels of competence. (a) The robot wanders randomly while learning a model of its sensorimotor apparatus. (b) The robot explores by randomly choosing applicable homing and open-loop path-following behaviors based on the static action model while learning the dynamic action model (see text). (c) The robot explores by randomly choosing applicable homing and closed-loop path-following behaviors based on the dynamic action model. From Pierce and Kuipers [39].

that constructed model. Meanings therefore do not reside in the external world, but in the constructed model we build of it. If the agent's constructed model corresponds sufficiently well with the external world, then it can function effectively, and it has created its own intentionality. When the internal model and external world diverge, plans and predictions fail. The sensory system provides relevant information that may be used to correct the model. If the divergence is sufficiently serious, the agent becomes non-viable, so natural selection ensures the quality of the constructed model.

4. The Hard Problem

We are not zombies (or at least I am not). Why not? It is undeniable that many experiences "feel like" something. Pain hurts, sugar tastes sweet, the sight of a loved one raises feelings that are strong and real, even though they cannot be fully articulated. In the words of Francisco Varela, "...*consciousness feels so personal, so intimate, so central to who we are*, ..." ([42], p. 226).

The "Hard Problem" of consciousness is: "*Why does consciousness feel like anything at all*?" Suppose we accept that the mind is a computation, running on the physical substrate of the brain. Why should a computational process — even one that creates trackers, builds a plausible coherent narrative from its experience, and constructs its own intentionality — feel like anything at all to the owner of the brain?

This problem is Hard, even to a philosopher. So far it has resisted all attempts even to state what it would mean to provide a solution. It's not just that we cannot find a solution. We cannot even figure out what a solution would look like.

However, perhaps we can sneak up on the Hard Problem and get a closer peek at what makes it tick. Rather than directly approach the question of why anything feels like anything at all, we will ask why some experiences are more vivid than others. My claim is that the vividness of subjective experience — of qualia — is directly related to its information content.

4.1. Why are some experiences more vivid than others?

Looking *now* at the apple sitting in front of me, I experience a vivid perceptual image. This apple is very round, and a light greenish-yellow with a few brownish-red streaks. (Someone else might describe it as yellowish-green or with reddish-brown streaks.) This is a classic *quale*, a primary sensory experience. I claim that it is vivid, in part, because my internal symbolic concept of this apple is directly bound to the corresponding region in my sensory input stream, which provides a huge flow of information in real time.

Revising these words *now*, some months later, my recalled image of that apple is not nearly as vivid as the experience itself. I can recall fragments of the sensory experience of perceiving the apple, but they are clearly incomplete. Even with effort, there are questions about the apple that I cannot answer now, with the amount of stored information available to me, that could have been effortlessly answered during the experience itself, simply by shifting my focus of attention within the sensor stream.

Just *now*, you the reader have read descriptions of my experience with a particular apple. But it is not your apple, or your experience, so your concept of this apple is much less vivid than either direct experience or personal memory.

These stories illustrate three widely separated points on a spectrum of the vividness of subjective experience. In each case, the information content of a particular experience is determined by the number of alternatives in the universe of possible experiences this one was drawn from. Ongoing visual experience of a particular object has the information content of the support for the tracker for that object in the sensory stream. The perception of an object is much more vivid when the observer fixates on it, which means that its image falls on the foveas in the two retinas, so many more receptive fields contribute information to the object's tracker.

The stored memory of a personal experience must contain far less information than the original experience, but that amount of information must still be very large, since it can apparently contain some sort of sensory snapshots as well as symbolic descriptions. A written word such as "apple" is encoded in a few dozen bits. Thus, differences in vividness appear to correspond well with differences in information content.

4.2. Why is subjective experience so personal?

Like the proverbial snowflakes, no two separately-created multi-mega-pixel digital camera images are ever identical, simply because of the huge number of bits they encode, and the number of unpredictable physical processes that determine those bits. Likewise with sensory experiences.

Each fragment of the sensory stream during direct experience, and even each sensory snapshot stored in long-term memory, has huge information content. The large number of bits in a particular sensory experience makes it astronomically unlikely that any other individual could have precisely the same experience.[5]

The long-term episodic memory of an adult contains an enormous number of representations of particular sensory experiences. If the stored representation of one experience is astronomically unlikely to be duplicated, the entire collection is far more likely to be unique and personal to the individual agent.

