The Handbook on Socially Interactive Agents

20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics

Volume 2: Interactivity, Platforms, Application

Birgit Lugrin,
Catherine Pelachaud,
David Traum,
(Editors)

ASSOCIATION FOR COMPUTING MACHINERY

The Handbook on
Socially Interactive Agents

ACM Books

ACM Books is a new series of high-quality books for the computer science community, published by ACM in collaboration with Morgan & Claypool Publishers. ACM Books publications are widely distributed in both print and digital formats through booksellers and to libraries (and library consortia) and individual ACM members via the ACM Digital Library platform.

Probabilistic and Causal Inference: The Works of Judea Pearl
Hector Geffner, *ICREA and Universitat Pompeu Fabra*
Rina Dechter, *University of California, Irvine*
Joseph Y. Halpern, *Cornell University*
2021

Event Mining for Explanatory Modeling
Laleh Jalali, *University of California, Irvine (UCI), Hitachi America Ltd.*
Ramesh Jain, *University of California, Irvine (UCI)*
2021

Intelligent Computing for Interactive System Design: Statistics, Digital Signal Processing, and Machine Learning in Practice
Parisa Eslambolchilar, *Cardiff University, Wales, UK*
Andreas Komninos, *University of Patras, Greece*
Mark Dunlop, *Strathclyde University, Scotland, UK*
2021

Semantic Web for the Working Ontologist: Effective Modeling for Linked Data, RDFS, and OWL, Third Edition
Dean Allemang, *Working Ontologist LLC*
Jim Hendler, *Rensselaer Polytechnic Institute*
Fabien Gandon, *INRIA*
2020

Code Nation: Personal Computing and the Learn to Program Movement in America
Michael J. Halvorson, *Pacific Lutheran University*
2020

Computing and the National Science Foundation, 1950–2016: Building a Foundation for Modern Computing
Peter A. Freeman, *Georgia Institute of Technology*
W. Richards Adrion, *University of Massachusetts Amherst*
William Aspray, *University of Colorado Boulder*
2019

Providing Sound Foundations for Cryptography: On the work of Shafi Goldwasser and Silvio Micali
Oded Goldreich, *Weizmann Institute of Science*
2019

Concurrency: The Works of Leslie Lamport
Dahlia Malkhi, *VMware Research* and *Calibra*
2019

The Essentials of Modern Software Engineering: Free the Practices from the Method Prisons!
Ivar Jacobson, *Ivar Jacobson International*
Harold "Bud" Lawson, *Lawson Konsult AB (deceased)*
Pan-Wei Ng, *DBS Singapore*
Paul E. McMahon, *PEM Systems*
Michael Goedicke, *Universität Duisburg–Essen*
2019

Data Cleaning
Ihab F. Ilyas, *University of Waterloo*
Xu Chu, *Georgia Institute of Technology*
2019

Conversational UX Design: A Practitioner's Guide to the Natural Conversation Framework
Robert J. Moore, *IBM Research–Almaden*
Raphael Arar, *IBM Research–Almaden*
2019

Heterogeneous Computing: Hardware and Software Perspectives
Mohamed Zahran, *New York University*
2019

Hardness of Approximation Between P and NP
Aviad Rubinstein, *Stanford University*
2019

The Handbook of Multimodal-Multisensor Interfaces, Volume 3: Language Processing, Software, Commercialization, and Emerging Directions
Editors: Sharon Oviatt, *Monash University*
Björn Schuller, *Imperial College London and University of Augsburg*
Philip R. Cohen, *Monash University*

Daniel Sonntag, *German Research Center for Artificial Intelligence (DFKI)*
Gerasimos Potamianos, *University of Thessaly*
Antonio Kruğer, *Saarland University and German Research Center for Artificial Intelligence (DFKI)*
2019

Making Databases Work: The Pragmatic Wisdom of Michael Stonebraker
Editor: Michael L. Brodie, *Massachusetts Institute of Technology*
2018

The Handbook of Multimodal-Multisensor Interfaces, Volume 2: Signal Processing, Architectures, and Detection of Emotion and Cognition
Editors: Sharon Oviatt, *Monash University*
Björn Schuller, *University of Augsburg and Imperial College London*
Philip R. Cohen, *Monash University*
Daniel Sonntag, *German Research Center for Artificial Intelligence (DFKI)*
Gerasimos Potamianos, *University of Thessaly*
Antonio Kruğer, *Saarland University and German Research Center for Artificial Intelligence (DFKI)*
2018

Declarative Logic Programming: Theory, Systems, and Applications
Editors: Michael Kifer, *Stony Brook University*
Yanhong Annie Liu, *Stony Brook University*
2018

The Sparse Fourier Transform: Theory and Practice
Haitham Hassanieh, *University of Illinois at Urbana-Champaign*
2018

The Continuing Arms Race: Code-Reuse Attacks and Defenses
Editors: Per Larsen, *Immunant, Inc.*
Ahmad-Reza Sadeghi, *Technische Universität Darmstadt*
2018

Frontiers of Multimedia Research
Editor: Shih-Fu Chang, *Columbia University*
2018

Shared-Memory Parallelism Can Be Simple, Fast, and Scalable
Julian Shun, *University of California, Berkeley*
2017

Computational Prediction of Protein Complexes from Protein Interaction Networks
Sriganesh Srihari, *The University of Queensland Institute for Molecular Bioscience*
Chern Han Yong, *Duke-National University of Singapore Medical School*
Limsoon Wong, *National University of Singapore*
2017

