Introduction: the constructive interdisciplinary viewpoint for understanding mechanisms and models of imitation and social learning

Chrystopher L. Nehaniv and Kerstin Dautenhahn

1 Introduction

Social learning, matched behaviour and imitation are important classes of mechanisms whereby knowledge may be transferred between agents (biological, computational or robotic autonomous systems). They comprise key mechanisms necessary for the evolution and development of social intelligence and culture. Researchers from across disciplines have begun coming together to understand these mechanisms with ever more sophisticated models.

While the importance of Social Learning has grown increasingly apparent to psychologists, ethologists, philosophers, linguists, cognitive scientists and computer scientists, biologists, anthropologists and roboticists, the workers in the field are often unaware of relevant research by others in other disciplines. Social learning has lacked a rigorous foundation and only very few major interdisciplinary publications have been available on the subject for researchers in artificial intelligence or psychology interested in realizations of the mechanisms they study. By bringing social learning techniques into computer and robotic systems, the door is being opened for numerous applications that will allow the acquisition of skills, programs and behaviours automatically by observation in human–computer interfaces (e.g. Lieberman, 2001), human–robot interaction important in service robotics and other applications where robot assistants or companions need to learn from humans, and industrial applications such as automated factory floors in which new robots can acquire skills by observing the behaviour of other robots or humans. Models from psychology and biology are being validated and extended.
as scientists from these fields interact with collaborators from sciences of the artificial, while the latter benefit from the insight of their colleagues in the natural and social sciences in the harnessing of social learning in constructed systems.

Increasingly, it has been recognized that such a constructive approach towards imitation and social learning via the synthesis of artificial agents (in software or robots) can (a) yield important insights into mechanisms that can inform biologists and psychologists by fleshing out theory, as well as (b) help in the creation of artifacts that can be instructed and taught by imitation, demonstration and social interaction rather than by explicit programming.

2 Models and mechanisms: a constructivist viewpoint

Rather than develop approaches to imitation and social learning that are discipline-specific, we seek to understand the possible mechanisms that could generate them, independent of whether they may be realized in robots, humans or other animals. Models and proposed mechanisms that result in matching behaviour and social learning are thus the major focus of this book. By unifying the discussion, deep and sometimes abstract theories from psychology, ethology and neuroscience take their place in and are viewed from a constructivist perspective.

The constructive perspective must be adopted explicitly in the work of authors from robotics and computer science, while psychologists and experts in the behaviour and biology of animals are primarily concerned with whether the particular mechanisms and models they propose reflect the reality within the organisms they study. Neuroscience, by examining the neural basis realizing particular capacities, bridges the two.

There are benefits to be reaped by extending the constructive viewpoint to all aspects within the field of social learning and imitation, and many more from the resulting questions that generally require an interdisciplinary approach for answers. Immediately a constructive approach leads one to ask:

Could an explanation proposed in a theory of imitation actually be validated by building an artifact that exhibits the behaviour the theory is supposed to address? Is the theory or model explicit and detailed enough for one to build artificial systems as instances embodying it? Are there multiple, non-equivalent ways to build realizations of the mechanisms? Are there any gaps in the theory revealed by attempts at the explicit construction of models?

On the other hand, does a given constructive model constitute a full realization of a proposed explanation, or is it limited in scope, validating...
Introduction

only a portion of the theory? Is the constructed artifact realizing a mechanism thought to occur in animals, or does it instead provide an alternative method to realize a given social learning or imitative phenomenon that may be used in applications of artificial intelligence in engineering and technology for their own sake? Is the construction scalable, and to what extent is this limited essentially or only due to the limitations of the current materials available (generally silicon and steel, and unlike biological brains, currently running at best only with limited concurrency of components)? Are the predictions of theory confirmed in a constructed model? Does the model lead to predictions about the behaviour of organisms?

3 The book

In psychology, biology, ethology and other areas, the study of social learning and imitation in humans and other animals is a very lively and growing area of research which has resulted over the past few years in a number of important books, two of which are published by Cambridge University Press within the series Cambridge Studies in Cognitive and Perceptual Development:

• Andrew N. Meltzoff and Wolfgang Prinz, eds., The Imitative Mind, Cambridge University Press, 2002.
• Susan Hurley and Nick Chater, eds., Perspectives on Imitation: From Neuroscience to Social Science (2 volumes), MIT Press, 2005.

In addition, a number of relevant special journal issues have been published, e.g.:

• B. G. Galef, Jr and C. M. Heyes, guest editors, Social learning and imitation, special issue of Learning and Behavior, 32(1), 2004.