This combinatorial argument is strengthened further by the observation that long-term memory includes not only snapshots of experience, but also an intricate web of associative links among concepts. Many of these links correspond to familiar semantic relationships (e.g., generalization, specialization, part-of, plus concept-specific relations). These are important, but they reflect structures in the external world that might be learned in similar form by different agents. However, there are also associations from vivid sensory images (for example, a characteristic smell) to episodes buried deep within long-term memory. Not only does the configuration of such links further increase the information content of long-term memory, but the unpredictability and distinctiveness of that collection of links would be evident to the individual agent from its own experience.

Note that this argument depends on the number and unpredictability of the low-level processes that create sensory experiences (and to a lesser extent, on the richness of the environment). If the memory of an artificial agent is created purely through symbolic input, or if bulk memory can be backed up and restored as with a disk drive, then it may become possible to create complex individuals that have identical memory states, at least momentarily.

4.3. Why does subjective experience feel like anything at all?

In sneaking up on the Hard Problem, we have focused on why different experiences feel different. We have not solved the Hard Problem of why subjective experience feels like *anything at all*.

[5] Furthermore, as Sloman and Chrisley observe [43], even bit-by-bit identity of stored representations would not imply that the experiences of two agents were the same, since the interpretation of those representations depends on the entire computational context of each agent.

When a question is this difficult, perhaps it is a non-question. As molecular biologists continue to tease out the mysteries of the genetic code, molecular self-replication, and the functioning of the complex machinery by which the genes influence the cells, they have not identified a specific answer to the question, *What is life?* There is no magical elixir that distinguishes non-living from living matter. Rather, when examined closely, there is a vast spectrum of complexity of molecular behavior.

Perhaps the same is true of subjective experience. Thermostats and autofocus cameras interact with their environments in response to their perceptions, but their versions of Fig. 1 are too simple to have recognizable subjective experience. We humans are vastly more complex in terms of the numbers and variety of processes taking place. Perhaps subjective experience just *is* the operation of a highly complex information-processing mechanism with very high information content as well as input and output.

Subjective experience feels like it does because, in the first instance, our bodies are designed to process certain information in certain ways. Pain is insistent and intrusive because its biological purpose is to attract our attention to a threat. We learn to experience higher-level sensations through their associations with everything else in our cognitive states.

By this argument, just as biologists seldom ask themselves, "What is life?", cognitive scientists (including roboticists) need not dwell on the Hard Problem. The Easy Problem and the problem of Intentionality both have functional import, and are likely to provide plenty for us to do.

5. Evaluating a theory of consciousness

It is not yet possible to build a robot with sufficiently rich sensorimotor interaction with the physical environment, and a sufficiently rich capability for tracking and reasoning about its sensor and motor streams, to be comparable with human consciousness. The remaining barriers, however, appear to be technical rather than philosophical.

We begin evaluating this theory of consciousness by discussing how well such a computational model might account for 11 central features of consciousness "that any philosophical—scientific theory should hope to explain" ([15], pp. 134—145). Each of the following subsections is titled with Searle's name for a feature, followed by a quote from his description of it.

For some of these features — Qualitativeness, Subjectivity, Intentionality, Distinction between Center and Periphery, and Active and Passive — the dynamical tracker model of consciousness provides a specific *explanation*. For other features — Sense of Self, Unity, Situatedness, Gestalt Structure, Mood, and Pleasure/Unpleasure — there may be several possible explanations, all of which are *expressible* within the dynamical tracker model.

5.1. Qualitativeness

Every conscious state has a qualitative feel to it. . . .[This includes] *conscious states such as feeling a pain or tasting ice cream* . . .[and also] *thinking two plus two equals four* ([15], p. 134).

The vividness, intensity, and immediacy of subjective experience are due to the enormous information content of the sensor stream $\mathbf{z}(t)$. There's a difference between thinking about the color red with my eyes closed in a dark room, and the immediate experiences (*qualia*) of seeing a red rose or apple or sunset. The intensity of subjective experience increases with the information content of the input: from text or verbal descriptions, to viewing a color photograph, to memories or dreams of experiences, to live multisensory experience.