The Handbook of Multimodal-Multisensor Interfaces, Volume 1: Foundations, User Modeling, and Common Modality Combinations
Editors: Sharon Oviatt, *Incaa Designs*
Björn Schuller, *University of Passau and Imperial College London*
Philip R. Cohen, *Voicebox Technologies*
Daniel Sonntag, *German Research Center for Artificial Intelligence (DFKI)*
Gerasimos Potamianos, *University of Thessaly*
Antonio Krüger, *Saarland University and German Research Center for Artificial Intelligence (DFKI)*
2017

Communities of Computing: Computer Science and Society in the ACM
Thomas J. Misa, Editor, *University of Minnesota*
2017

Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining
ChengXiang Zhai, *University of Illinois at Urbana–Champaign*
Sean Massung, *University of Illinois at Urbana–Champaign*
2016

An Architecture for Fast and General Data Processing on Large Clusters
Matei Zaharia, *Stanford University*
2016

Reactive Internet Programming: State Chart XML in Action
Franck Barbier, *University of Pau, France*
2016

Verified Functional Programming in Agda
Aaron Stump, *The University of Iowa*
2016

The VR Book: Human-Centered Design for Virtual Reality
Jason Jerald, *NextGen Interactions*
2016

Ada's Legacy: Cultures of Computing from the Victorian to the Digital Age
Robin Hammerman, *Stevens Institute of Technology*
Andrew L. Russell, *Stevens Institute of Technology*
2016

Edmund Berkeley and the Social Responsibility of Computer Professionals
Bernadette Longo, *New Jersey Institute of Technology*
2015

Candidate Multilinear Maps
Sanjam Garg, *University of California, Berkeley*
2015

Smarter Than Their Machines: Oral Histories of Pioneers in Interactive Computing
John Cullinane, *Northeastern University; Mossavar-Rahmani Center for Business and Government, John F. Kennedy School of Government, Harvard University*
2015

A Framework for Scientific Discovery through Video Games
Seth Cooper, *University of Washington*
2014

Trust Extension as a Mechanism for Secure Code Execution on Commodity Computers
Bryan Jeffrey Parno, *Microsoft Research*
2014

Embracing Interference in Wireless Systems
Shyamnath Gollakota, *University of Washington*
2014

The Handbook on Socially Interactive Agents

20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition

Birgit Lugrin

Julius-Maximilians-Universität of Würzburg

Catherine Pelachaud

CNRS-ISIR, Sorbonne Université

David Traum

University of Southern California

ACM Books #37

The Handbook on Socially Interactive Agents
Birgit Lugrin, Catherine Pelachaud, and David Traum, Editors

books.acm.org
http://books.acm.org

ISBN: 978-1-4503-8720-0 hardcover
ISBN: 978-1-4503-8721-7 paperback
ISBN: 978-1-4503-8722-4 EPUB
ISBN: 978-1-4503-8723-1 eBook

Series ISSN: 2374-6769 print 2374-6777 electronic

DOIs:

10.1145/3477322 Book
10.1145/3477322.3477323 Foreword
10.1145/3477322.3477324 Chapter 1
10.1145/3477322.3477325 Chapter 2
10.1145/3477322.3477326 Chapter 3
10.1145/3477322.3477327 Chapter 4
10.1145/3477322.3477328 Chapter 5
10.1145/3477322.3477329 Chapter 6
10.1145/3477322.3477330 Chapter 7
10.1145/3477322.3477331 Chapter 8
10.1145/3477322.3477332 Chapter 9
10.1145/3477322.3477333 Chapter 10
10.1145/3477322.3477334 Chapter 11
10.1145/3477322.3477335 Chapter 12
10.1145/3477322.3477336 Chapter 13
10.1145/3477322.3477337 Bios/Index

A publication in the ACM Books series, #37

This book was typeset in Arnhem Pro 10/14 and Flama using pdfTEX.
Front cover design by Birgit Lugrin and Sophia C. Steinhaeusser (copyright remains with the designers).

First Edition

10 9 8 7 6 5 4 3 2 1

Contents

Foreword

In preparation for writing this foreword, I looked through old emails (*really* old emails) dating back to early 1998, when we were planning the "First Workshop on Embodied Conversational Characters." In and amongst detailed menu planning for the workshop (I haven't changed a bit since those days) are emails floating the idea of publishing a book with the best papers from the workshop. We were already starting to see a shift in the literature, away from "lifelike computer characters" (Microsoft's Clippy was presented at a workshop with that name) and "believable computer characters" (characters whose behavior was believable, but that did not *do* anything for people), and we wanted the book to reflect that shift.

We particularly wanted to highlight the fact that embodied conversational characters did not only talk but also listened. They were capable of understanding as well as generating language and non-verbal behavior, and they did so in the service of humans—they were *agents*, like travel agents or real estate agents. To that end, I sent an email to the chapter authors with the following tidbits. I wrote:

> Next, a note about terminology. After long debate, we've decided to call the book Embodied Conversational ***Agents***, and not ***Characters*** (for marketing reasons, in part) so you might want to follow this terminology in your chapter.
>
> Finally, do make sure to focus on the ***communicative*** abilities of your systems, since this is what distinguishes this work—and this book—from previous volumes on believable characters, software agents and so forth.

It's amusing to read this today when we take for granted the *agentive* nature of our conversational systems. At this point, we assume that embodied conversation agents (ECAs) are designed primarily to accomplish work for people. We also take for granted that ECAs must both understand *and* talk. However, when the *Embodied Conversational Agents* book was conceived, both of those features were only newly possible. In turn, the title of the current volume highlights the most recent technological innovation, which is the ability of the systems not just to do work for humans but to interact socially with them in the process, in many cases using social interaction as a way to bootstrap task performance.