Imitation in Animals and Artifacts (Dautenhahn and Nehaniv, 2002) was the first book to bring together research in biology and psychology
with research in computer science and robotics on the particular topics of imitation and social learning, whereby many chapters discuss definitions of imitation and social learning, and study which species of animals imitate and in what way. The robotics/artificial intelligence chapters in this previous publication, although mostly biologically inspired, are similarly focused on particular research questions relevant to the domain of robotics and artificial intelligence. While this had helped set the groundwork for an interdisciplinary dialogue, the present volume, *Imitation and Social Learning in Robots, Humans and Animals: Behavioural, Social, and Communicative Dimensions*, reaches beyond single disciplines and reflects the emergence of numerous genuine cross-disciplinary collaborations between psychologists and biologists working together with roboticists and computer scientists. This new development of truly interdisciplinary work marks an important next step towards the advancement of the research field of imitation and social learning.

Unlike most books in the area of social learning and imitation, this particular volume thus emphasizes the interdisciplinary perspective on imitation in the context of human, animal and robotic behaviour. Many chapters are co-authored by experts pioneering collaborations across traditional disciplinary boundaries.1 Several chapters are co-authored by researchers from different disciplines who have brought cross-disciplinary expertise to bear on models and mechanisms of social learning.2 This serves both to consolidate and to provide a valuable reference for the increasing number of researchers entering this interdisciplinary field.

4 **Organization and themes**

Next we overview the major themes, organization and contents of the book. Chapters are grouped into a number (in the cognitively manageable range 7±2) of overarching and overlapping thematic parts, numbered by Roman numerals (I–VIII), while chapters are numbered continuously (from 1 to 21, excluding this introduction).

1 All book chapters have been peer reviewed anonymously.

2 The first forum that deliberately brought together participants from the natural sciences as well as the sciences of the artificial was the First International Symposium on Imitation in Animals and Artifacts, organized in 1999 by Professors Kerstin Dautenhahn and Chrystopher Nehaniv as part of the AISB’99 convention in Edinburgh, Scotland. The Second International Symposium on Imitation in Animals and Artifacts organized by Dautenhahn and Nehaniv was held in April 2003 in Aberystwyth as part of AISB’03. In 2005, as part of AIBS’05, the Third International Symposium on Imitation in Animals and Artifacts was held in April at University of Hertfordshire, Hatfield, organized by Dr Yiannis Demiris and the editors.
4.1 Thematic sections

The book is structured according to themes, reflected in the different book sections.

I Correspondence problems and mechanisms
II Mirroring and ‘mind-reading’
III What to imitate
IV Development and embodiment
V Synchrony and turn-taking as communicative mechanisms
VI Why imitate? – Motivations
VII Social feedback
VIII Ecological context

Note, these themes are multidisciplinary, so that the work on robots, humans and animals is represented across different themes. Each theme is introduced at the beginning of the corresponding section by a short text written by the editors, setting the stage for the chapters belonging to that theme.

Within each theme, we have sought to bring together the perspectives of authors from radically different areas of scientific inquiry (in some cases within a single chapter).

Across the eight themes of the book, several meta-groupings can be identified, cross-linking themes and chapters. In Parts I, III and IV we find chapters addressing the embodied nature of behavioural mapping between observer and learner. Parts II, IV, V and VII investigate neurobiological and developmental foundations of imitation and understanding others’ minds. Social and communicative perspectives of imitation and social learning are highlighted in sections V, VI and VII. Parts VI and VIII address motivational and evolutionary aspects of social learning and imitation.

Note, significant progress has been made over the past ten years in fields of research investigating social learning and imitation in humans, animals and artifacts. However, in none of the key themes as identified in the sections of this book are we close to a complete understanding of the issues involved. Thus, imitation and social learning will pose exciting challenges for many years to come, it will continue to challenge our understanding of the psychology and biology of important aspects of animal (social) minds and behaviour, and it will remain a demanding ‘benchmark challenge’ for artificial intelligence and robotics researchers concerned with socially intelligent artifacts. Cross-disciplinary approaches whereby robots serve as models and tools for the investigation of social learning and imitation in humans and animals could potentially help to synthesize ideas from different fields, that may ultimately lead to unified frameworks.
6 Nehaniv and Dautenhahn

4.2 Chapters

To allow readers to quickly acquire an overview of the organization of the book, the contents of the chapters are summarized below based on abstracts or introductions provided to the editors by the authors of the individual chapters. A short editorial essay at the beginning of each section of the book introduces the respective theme and chapters in more detail.