Trackers do not capture the experience itself, but they provide structure to William James' "one great blooming, buzzing confusion". By providing rapid access to specified parts of the sensory stream, trackers (in vision at least) maintain the illusion that the entire visual field is perceived with the same high fidelity as the point of foveal attention [30,33].

If an attribute value such as the color red is stored as a symbol "red" in memory, its information content is determined by the number of other possible color symbols that could have been stored as a value of that attribute: at most a dozen bits or so. On the other hand, if a tracker is bound to a region in the sensor stream, the number of bits of color information streaming past, even in a small region, is orders of magnitude larger.

The higher information content of the sensor stream means that attribute values drawn from sensory experience necessarily include more distinctions than are available in, for example, common vocabulary. The reds of roses, apples, and sunsets are different, though their distributions may overlap. The agent who has experienced these qualia possesses more distinctions than one who has not, and can recognize the rarity of a sunset-colored rose.

Although it is implausible for the entire sensory stream to be stored in long-term memory, at least some qualia (e.g., the pain of a kidney stone or the smell of a freshly-baked madeleine) can be stored in memory and can serve as associative links to memories of previous experiences. The high information content of a quale makes it possible to select out a single distinctive association from the huge contents of long-term memory.

There is a compelling argument that perception requires abduction [44]. There must be a process that proposes hypotheses to account for ongoing sensor data. If this process were purely bottom—up (i.e., driven by the sensor data), then in the absence of the sensor stream, no hypotheses would be generated and no perception would take place. However, experience suggests that there are significant top—down and perhaps random processes for generating hypotheses. Under conditions of sensory deprivation, people tend to hallucinate, that is, to generate perceptual hypotheses poorly grounded in sensor input [45].

The practical value of qualia is that they help keep the hallucinations down. Symbolic (logical) theories are subject to multiple interpretations. Larger theories have fewer interpretations. Sensory grounding through trackers provides a huge number of additional axioms to such a theory, and thereby constrains its possible interpretations. Active trackers provide strong evidence, solidly grounded in the sensor stream, to eliminate incorrect perceptual hypotheses. Thus, the qualitativeness of experience corresponds to the high information content of the sensor stream, which is pragmatically important to the cognitive agent because it helps to keep the generation and refutation of perceptual hypotheses in balance.

5.2. Subjectivity

Because of the qualitative character of consciousness, conscious states exist only when they are experienced by a human or animal subject. ... Another way to make this same point is to say that consciousness has a first-person ontology ([15], p. 135).

Consciousness is experienced exclusively from a first-person point of view. (I reject Searle's explicit restriction of conscious experience to "a human or animal subject".)

What it means for an agent to have a first-person point of view is for it to have access to the sensor and motor streams from its own body. That is, its body is physically embedded in the world, and Eqs. (1) and (2) describe the causal path from its actions $\mathbf{u}$ to its perceptions $\mathbf{z}$. By selecting a control law H_i, the agent creates a causal path from its sensory input $\mathbf{z}$ to its motor output $\mathbf{u}$, closing the loop and giving it some degree of control over its body. Only the agent itself has access to the sensor stream $\mathbf{z}(t)$ from its own body, or to the motor output stream $\mathbf{u}(t)$, and only the agent is in a position to select and impose a control law H_i relating them. (This individuation reflects biology. Robots may not have the same constraints. Also see the 1999 movie, *Being John Malkovich*.)

The agent can learn from experience which aspects of its perceptions are under its direct control, and which are not, therefore learning to separate its concept of itself ($\mathbf{x}$) from its concept of its environment ($\mathbf{w}$) [46,47]. This distinction comes not from anatomy, but from the existence of tight control loops. Virtual reality and telepresence are subjectively compelling exactly because humans are quickly able to learn novel models of senses, actions, body, and world from interactive experience.

Exceptions to the agent's privileged access to its own sensorimotor system confirm this description of the first-person point of view. Searle ([15], p.142) cites an experiment by the neurosurgeon Wilder Penfield, who was able to stimulate a patient's motor neurons directly, to raise the patient's arm, prompting the patient to say, "I didn't do that, you did." This corresponds to the surgeon being able to set $\mathbf{u}(t)$ directly, without the patient selecting a control law.