It's illuminating to look at two other debates that took place during this same period. The first concerns what kinds of data are used to create the most natural behaviors for an ECA. The second concerns whether it is ethical to build natural-acting ECAs.

While there was beginning to be consensus in the late 1990s on the idea that conversational characters could do more than just look pretty, there were three schools of thought about the proper inspiration for the conversational behaviors of ECAs (as they were called). Some of the authors in the original volume worked with actors to understand what kinds of language and non-verbal behaviors were most evocative of normal human conversation. These researchers hewed to the belief that ECAs should behave in a somewhat exaggerated fashion, like actors on a stage, in order to seem natural to their human interlocutors. Other authors believed that, being native speakers of their own language, and acculturated to the customs of their own society, the simple intuitions of the researcher were sufficient to design human-like conversational behaviors. A third group believed that psychological and linguistic studies of human conversation were the only proper inspiration for the behaviors of ECAs. Today, while a few researchers still work with actors or rely on their own intuitions, the community of researchers in ECAs (and in today's socially interactive agents) mostly rely on empirical psychological and linguistic studies of human behavior as their inspiration. Some of these researchers carry out their own studies, and some rely on extant literature, but in both cases they rely on normal everyday humans for inspiration rather than actors or computer scientists. The debate is interesting in the face of today's focus on big data. In fact, the increasing reliance in the field of artificial intelligence (AI) on machine learning techniques to analyze human behavior has led to a parallel increase in ECA systems that rely on deep learning techniques applied to large corpora, often of naturally produced human conversational behavior, to generate appropriate verbal and non-verbal conversational behavior. In other words, AI has brought us closer to the human-inspired ECAs of the past by bringing a focus on corpora of natural behaviors. At the same time, however, it has taken us further away from those human-inspired ECAs of the past because the corpora are too large to be examined by the human eye.

Another debate that evoked heated interchanges in the late 1990s, and that is useful to contemplate today, was whether we *should* even contemplate deploying ECAs as interfaces to computational systems in the first place. Many if not most of the authors in the 1998 volume believed that ECAs represented a more natural way of interacting with computational systems than a keyboard and a mouse.

Their work was predicated on the assumption that interacting with a human-like agent was a more intuitive manner of accessing technical systems. To other computer scientists of the era, however, ECAs were downright evil. Perhaps most famously, human–computer interaction researcher Ben Schneiderman saved his strongest invectives for human-like agents and their designers. In 1995, he wrote

> Anthropomorphic terms and concepts have continually been rejected by consumers, yet some designers fail to learn the lesson. Talking cash registers and cars, SmartPhone, SmartHome, Postal Buddy, Intelligent Dishwasher, and variations have all come and gone. Even the recent variation of Personal Digital Assistant had to give way to the more service oriented name now used in the Apple Newton ads: MessagePad. We'll leave it to the psychoanalysts to fathom why some designers persist in applying human attributes to their creations … But, possibly, just possibly, all this heated debate is excessive and agents will merely become the Pet Rock of the 1990s—everyone knows they're just for fun (Ben Shneiderman 1995. *ACM Interactions*. 2, 1, 13–15).

Today, fears about whether robots will steal jobs, and whether machine learning will make it hard to tell who is human and who is an AI, have once again launched debates on whether human-like agents are a good or bad influence on society. These debates have led to a stronger focus on transparency in AI, a concern with bias in data, and a much-needed conversation on the ethics of where ECAs should and should not be used. These contemporary debates, however, and contra Shneiderman's predictions, show that anthropomorphic agents have stood the test of time. The topic has inspired passion and dedication in a whole new generation of researchers. To that end, here 20 years later is a two-volume follow-up from our 1998 *Embodied Conversational Agents* book, with more than 25 chapters, showing the depth, breadth, innovation, creativity, and, yes, effectiveness, of human-inspired agents.

Justine Cassell

1 Introduction to Socially Interactive Agents

Birgit Lugrin

Since the commercialization of graphical user interfaces in the late 1980s, the way humans interact with computers has been dominated by their interactions through windows, icons, menus, and pointers (WIMP) interfaces, with buttons that can be clicked and information that can be read or watched in separate windows. The research discipline of human–computer interaction (HCI) is constantly developing new and creative systems that go beyond this traditional interaction for a more intuitive usage, for example, with technology such as touch interaction, virtual reality, tangible computing, and many more.

Taking a different approach to realize natural and intuitive interaction, the research area of *socially interactive agents* (SIAs) aims to develop artificial agents that can interact via communication channels that come more natural to human interactants by equipping the interface with a body that interacts multi-modally by using verbal, para-verbal, and non-verbal behaviors. With it, communication styles that are known from human face-to-face interaction can be transferred to interaction with machines.

SIAs (see Figure 1.1 for examples) have been developed under different names in different research fields such as *intelligent virtual agents* (IVAs), *embodied conversational agents* (ECAs), or *social robotics* (SRs) (see below for definitions of the respective terms). More than 20 years of research and development in these fields have drastically advanced the state of the art. For this book, we chose to use the term *socially interactive agents* (or *SIAs*) as it includes both physical and virtual embodiments, while highlighting their ability for social interaction as well as the need to realize socially intelligent, autonomous behaviors.