I Correspondence problems and mechanisms

1 Imitation: thoughts about theories
Geoffrey Bird and Cecilia Heyes

How does perception of an action enable the perceiver to produce a matching movement? This chapter examines three functional theories addressing this ‘correspondence problem’: active intermodal matching, associative sequence learning and ideomotor theories. Bird and Heyes first review behavioural and neurological evidence which is consistent with all three models. These experiments indicate that imitation involves the combination of perceptual and motor representations of action and that, once combined, these representations will support learning of new motor skills, in addition to activation of pre-existing motor patterns. Attention is then turned to evidence bearing on three foci of disagreement between the theories: effector-dependence of imitation, the extent to which observers are aware of information acquired through imitation and the role of experience in the formation of an imitative ability. In the light of this evidence, the authors identify current strengths and weaknesses of the three models, and discuss priorities for future research addressing the correspondence problem.

2 Nine billion correspondence problems
Chrystopher L. Nehaniv

The character and quality of matching behaviour depend crucially on the granularity and measure of similarity. The author outlines a taxonomy of social learning, imitative and matching behaviour according to these dimensions. This yields 24 classes of correspondence problems, in which the particular specialization of the model and ‘imitator’ embodiment and of the metric yields myriad particular correspondence problems, problems of learning how to imitate. Nehaniv describes how traditional and more recent taxonomies of social learning and behaviour matching are refined and clarified by this system, applicable to both animals (including humans) and constructed agents such as robots. New classes of matched behaviour are also distinguished and illustrated.

Attempts at matched behaviour according to a selected granularity and harnessing reinforcement using the chosen metrics gives rise to generic,
schematic mechanisms for these myriad correspondence problems in model systems. By considering measures of salience, it may be possible to adaptively and dynamically choose between possible correspondence problems, in automatic systems that learn what to imitate.

3 Challenges and issues faced in building a framework for conducting research in learning from observation

Darrin Bentivegna, Christopher Atkeson and Gordon Cheng

This chapter presents a framework that allows an agent to learn from observing a task. The authors describe a memory-based approach to learning how to select and provide sub-goals for behavioral primitives, given that an existing library of primitives is available. The framework, algorithms and test environments are used to highlight the challenges that are faced when building robotic systems that will learn from observation and practice. The details of implementing this framework on robots that operate in an air hockey and a marble maze environment are presented. Challenges involved with using observation data and primitives to increase the learning rate of agents are discussed.

II Mirroring and ‘mind-reading’

4 A neural architecture for imitation and intentional relations

Marco Iacoboni, Jonas Kaplan and Stephen Wilson

This chapter discusses the neural architecture that allows imitative behaviour and intentional relations. The discussion is at three levels: neural, computational and theoretical. The architecture comprises superior temporal cortex, the rostral part of the posterior parietal cortex and inferior frontal cortex. These three brain areas are mostly concerned with coding actions of self and others. The neural properties of these areas are mapped onto a computational architecture of paired forward and inverse internal models. From a theoretical perspective, this architecture permits a common framework for third-person knowledge (i.e. the observation of actions of others) and first-person knowledge (i.e. internal motor plans), thus facilitating social understanding. Within this framework, links between imitation and other cognitive domains are discussed.

5 Simulation theory of understanding others: a robotics perspective

Yiannis Demiris and Matthew Johnson

When we observe the behaviour of agents in our social environment, we are usually able to understand their actions, the reasoning behind them, and the goals of these actions. This is known as ‘having a theory of mind’,
i.e. a ‘mind-reading’ ability, which has evolved to allow us to be able to efficiently understand, and as a result, collaborate or compete with others.

There is intense debate with respect to the nature of this ability (Carruthers and Smith, 1996). Although there are intermediate positions between them, the two opposing theories that have been advocated are the ‘simulation theory’ and the ‘theory theory’ of mind reading (Gallese and Goldman, 1998). In the former, we understand the behaviour of others by simulating it using our own cognitive mechanisms underlying behaviour generation, while in the latter, we understand others’ behaviours by using an acquired (or for some researchers, innate) theory about behaviours.

This chapter focuses on the simulation theory of mind, starting by briefly examining related biological data, including neurophysiological experiments with primates, and human brain imaging and psychophysical data. The authors subsequently proceed to review their research designing and implementing computational architectures that embody principles of the simulation theory. Experiments are described that implement this theory onto robots, that observe and imitate simple actions by humans, before outlining important remaining challenges.