5.3. Unity

At present, I do not just experience the feelings in my fingertips, the pressure of the shirt against my neck, and the sight of the falling autumn leaves outside, but I experience all of these as part of a single, unified, conscious field ([15], p. 136).

We experience the audio—visual surround as a single unified field, continuous in space and time, in spite of a variety of contradictory findings about our actual sensory input [7]. For example, the fovea has vastly higher resolution than the periphery of the retina, and the blind spot has no resolution at all. The density of color-receptive cones is even more strongly biased toward the foveal area and away from the periphery. Auditory and visual evidence from the same event reaches the brain at different times.

The apparent unity of perception is a plausible coherent narrative, constructed 50-500 ms after the fact from evidence from parallel and irregular sources [30,33]. Several mechanisms and cognitive architectures have been proposed to explain how this narrative is constructed. For example, Minsky's "Society of Mind" [8], Baars' "Global Workspace

Theory" [9], Dennett's "Multiple Drafts Model" [10], and others, propose that consciousness arises from the interaction of many simple cognitive modules that observe and control each other. The generally-linear stream of conscious thought is constructed, typically in fragments, by these modules from each others' outputs. Within this kind of architecture, trackers are the modules that interface between the sensor stream and the symbolic cognitive modules.

A number of technical and scientific questions remain to be answered about how the coherent conscious narrative is actually constructed from parallel and irregular sources of input. Global Workspace Theory [9,34] appears to be the current best detailed computational model of this process.

In robotics, the Kalman Filter [21] is often used to predict the most likely trajectory of a continuous dynamical system (along with its uncertainty), given a model and an irregular collection of sensor observations (along with their uncertainties). The technical methods are different, but philosophically, the slightly retrospective construction of a plausible coherent narrative from irregular observations is no more problematical than a Kalman Filter.

5.4. Intentionality

My present visual perception, for example, could not be the visual experience it is if it did not seem to me that I was seeing chairs and tables in my immediate vicinity. This feature, whereby many of my experiences seem to refer to things beyond themselves, is the feature that philosophers have come to label "intentionality" ([15], p. 138).

The core of Searle's "Chinese room" argument [14] is that strong AI commits a category error with regard to intentionality. The mind necessarily *has* intentionality (the ability to refer to objects in the world), while computation (the manipulation of formal symbols according to syntactic rules) necessarily *lacks* intentionality. Therefore, the mind cannot be a computation.

However, intentionality is exactly what the tracker for a high-level concept delivers: it binds a portion of the current sensor stream to the symbolic description of an object (believed to be) in the external world. The relationship of intentionality follows from the causal connection from the external, physical world to the contents of the sensor stream, and thence to the internal symbols created by the trackers.

Searle's response to the "Robot Reply" [14] acknowledges the importance of the causal connection between a robot's sensorimotor system and the world, but he claims that uninterpreted sensor and motor signals are just as free of intentionality as any formal symbols.

Presumably, Searle would argue that the intentionality provided by a tracker is merely "derived intentionality," coming from the mind of the human who programmed the algorithms and control laws that make the tracker work. This argument is vulnerable to a demonstration that effective trackers can be learned automatically from experience with uninterpreted sensors and effectors. As discussed in Section 3.1, Pierce and Kuipers [39] have made a preliminary demonstration of just this. Other highly relevant work on the same problem includes [33,46,48,49].

We have taken significant steps toward learning intentionality. The Spatial Semantic Hierarchy [36] maps an unknown environment by identifying *locally distinctive states* and linking them into a topological map. The ability of a symbol to refer to a distinctive state in the physical environment depends on the behaviors of the dynamical systems defined by the control laws, not on intentionality in the pre-existing set of symbols. Pierce and Kuipers [39] showed that these control laws could be learned from the dynamical regularities in the robot's own experience with its uninterpreted sensors and effectors, constrained by their causal connections with the environment. Modayil and Kuipers [50,41] have used related methods to learn to individuate and describe coherent objects from the "blooming, buzzing confusion" of sensory input.

We believe that learning methods like these can be extended to learn trackers for many kinds of distinctive configurations in the sensory stream. New symbols are defined, and their properties are learned, to refer to the objects of the trackers in the external world. The agent thus acquires intentionality *of its own*.