We define SIAs as follows:

> *SIAs are virtually or physically embodied agents that are capable of autonomously communicating with people and each other in a socially intelligent manner using multi-modal behaviors.*

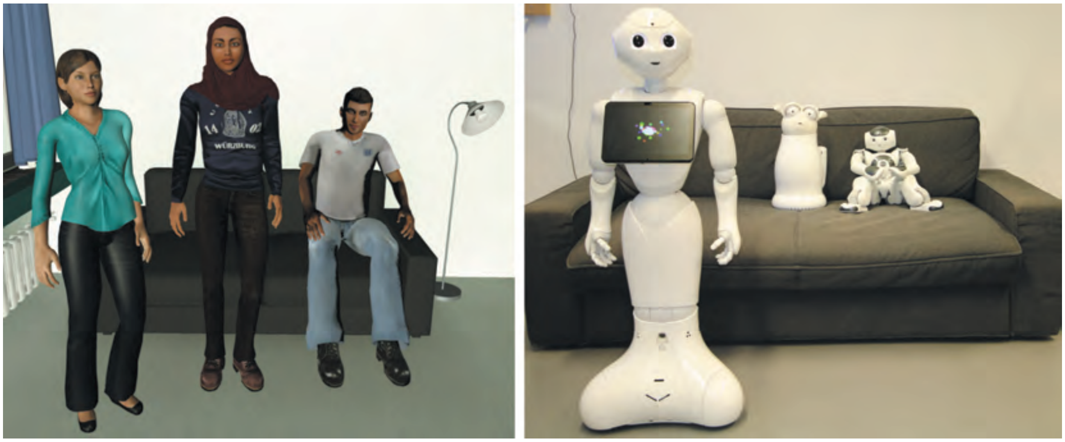

Figure 1.1 Examples of socially interactive agents (SIAs): intelligent virtual agents (left) and social robots (right). SIAs in both figures are located in the same virtual versus physical office space (reflected reality) [Eckstein et al. 2019], used for various research in the Media Informatics Laboratory of Wuerzburg University (left to right: two female agents and a male agent by Autodesk, partly adapted by features such as clothing style, Pepper by SoftBank Robotics, Reeti by Robopec, Nao by SoftBank Robotics).

In order to interact with humans in a socially intelligent manner, underlying concepts such as emotions, empathy, or how to behave in a group are essential for SIAs to interpret. To be part of the interaction, observed input must be reasoned about, and decisions to be taken upon that resemble a cognitive process. The SIA's (re)actions need to be externalized by natural language, expressive speech, and non-verbal behaviors.

1.1 Potential of SIAs

The right choice of interface is not a simple one. While traditional WIMP interaction is certainly fast and well established, it is still rather well suited for simple, repetitive tasks, for example, in office work. Many of today's challenging tasks have led to novel solutions, such as sophisticated three-dimensional (3D) interfaces to help visualize 3D problems. Analog, for scenarios, where social sensitivity and conversation are paramount, the natural communication with SIAs might be the best solution.

Thanks to extensive research, today prototypes including SIAs are used in many application domains that are helpful for individuals or society, with SIAs serving as companions or assistants in aging support, health education, life-long learning, or training of specific skills. In the long run, SIAs are envisioned to unobtrusively support humans in their daily lives. Figure 1.2 illustrates that vision by extending a

Figure 1.2 Vision of using SIAs in the future: progressing from having to adapt to interact with technology to a more natural communication with SIAs that assist people in their everyday lives (based on Zallinger [1965], humoristic extensions and silhouette versions of the Zallinger image have become known as "The March of Progress" that has achieved iconic status so that there are many different versions of the March of Progress used today).

well-known humoristic illustration in HCI, of how humans had to adapt for interaction with machines by adding a future perspective of how technology in the form of a SIA adapts to human-style interaction.

In some cases, using a SIA might even have advantages over a human communication partner. For example, in a tutoring scenario with a SIA an emotional distance can be kept, and a user might not feel embarrassment, for example, to admit that he or she cannot read. In addition, a training task can be repeated as often as preferred without the risk of annoying a human training partner, or having to pay for each additional lesson, providing individualized sessions to social groups that usually might not have access to private training. Also, there is a dichotomy between appearance and behavior for SIAs, allowing modification of background factors such as age, gender, personality, or ethnic background separately from the implemented role and communicative behavior. This can be useful in personalizing a SIA to provide the best possible solution for each user's specific requirements or preferences.

Besides the many useful applications SIAs are (envisioned to be) employed in, they can serve as a research paradigm. In perception studies SIAs can serve as stimulus material. With it, they can help learn more about humans, their judgements, preferences, or emotional reactions to artificially created, yet very standardized variations of social situations. In interaction studies SIAs can serve as communication partners, allowing for high control over the experiment, ensuring detailed consistent behavior over many sessions. That way, the social behavior of humans, and the effects of different behaviors they are confronted with, can be studied.

A concern or fear that many researchers in the research area of SIAs are confronted with is the conception that these agents might be developed to replace humans in the workplace or even in social relationships. It is very important to note here that the replacement of humans is not, and has never been, a goal in the development or research on SIAs. On the contrary, SIAs are developed to support humans and assist in situations where no human support can be provided or is not desired, and to offer additional functionalities or support in social domains. As they aim to provide a more human-like interface that is intuitive to understand and interact with, they might be replacing other devices that might appear complicated to certain users for certain tasks. We want to further highlight that, particularly since SIAs enter social domains, development has to follow interdisciplinary approaches and methods, and needs to include, besides the technical know-how, expertise in psychology, sociology, and ethics.

1.2 Terminology

Since research on SIAs is manifold and researchers are coming from different disciplines and research areas, a number of terms exist that can be found in the literature. In the following, we aim to shed light on the terminology (in alphabetical order), and highlight their origin and different foci, albeit you might find some of the definitions quite similar:

Agent: "An agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators" [Russell and Norvig 2009]. This very classic and well-known definition looks at agents from the perspective of artificial intelligence (AI), highlighting the autonomy of the artificial entities. These agents can be, but are not necessarily, embodied. Examples include softbots, thermostats, robots, or humans.