6 Mirrors and matchings: imitation from the perspective of mirror-self-recognition, and the parietal region’s involvement in both

Robert W. Mitchell

Given that mirror-self-recognition and generalized bodily imitation both require recognition of matching between bodies, it is likely that explanations for mirror-self-recognition would be useful for understanding imitation. In this chapter Mitchell presents two models he developed to explain self-recognition (Mitchell, 1993) which should provide a glimpse at imitation from different vantage points. The models should be of interest to researchers attempting to distinguish potential components of imitation. The first model focuses attention on self-recognition (and by extension generalized bodily imitation) as a form of intermodal matching. In this first model, derived from Guillaume’s (1925) analysis, self-recognition occurs for the first time because the organism: (1) understands mirror-correspondence; (2) has kinesthetic-visual matching; and (3) can make inferences, such as that the image is ‘the same’ as its own body. The ability to match between kinesthetic and visual modalities seems essential for generalized bodily imitation as well as self-recognition. In the second model, self-recognition occurs for the first time because the organism: (1) understands mirror-correspondence (as in the first model); (2) objectifies body parts; and (3) understands (sixth-stage) object permanence,
Introduction

such that it recognizes that parts of things are indicative of the whole and has low-level deductive abilities. From these abilities, the organism recognizes (via understanding mirror-correspondence) that the hand (e.g.) in the mirror is a contingent accurate image of its hand, and then recognizes (via its understanding of object permanence) that what is continuous with the hand in the mirror (i.e. the body-image) must be a contingent accurate image of what is continuous with its own hand (i.e. its body). This same sort of process could lead an organism to recognize similarities between its own actions and those of another, an essential component of generalized bodily imitation. The chapter examines the implications of these models for understanding generalized bodily imitation.

III What to Imitate

7 The question of ‘what to imitate’: inferring goals and intentions from demonstrations

Malinda Carpenter and Josep Call

A difficult issue for robotics researchers is the question of what to imitate, that is, which aspects of a demonstration robots should copy. Sometimes it is appropriate to copy others’ actions, sometimes it is appropriate to copy others’ results and sometimes copying both or even neither of these is the most appropriate response. The chapter discusses the advantages of using an understanding of others’ goals and intentions to answer this question, copying what the demonstrator intended to do rather than what he actually did. Whereas some animals focus mainly on demonstrators’ results or actions, one-year-old human infants appear to use an understanding of others’ goals to decide what to imitate. The authors identify some specific ways in which infants can infer the goal of a demonstrator in imitation situations, using such information as the demonstrators’ gaze direction, emotional expressions, actions and the context. This is followed by a brief review of what robots currently can do in this regard, proposing some further challenges for them.

8 Learning of gestures by imitation in a humanoid robot

Sylvain Calinon and Aude Billard

In this chapter, Calinon and Billard explore the issue of encoding, recognizing, generalizing and reproducing gestures. They address one major and generic issue, namely how to discover the essence of a gesture, i.e. how to find a representation of the data that encapsulates only the key aspects of the gesture, and discards the intrinsic variability across people’s motions. The model is tested and validated in a humanoid robot, using kinematics data of human motion. In order for the robot to learn new skills by
imitation, it must be endowed with the ability to generalize over multiple demonstrations. To achieve this, the robot must encode multivariate time-dependent data in an efficient way. Principal component analysis and hidden Markov models are used to reduce the dimensionality of the dataset and to extract the primitives of the motion.

The model takes inspiration from a recent trend of research that aims at defining a formal mathematical framework for imitation learning. In particular, it stresses the fact that the observed elements of a demonstration, and the organization of these elements should be stochastically described to have a robust robotic application. It bears similarities with theoretical models of animal imitation, and offers at the same time a probabilistic description of the data, more suitable for a real-world application.

9 The dynamic emergence of categories through imitation
Tony Belpaeme, Bart de Boer and Bart Jansen
The authors propose an extension of the study of imitation in artifacts in which imitation takes place at the population level. Instead of assuming a one-to-one teacher–student relation between agents, agents in their set-up do not take a pre-assigned role and have to build up a repertoire of actions through imitative interaction with all other agents in the population. The action–observation categories used by the agents emerge in the population via a process of self-organization in the course of iterated ‘imitation games’, in which agents take turns controlling a physical robot arm and observing another agent using it. Successful or unsuccessful recognition of category membership of observed actions drive the autonomous formation and revision of the individual agents’ categories without direct communication. Thus interaction via imitation games drives the self-organization of categories in a population, where social learning serves a mechanism for ‘cultural’ transmission.

IV Development and embodiment
10 Copying strategies by people with autistic spectrum disorder: why only imitation leads to social cognitive development
Justin H. G. Williams
Imitation seems likely to be an important process in the development of social cognition and theory of mind. Imitation and mental state communication both depend upon the development of skills through modification of an existing action repertoire, as well as an ability to represent such modifications as resulting from the influences of inner mental states. Therefore, impairment in forming representations of the cognitions