5.5. The Distinction between the Center and the Periphery

Some things are at the center of my conscious field, others are at the periphery. A good mark of this is that one can shift one's attention at will. I can focus my attention on the glass of water in front of me, or on the trees outside the window, without even altering my position, and indeed without even moving my eyes. In some sense, the conscious field remains the same, but I focus on different features of it ([15], p. 140).

This is quite a good description of the "spotlight of attention" from Baars' Global Workspace Theory [9]. There are many computational processes at work in the mind, some of them trackers grounded in the sensor stream, others using the output of trackers, directly or indirectly. The "spotlight of attention" amounts to sending the output of certain processors to a globally accessible workspace (the "blackboard" metaphor is sometimes used), available to every process in the system.

An individual tracker maintains a set of pointers into the sensor input stream that defines the features it attends to. The rest of the sensor stream is examined only enough to continue to track successfully, and to allow "pop-out" detection. When certain trackers are within the agent's focus of attention, they constitute the "figure" part of the "figure-ground" distinction in the visual field. Other trackers outside the current focus of attention may track "ground" objects that could be attended to later, or they may contribute to maintaining *Situatedness* (next section).

Thermostats and robot vacuum cleaners are coupled with the world to form simple dynamical systems. However, they fail to be conscious because they have a single fixed "figure", no "ground" at all, no plausible coherent narrative, no ability to store qualia or use them for retrieval, and no ability to shift focus of attention.

5.6. Situatedness

All of our conscious experiences come to us with a sense of what one might call the background situation in which one experiences the conscious field. The sense of one's situation need not be, and generally is not, a part of the conscious field. But, normally I am in some sense cognizant of where I am on the surface of the earth, what time of day it is, what time of year it is, whether or not I have had lunch, what country I am a citizen of, and so on with a range of features that I take for granted as the situation in which my conscious field finds itself ([15], p. 141).

While the concept of the tracker is particularly clear when applied to images of objects that move within the visual field, it applies equally well to tracking the location of the robot within a given frame of reference, for example, localization within the current enclosing room. This concept of tracker can, in turn, be generalized to track motion through an abstract space such as time or a goal hierarchy. Such background situation trackers could potentially continue tracking with little or no attention.

5.7. Active and Passive Consciousness

The basic distinction is this: in the case of perception (seeing the glass in front of me, feeling the shirt against my neck) one has the feeling, I am perceiving this, and in that sense, this is happening to me. In the case of action (raising my arm, walking across the room) one has the feeling, I am doing this, and in that sense, I am making this happen ([15], p. 142).

The agent's sensorimotor interface (Eqs. (2) and (6)) clearly divides into the sensor stream $\mathbf{z}(t)$, which is happening to the agent, and the motor stream $\mathbf{u}(t)$, by which the agent makes things happen.

Active control of perception by moving the eyes to bring a target into the fovea is accomodated by the current model, since the state of the eyes would be part of the robot's state vector $\mathbf{x}(t)$, and would be controlled by the motor vector $\mathbf{u}(t)$. Attentional processes such as giving a particular tracker more resources and allowing it to fill out its hierarchical structure more fully, could also be modeled as control laws whose effect is on the internal state $\mathbf{m}(t)$ of the agent.

5.8. The Gestalt structure

We do not, for example, in normal vision see undifferentiated blurs and fragments; rather, we see tables, chairs, people, cars, etc., even though only fragments of those objects are reflecting photons at the retina, and the retinal image is in various ways distorted. The Gestalt psychologists investigated these structures and found certain interesting facts. One is, the brain has a capacity to take degenerate stimuli and organize them into coherent wholes. Furthermore, it is able to take a constant stimulus and treat it now as one perception, now as another ([15], p. 143).

Each tracker looks for a certain structure in the sensor stream. When it finds it, that structure is foreground for that tracker, and the rest of the sensor stream is background. The findings of the Gestalt psychologists provide clues about the properties of individual trackers, of the process by which potential trackers are instantiated and become active, of the ensemble of active trackers, and perhaps even of the learning process by which trackers for new types of objects are learned.

For example, interpretation-flipping figures such as the Necker cube or the duck/rabbit figure suggest properties of the ensemble of active trackers, such as mutual exclusion and continued competition among the higher level of hierarchical trackers, while lower levels preserve their bindings and can be used by either competing interpretation.