Avatar: An avatar represents a game unit that is under the player's control [Kromand 2007], which is usually the graphical representation of the user in the virtual environment [Trepte and Reinecke 2010]. Unfortunately, this term is often confused with virtual or robotic agents in communities other than SIAs. Note that an avatar is not behaving or interacting autonomously with a user but representing the user in the virtual or real world. The term embodiment also has a different meaning concerning avatars, and describes the physical process to substitute (parts of) a person's body with a virtual one by the deployment of virtual reality hardware and software [Spanlang et al. 2014].

Embodied conversational agent: "Embodied conversational agents are computer-generated cartoonlike characters that demonstrate many of the same properties as humans in face-to-face conversation, including the ability to produce and respond to verbal and nonverbal communication" [Cassell et al. 2000]. The term was defined by Cassell and colleagues in their same named book on the topic in

2000. The authors highlight the importance of the combination of the multi-modal interface, with a software agent and a dialog system, to assure natural conversation. While the original focus was on virtual embodiments, the term also allows robotic embodiments, and is used in both fields.

Intelligent virtual agent: "Intelligent virtual agents are interactive digital characters that exhibit human-like qualities and can communicate with humans and each other using natural human modalities like facial expressions, speech, and gesture. They are capable of real-time perception, cognition, emotion, and action that allow them to participate in dynamic social environments" [IVA 2019]. This term focuses on communicative, digital characters, and is mainly used by researchers that are affiliated with the IVA conference series. An important fact lies on the character's intelligence that allows them to dynamically interact, as opposed to scripted behavior.

Socially assistive robot: Socially assistive robots were defined by Feil-Seifer and Matarić [2005] as robots that share characteristics with assistive robots, in particular to provide assistance to users, but are distinguished by their focus on social interaction while assisting people.

Socially intelligent agent: "The field of socially intelligent agents is characterized by agent systems that show human style social intelligence" [Dautenhahn et al. 2002]. The term was coined by Dautenhahn in the late 1990s and highlights the specific social intelligence of the agent, relying on "deep models of human cognition and social competence" [Dautenhahn 1998] that needs to comprise strongly interdisciplinary approaches. Different embodiments of these agents are possible, virtual, or robotic.

Socially interactive robot: Socially interactive robots were defined as "robots for which social interaction plays a key role" [Fong et al. 2003] in order to "distinguish these robots from other robots that involve 'conventional' human–robot interaction, such as those used in teleoperation scenarios" [Fong et al. 2003]. This term was defined after the definition of socially intelligent agents, to highlight the need for social interaction.

Socially interactive agent: The term *socially interactive agent* extends the term socially interactive robot by allowing virtual and physical embodiments. This term was used by the AAMAS (autonomous agents and multi-agent systems) community and conference series, where they are described as "capable of interacting with people and each other using social communicative behaviors common to human–human interaction. Example applications include social assistants on mobile devices, pedagogical agents in tutoring systems, characters in interactive games, SRs collaborating with humans and multimodal interface agents for smart appliances and environments" [AAMAS 2019].

Social robot: "Social (or Sociable) robots are designed to interact with people in a natural, interpersonal manner [...] They will need to be able to communicate naturally with people using both verbal and non-verbal signals. They will need to engage us not only on a cognitive level, but on an emotional level as well in order to provide effective social and task-related support to people" [Breazeal et al. 2016]. Social robotics is distinguished from robotics through its socially interactive focus with applications in domains such as education, ageing support, or entertainment. This term is predominantly used by the social robotics community and the same named conference series and journal.

Virtual character: The term *virtual character* focuses on a virtual representation of a figure along with its animations. "Virtual characters in animated movies and games can be very expressive and have the ability to convey complex emotions" [McDonnell et al. 2008]. Note, that they do not necessarily have to be intelligent or interactive, cf. characters of a movie. Thus, the term is often used by researchers who focus on the character's appearance, graphics, animation, or background story.

Virtual human: "Virtual humans are artificial characters who look and act like humans but inhabit a simulated environment" [Traum 2008]. The term focuses on human-like appearance and behavior and is frequently used by American authors and research groups. Research on virtual humans often relies on highly realistic graphical representations of the characters and their animations.

Please note that the terms introduced above are the ones most commonly used. Other variations, for example, affective embodied agent, companion robot, conversational robot, relational agent, social embodied agent, socially intelligent robot, socially intelligent virtual agent, virtual agent, and so on, are also found in the literature and address similar research topics.

For the scope of this book, we use the term *socially interactive agents* (or *SIAs*) when we talk about both kinds of embodiment, virtual or robotic. We chose this term as we think it highlights the socially interactive nature as well as the intelligent background of the agent. We use the term *intelligent virtual agent* (or *IVA*) in instances where we discuss virtual representations of SIAs solely. We use the term *social robot* (or *SR*) when we discuss robotic representations of SIAs solely.

1.3 Origin and Embodiment

To understand why there is a large variety of terms, and why research in the fields of IVAs and SRs might seem distinct sometimes, it is informative to have a look at their different origins.

IVAs originated from the idea of simulating human communication with displays of human-like communication channels such as facial expressions or gestures, and became feasible in the 1990s due to advances in computer graphics. With constant advances in computer graphics and the integration of AI methods and cognitive modeling, the communicative abilities and social behaviors of IVAs have constantly been driven further.