5.9. Mood

All of my conscious states come to me in some sort of mood or other. . . . there is what one might call a certain flavor to consciousness, a certain tone to one's conscious experiences ([15], p. 139).

The relation between a human agent's psychochemical state (a component of $\mathbf{x}(t)$), mood (a component of $\mathbf{z}(t)$), and the rest of the agent's perception, is presumably embedded in the complex and unknown functions F and G. How mood affects behavior is embedded (in part) in the mechanism for selecting the next control law H_i.

5.10. Pleasure/Unpleasure

Related to, but not identical with, mood is the phenomenon that for any conscious state there is some degree of pleasure or unpleasure. Or rather, one might say, there is some position on a scale that includes the ordinary notions of pleasure and unpleasure ([15], p. 141).

The pleasure/unpleasure scale has a natural role as a reward signal during reinforcement learning. The links between particular qualia and their positions on the pleasure/unpleasure scale are very likely determined by evolutionary learning [10]. For example, pain is unpleasant and sex is pleasant, surely because of their functional roles in the survival of the individual and the species.

5.11. The Sense of Self

It is typical of normal conscious experiences that I have a certain sense of who I am, a sense of myself as a self ([15], p. 144).

In many ways, the most pragmatically useful aspect of consciousness is the ability to observe, describe, store, recall, reflect on, and in some ways control one's own thoughts, experiences, goals, plans, and beliefs.

As we have seen, the apparently sequential and continuous nature of conscious experience is the post-hoc construction of a plausible coherent narrative to explain a somewhat irregular collection of sensory inputs.

Once such a narrative exists, it can be stored in long-term memory, recalled, and reasoned about like any other piece of symbolic knowledge. The construction, storage, recall, and manipulation of this kind of knowledge poses no fundamental difficulties for computational modeling [51,43].

It is important to acknowledge that memory can store more than abstracted symbolic descriptions of experience. Memory can include qualia such as snapshots or fragments of the sensory stream with high information content. The *content* of the conscious narrative, as well as the content of these qualia, and the associations they provide into the agent's long-term memory, are highly specific to the agent whose experience they reflect, so they contribute to a unique "sense of self."[6]

6. Conclusions

We approach the problem of consciousness from the pragmatic design perspective of AI and robotics. One of the major requirements on an embodied agent is the ability to cope with the overwhelming information content of its own sensory input (the "firehose of experience"). One cognitive architecture that meets this requirement includes *trackers* that ground dynamic symbolic descriptions in spatio—temporal regions of the sensory stream, and a *plausible coherent narrative* that explains the objects and events from the external world that are observed in a temporal window on the sensory stream. Researchers from a variety of perspectives appear to be converging on such an architecture, which would be a solution to the Easy Problem of consciousness.

The claim presented here (a "strong AI" claim) is that the conditions for consciousness are expressible as a computational model, including dynamical trackers maintaining symbolic references to perceptual images in the sensor stream. The phenomenal character of consciousness ("what it is like") comes from the enormous flow of information in the sensory stream, and from the turbulent "churn" of process activation, on the way to being serialized as conscious thought [9].

Qualia reflect the information density of the sensor stream. Trackers ground a symbolic knowledge representation in the "firehose of experience" and constrain its interpretations. Intentionality is intrinsic if useful trackers can be learned by bootstrapping up from uninterpreted experience. And the sequential stream of subjective consciousness is a plausible coherent narrative, constructed retrospectively (by 500 ms or so).

The Intentionality Problem applies to any embodied agent, human or robot, that interacts with the world only through coded sensor and motor signals. We argue that there is no magic, for humans or robots, whereby symbols inside the mind can refer, directly and correctly, to corre-

[6] *"What do you see when you turn out the light? I can't tell you, but I know it's mine."*— John Lennon.

sponding objects in the outside world. On the other hand, we can exhibit early versions of learning algorithms that can construct explanations for the regularities of pixel-level sensorimotor interaction in terms of higher-level entities such as places, paths, objects and actions. The "meaning" of a symbol in the internal knowledge representation is an entity hypothesized by such a learning algorithm, that is, another internal construct. If these internal entities correspond usefully with the external world, the agent will be able to plan and act effectively. If not, it is unlikely to survive.