SRs, on the other hand, originated from robotics. Since robots are leaving industrial applications and are entering private households, a closer interaction with the user in a private and social domain is imposed. Social behavior and skills that are acceptable for humans are becoming a requirement (referred to as a "robotiquette" by Dautenhahn [2007]). Thus, SRs are robots that are specifically designed to interact with people in an inter-personal manner, including the need to recognize and generate verbal and non-verbal signals, for example [Breazeal 2002].

In simple terms, one could say that in IVA research the virtual body was introduced to be able to simulate human behavior, while in SR research the physical body of the robot was naturally there and needed to be adapted for human-like behavior when interpersonal interaction was desired.

Despite their different origins, the fields of IVAs and SRs today follow the same goals and are employed in similar domains. To a certain extent, they also share common underlying technologies, such as text-to-speech systems, computer vision, or emotion detention. Also, the theoretical background from psychology or the social sciences are shared for the computational modeling of cognitive processes such as empathy. Particularly in these areas, research from IVAs and SRs can benefit greatly from one another.

Other aspects are not as easily transferable from one field to the other. The key difference is the environment that the SIA inhabits and with it whether or not they share the same physical space with their human interaction partners. Particularly when it comes to human perception of the SIA or the acceptance thereof, the different type of embodiment seems to play a key factor. Also, the translation from high-level behavior (e.g., show agreement) to the concrete execution with the particular body part (e.g., nod head and smile) differs across embodiments. Furthermore, hybrid versions are available today, where an SR contains a display on the head that shows a virtual face.

Either type of embodiment has certain advantages that might be a disadvantage of the other but does not necessarily have to be. Some of the characteristics are listed below:

Characteristics of a virtual embodiment

- Appearance: The look of an IVA can be freely customized and adapted for different users, applications, or contexts.
- Animation: The virtual face and body can be animated very fine grained and in a very realistic manner and show a large variety of emotional expressiveness.
- Acceptance: IVAs are often described as non-threatening. A “safe setting” can be created through the separation of the environment that is inhabited by the human and the IVA.
- Duplication: Applications with IVAs can be duplicated easily and provided to many users.
- Easy Access to implementation: Since very good tools are available for free, students and practitioners who want to get acquainted with the research area of SIAs can have easy access to build their own IVA architecture, or application. The only requirement to get started is the availability of a computer.
- Easy Access to use: Applications with IVAs can be implemented for usage with traditional computers or mobile phones. This way the IVA can be deployed anywhere, anytime on people’s private devices.

Characteristics of a physical embodiment

- Appearance: With commercial SRs, the options to customize the appearance are limited (e.g., by adding stickers or accessories). However, today’s opportunities with 3D printers and single-board computers allow designing individual SRs at a rather low cost.
- Animation: The options to animate an SR is dependent on the particular model and its individual degrees of freedom (e.g., whether it has limbs or an animate-able face). Due to hardware limitations subtle emotional expressions might not be feasible.
- Acceptance: It has been widely reported that the physical presence of an SR has a positive effect on the perception of users, and in particular their feeling of social presence, for example [Breazeal et al. 2016].

- Mobility: The most dominant advantage of a robot's physical body lies in its ability to move around in the real world and conduct physical interaction with the environment. An SR can, for example, provide services such as serving food or beverages. However, the physical body also provides challenges, such as the risk of accidentally falling over.
- Physical interaction: In addition to conversational interaction, physical interaction with the human user is possible (e.g., by performing social touch). The shared space can additionally be used for conversational purposes (e.g., to gain someone's attention).

Despite the characteristics that are implied by the embodiment of a SIA, a number of studies have directly compared physical and virtual embodiments to evaluate the outcomes of similar interactions with users, see, for example, Deng et al. [2019] for an overview. Mainly in these comparisons, the virtual SIA is a direct transfer of an SR into a virtual representation of the same robot. It seems the physical embodiment of an SR outperforms a virtual one, both in task performance and the perception of the users. However, the results are more inconclusive if the concepts of physical presence and embodiment are separated, by either comparing physically present SIAs to virtually present SIAs, or comparing physical SIAs with virtual SIAs both presented on a screen [Li 2015].

While directly comparing virtual and physical representations of SRs is a valid research paradigm that allows comparing between embodiments and the impact of physical presence in very controlled settings, from a practical perspective the design of a "virtual social robot" would not be beneficial. With it, most advantages of an IVA are out-ruled, and the virtual representation is artificially bound to non-existing, virtual hardware limits. Advantages such as subtle animations, duplication, or potential usage on smartphones are neglected. To date, the number of studies that compare three or more representations, or compare a state-of-the-art SR against a state-of-the-art IVA are rare.

It also needs to be noted that moderating factors such as the interaction scenario and task, and the user's perception of the SIA's body-related capabilities, seem to play a crucial role in people's ratings of the SIAs [Hoffmann et al. 2018]. The right choice of embodiment of a SIA is thus highly complex and dependent on many factors such as the situational context, role of the SIA, purpose of the application, or user's preference.

1.4 Purpose of the Book

The fields of IVAs and SRs face similar research issues and challenges and are further developed in universities and research facilities across the world. Research on

IVAs and SRs can benefit greatly from one another and have contributed to each other's advancement in the past. However, substantial work in both research fields is sometimes overlooked by researchers in the other area. This is partly due to the fact that different wordings are used and there exists a large number of journals and conferences that publish works on SIAs, making it very difficult to maintain a good overview.

The interdisciplinary nature of SIA research also contributes to the very diverse venues where you can find relevant findings on SIAs. While researchers from the cognitive sciences bring expertise in underlying processes, communication, and interaction, computer scientists bring expertise in conceptualizing computational models and implementation. Even within a single discipline, approaches, methods, and wording can be used differently, thereby complicating cooperation. In computer science, for example, many areas are involved in SIA research, such as AI, HCI, robotics, computer graphics, or software engineering. Only through communication and research in interdisciplinary teams can the field be advanced. This constitutes one major challenge by itself, as researchers sometimes do not have enough insights into other areas (or even disciplines), and thus might not appreciate each other's work enough.

We hope that this handbook will help raise the visibility of the research in the fields involved and further close the gap between the IVA and SR communities. At the same time, we hope that in the future reinventing the wheel can be avoided. This comprehensive handbook on SIAs summarizes the research that has taken place over the last 20 years. We are referring to this period, since the first complete book on embodied conversational agents [Cassell et al. 2000, see above] appeared in 2000, although we are aware that research on this topic began earlier. By pointing out current challenges and future directions in the various topics involved, we hope to help directing future research and cooperation. In the book, we include views from an interdisciplinary perspective, containing theoretical backgrounds from human–human interaction, their implementation in computational models, their evaluation with human users, integration into applications, and ethical implications.

In a structured and easily accessible way, the book (hopefully) provides a valuable source of information on SIAs for research and education. Researchers in the research area of SIAs will find it a valuable overview of the field. Teaching staff will benefit from the handbook to structure courses for undergraduate or graduate students, and with it train the upcoming generations of young researchers.

Particularly now, public interest in SIAs is increasing. The book will also help professionals, and interested lay public readers, to get acquainted with this research area.

1.5 Structure of the Book

This handbook is divided into two volumes, including 28 chapters that are grouped in five major parts, to cover the major topics in the area. For the book, we have relied on our connections to both fields, IVAs and SRs, providing a collection of surveys, each written by (an) acknowledged international expert(s) of their field.

Each chapter provides a survey that summarizes the theoretical background, approaches for implementation, history/overview of the topic, alongside current challenges and future directions. All the chapters discuss similarities and differences between IVAs and SRs and highlight important work of both fields. Where applicable, the chapters will follow a common structure to ensure internal consistency and facilitate understanding.

1.5.1 Volume 1

After this first chapter that introduces readers to the handbook, Volume 1 starts with Part I "Establishing SIA Research" that helps understand how research in this area is conducted and discusses the impact thereof on individuals and society.

Chapter 2 "Empirical Methods in the Social Science for Researching Socially Interactive Agents," by Astrid Rosenthal-von der Pütten and Anna M. H. Abrams, introduces the empirical methodology from the social sciences that is necessary for SIA research, particularly when it comes to research experiments including human participants.

Chapter 3, "Social Reactions to Socially Interactive Agents and Their Ethical Implications," by Nicole Krämer and Arne Manzeschke, looks at SIA research from a psychological and ethical perspective. It points to numerous studies demonstrating that people (unconsciously) react socially toward artificial entities, and that as soon as they display social cues people can also be manipulated or socially influenced.

Part II "Appearance and Behavior" deals with the impact of the looks of SIAs and the various aspects of multi-modal behavior that need to be taken into account when convincing SIAs behavior is modeled.

Chapter 4 "Appearance," by Rachel McDonnell and Bilge Mutlu, argues that, compared to voice assistants, embodied agents enable the use of appearance-based cues from human–human interaction, such as mutual gaze, that are known to improve social outcomes. The chapter shows that the appearance of a SIA can affect how people perceive, respond to, and interact with it.

Chapter 5 "Natural Language Understanding in Socially Interactive Agents," by Roberto Pieraccini, introduces natural language understanding as an essential part of any interactive agent and highlights its complexity, particularly for SIAs that need to react to user-initiated interactions across various application areas.

Chapter 6 “Building and Designing Expressive Speech Synthesis,” by Matthew Aylett, Leigh Clark, Benjamin R. Cowan and Ilaria Torre, gives an overview of definitions, methods, and state-of-the art in expressive voices, and critically discusses when and where expressive speech is beneficial.

Chapter 7 “Gesture Generation,” by Carolyn Saund and Stacy Marsella, discusses the complexity of communicative gestures and how they enhance communication in human–human conversation, and summarizes the research and their challenges in the transfer of this complexity in the implementation with SIAs.

Chapter 8 “Multi-modal Behavior Modeling for Socially Interactive Agents,” by Catherine Pelachaud, Carlos Busso, and Dirk Heylen, extends the theme of nonverbal behavior by adding additional modalities such as gaze, smiles, or social touch. Starting from introducing concepts from the social sciences, the chapter has a strong focus on the different computational models that can be employed for the implementation of multi-modal behaviors.

Part III “Social Cognition—Models and Phenomena” investigates internal processes known from human cognition that are driving forces in human–human interaction, and demonstrates how they are addressed in SIA systems.

Chapter 9 “Theory of Mind and Joint Attention,” by Jairo Perez-Osorio, Eva Wiese, and Agnieszka Wykowska, introduces the two crucial mechanisms of social cognition, and explains how they apply to the interaction between humans and SIAs from two angles: evoking human social cognition and modeling artificial social cognition.

Chapter 10 “Emotion,” by Joost Broekens, focuses on the computational representation of emotion and other related affective concepts such as mood, attitude, or appraisal, and highlights how SIAs can make constructive use of them.

Chapter 11 “Empathy and Prosociality in Social Agents,” by Ana Paiva, Filipa Correia, Raquel Oliveira, Fernando Santos, and Patrícia Arriaga, focuses on empathy and in particular on the related concept of prosociality (conducting positive and voluntary behavior that should benefit others). With it, the authors provide a framework including the main variables needed to design prosocial agents, for individual or dyadic interactions, or at the society level.