We have attempted to "sneak up" on the Hard Problem by offering relative information content as an explanation for why different experiences have different levels of vividness. This leaves open the Hard Problem itself: *Why should any amount of information transfer feel like anything at all*? However, we do know that information transfer in an embodied agent necessarily corresponds to some sort of physical state-changes, which must have physical correlates that can be sensed. Drawing on an analogy with classic models of emotion, we may speculate that it is the physical correlates of raw information transfer that "feels like anything at all", and that "what it feels like" depends on the content of the information and the cognitive and behavioral context of the agent.

The empirical and philosophical study of consciousness in humans helps clarify the nature of the phenomenon. The study of the brain helps us understand the one implementation of a conscious system in whose existence we have confidence. But according to our claim, consciousness is not restricted to biological implementation. The essential features of consciousness can, in principle, be implemented on a robot with sufficient computational power and a sufficiently rich sensorimotor system, embodied and embedded in its environment.

Acknowledgements

This work has taken place in the Intelligent Robotics Lab at the Artificial Intelligence Laboratory, The University of Texas at Austin. Research of the Intelligent Robotics lab is supported in part by grants from the Texas Advanced Research Program (3658-0170-2007), from the National Science Foundation (IIS-0413257, IIS-0713150, and IIS-0750011), and from the National Institutes of Health (EY016089). Portions of this paper have been presented previously in [1,2].

References

[1] Kuipers B. Consciousness: drinking from the firehose of experience. In: Proceedings of the 20th National Conf. on Artificial Intelligence (AAAI-05). AAAI Press; 2005. p. 1298—305.
[2] Kuipers B. Sneaking up on the Hard Problem of consciousness. In: Chella A, Manzotti R, editors. AI and consciousness: theoretical foundations and current approaches, AAAI Fall Symposium Series. AAAI Press; 2007. p. 84—9.
[3] Sacks O. The man who mistook his wife for a hat and other clinical tales. New York, NY: Simon & Schuster; 1985.
[4] Damasio A. The feeling of what happens. New York: Harcourt, Inc; 1999.
[5] Ramachandran VS. A brief tour of human consciousness. New York: Pi Press; 2004.
[6] Crick F, Koch C. A framework for consciousness. Nat Neurosci 2003;6(2):119—26.
[7] Koch C. The quest for consciousness: a neurobiological approach. Englewood CO: Roberts & Company Publisher; 2003.
[8] Minsky M. The society of mind. NY: Simon and Schuster; 1985
[9] Baars BJ. A cognitive theory of consciousness. New York: Cambridge University Press; 1988.
[10] Dennett D. Consciousness explained. Boston: Little, Brown & Co; 1991.
[11] Chalmers DJ. The conscious mind: in search of a fundamental theory. New York: Oxford University Press; 1996.
[12] Tononi G. An information integration theory of consciousness. BMC Neurosci 2004;5:42. http://www.biomedcentral.com/1471—2202/5/42.
[13] Humphrey N. Seeing red: a study in consciousness. Cambridge MA: Harvard University Press; 2006.
[14] Searle J. Minds, brains, and programs. Behav Brain Sci 1980;3:417—24.
[15] Searle JR. Mind: a brief introduction. New York: Oxford University Press; 2004.
[16] Koch K, McLean J, Segev R, Freed MA, Berry II MJ, Balasubramanian V, et al. How much the eye tells the brain. Curr Biol 2006;16:1428—34.
[17] Harnad S. Minds, machines and Searle. J Exp Theor Artif Intell 1989;1:5—25.
[18] Quine WVO. Two dogmas of empiricism. In: Quine WVO, editor. From a logical point of view. Second revised edition, Cambridge, MA: Harvard University Press; 1961 . p. 20—46.
[19] McCarthy J, Hayes PJ. Some philosophical problems from the standpoint of artificial intelligence. In: Meltzer B, Michie D, editors. Machine intelligence 4. Edinburgh: Edinburgh University Press; 1969. p. 463—502.
[20] Hutchinson S, Hager GD, Corke PI. A tutorial on visual servo control. IEEE Trans Robot Automat 1996;12(5):651—70.
[21] Gelb A. Applied optimal estimation. Cambridge, MA: MIT Press; 1974.
[22] Wiener N. Cybernetics or control and communication in the animal and the machine. Cambridge MA: MIT Press; 1948.
[23] Blake A, Yuille A. Active vision.. Cambridge, MA: MIT Press; 1992.
[24] Minsky M. A framework for representing knowledge. In: Winston PH, editor. The psychology of computer vision. NY: McGraw-Hill; 1975. p. 211—77.
[25] Marr D, Nishihara HK. Representation and recognition of the spatial organization of three-dimensional shapes. Proc R Soc B 1978;200:269—94.
[26] Ullman S. Visual routines. Cognition 1984;18:97—157.

[27] Agre PE, Chapman D. Pengi: An implementation of a theory of activity. In: Proceedings of the 6th National Conference on Artificial Intelligence (AAAI-87). MIT Press; 1987. p. 268—72.

[28] Pylyshyn ZW. The role of location indexes in spatial perception: a sketch of the FINST spatial-index model. Cognition 1989;32:65—97.

[29] Kahneman D, Treisman A. The reviewing of object files: object-specific integration of information. Cognit Psychol 1992;24:175—219.

[30] Ballard DH, Hayhoe MM, Pook PK, Rao RPN. Deictic codes for the embodiment of cognition. Behav Brain Sci 1997;20(4): 723—67.

[31] Coradeschi S, Saffiotti A. An introduction to the anchoring problem. Robot Autonomous Syst 2003;43(2—3):85—96.

[32] Ballard D, Hayhoe M, Pelz J. Memory representations in natural tasks. Cognit Neurosci 1995;7:66—80.

[33] O'Regan JK, Noë A. A sensorimotor account of vision and visual consciousness. Behav Brain Sci 2001;24(5):939—1011.

[34] Baars BJ. The conscious access hypothesis: origins and recent evidence. Trends Cognit Sci 2002;6(1):47—52.

[35] Edelman G. Neural Darwinism: the theory of neuronal group selection. NY: Basic Books; 1987.

[36] Kuipers B. The spatial semantic hierarchy. Artif Intell 2000;119:191—233.

[37] McDermott DV. Mind and mechanism. Cambridge MA: MIT Press; 2001.

[38] Kuipers BJ, Byun Y-T. A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations. J Robot Autonomous Syst 1991;8:47—63.

[39] Pierce DM, Kuipers BJ. Map learning with uninterpreted sensors and effectors. Artif Intell 1997;92:169—227.

[40] Kuipers B, Beeson P, Modayil J, Provost J. Bootstrap learning of foundational representations. Connection Sci 2006;18(2): 145—58.

[41] Modayil J, Kuipers B. Autonomous development of a grounded object ontology by a learning robot. In: Proceedings of the 22nd National Conference on Artificial Intelligence (AAAI-07). AAAI Press; 2007. p. 1095—101.

[42] Blackmore S. Conversations on consciousness. New York: Oxford University Press; 2006.

[43] Sloman A, Chrisley R. Virtual machines and consciousness. J Consciousness Studies 2003;10(4—5):133—72.

[44] Shanahan M. Perception as abduction: turning sensor data into meaningful representation. Cognit Sci 2005;29:103—34.

[45] Lilly JC. The center of the cyclone. New York: The Julian Press; 1972.

[46] Philipona D, O'Regan JK, Nadal J-P. Is there something out there? Inferring space from sensorimotor dependencies. Neural Comput 2003;15:2029—49.

[47] Kemp CC, Edsinger A. What can I control?: the development of visual categories for a robot's body and the world that it influences. In: International Conference on Development and Learning (ICDL); 2006.

[48] Choe Y, Smith NH. Motion-based autonomous grounding: inferring external world properties from encoded internal sensory states alone. In: Proceedings of the 21st National Conference on Artificial Intelligence (AAAI-06). AAAI Press; 2006. p. 936—41.

[49] Olsson LA, Nehaniv CL, Polani D. From unknown sensors and actuators to actions grounded in sensorimotor perceptions. Connection Sci 2006;18(2):121—44.

[50] Modayil J, Kuipers B. Bootstrap learning for object discovery. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-04). IEEE Computer Society Press; 2004. p. 742—7.

[51] Minsky ML. Matter, mind, and models. In: Minsky M, editor. Semantic information processing. Cambridge, MA: MIT Press; 1968. p. 425—32.