Chapter 12 “Rapport Between Humans and Socially Interactive Agents,” by Jonathan Gratch and Gale Lucas, introduces rapport (a fine-grained emotional communicational interplay) in the communication of humans and machines by approaching it from a theoretical, computational, and empirical side, and demonstrating its benefits.

Chapter 13 “Culture for Socially Interactive Agents,” by Birgit Lugrin and Matthias Rehm, introduces culture and its implementation in SIAs, and argues that implementing culture for SIAs can be beneficial not only to raise their acceptance in certain user groups but also to be able to teach about cultural differences and foster cultural diversity.

1.5.2 Volume 2

The second volume of this handbook starts with a preface that recaps the most important aspects and terminology of its introductory chapter. Part IV “Modeling Interactivity” explains how interaction with human users or other SIAs is modeled, and how the many detailed aspects of multi-modal, multi-party, adaptive interactivity are implemented.

Chapter 14 “Interaction in Social Space,” by Hannes Högni Vilhjálmsson, deals with the intricate social performance that inevitably takes place when SIAs and human users share the same social space (virtual or physical), regardless of their explicit intentions to connect with one another.

Chapter 15 “Dialogue for Socially Interactive Agents,” by David Traum, introduces several approaches to modeling the structure of extended verbal and multimodal interactions, with an emphasis on how different kinds of embodiment impact the communication affordances and requirements for SIA tasks.

Chapter 16 “The Fabric of Socially Interactive Agents—Multi-modal Interaction Architectures,” by Stefan Kopp and Teena Hassan, presents different SIA architectures and gives an extensive overview on how SIAs can engage in dynamic and fluid social interaction, discussing different approaches to deal with multi-modality and interactivity.

Chapter 17 “Multi-party Interaction Between Humans and Socially Interactive Agents,” by Sarah Gillet, Marynel Vázquez, Christopher Peters, Fangkai Yang, and Iolanda Leite, looks into SIAs that interact with a group of people for which the complex group dynamics need to be understood, and highlights that the SIA can affect and even explicitly influence the group’s dynamics.

Chapter 18 “Adaptive Artificial Personalities,” by Kathrin Janowski, Hannes Ritschel, and Elisabeth André, focuses on how a SIA can automatically adapt its personality in accordance with the user’s preferences, and with it make the interaction with them more enjoyable and productive.

Chapter 19 “Long-term Interaction with Relational Socially Interactive Agents,” by Jacqueline M. Kory-Westlund, Cynthia Breazeal, Hae Won Park, and Ishaan Grover, argues that strong relationships support people in achieving their goals in various domains, and thus relational SIAs have the potential to scaffold humans in their long-term endeavors.

Chapter 20 "Platforms and Tools for Socially Interactive Agent Research and Development," by Arno Hartholt and Sharon Mozgai, gives a practical introduction to the history of SIA platforms and tools directing to state-of-the-art technical solutions that support the development and implementation of SIAs.

Part V "Areas of Application" gives an overview of the major domains in which SIAs are employed, directing to systems and research findings, highlighting the benefits of SIAs to individuals and society.

Chapter 21 "Pedagogical Agents," by H. Chad Lane and Noah L. Schroeder, introduces work with SIAs in the domain of education, examining social aspects of teaching and learning and summarizing empirical research with pedagogical agents.

Chapter 22 "Socially Interactive Agents as Peers," by Justine Cassell, describes work that uses SIAs that are designed to work or play with children or teenagers at an eye level, discussing the benefits of SIAs that look and act like peers rather than teachers, tutors, or parents.

Chapter 23 "Socially Interactive Agents for Supporting Aging," by Moojan Ghafurian, John Edison Munoz Cardona, Jennifer Boger, Jesse Hoey, and Kerstin Dautenhahn, is centered on work with SIAs located in the area of aging support that aim to improve older adults' quality of life and wellbeing. The chapter provides methods and suggestions to address the many challenges that arise when designing SIAs that should successfully assist the targeted user group.

Chapter 24 "Health-related Applications of Socially Interactive Agents," by Timothy Bickmore, addresses another area of major societal importance, and highlights the potential of SIAs that have shown to have a positive impact on voluntary changes in health behavior.

Chapter 25 "Autism and Socially Interactive Agents," by Jacqueline Nadel, Ouriel Grynszpan, and Jean-Claude Martin, reviews work that uses SIAs to study or help improve the social skills of people with autism spectrum disorder. The chapter highlights the improvements that have been achieved throughout the last two decades and that, following a multi-disciplinary approach, more can be expected in the future.

Chapter 26 "Interactive Narrative and Story-telling," by Ruth Aylett, introduces narrative and storytelling as fundamental human capabilities, and outlines how SIAs are used in character- or plot-based systems, highlighting the great challenge of interactivity in this domain.

Chapter 27 "Socially Interactive Agents in Games," by Rui Prada and Diogo Rato, discusses the complexity in which SIAs have been used in games, and introduces their different roles alongside with their contributions to gameplay.

Chapter 28 "Serious Games with Socially Interactive Agents," by Patrick Gebhard, Dimitra Tzovaltzi, Tanja Schneeberger, and Fabrizio Nunnari, focuses on serious games that can partly be seen as a means to an end to achieve certain goals in various domains (such as education or health-behavior change) using specific methods from games and interactive narratives. Thus, the chapter focuses on learning gain as well as individual experience during game play.

1.6 Further Readings

Since the research area of SIAs is interdisciplinary, and researchers approach it from different angles and disciplines, a large number of books, conferences, and journals present work on SIAs. Below, we suggest further readings (in alphabetical order), but do not claim that the list is complete.

A few *books* have appeared that focus on SIAs:

- C. Bartneck, T. Belpaeme, F. Eyssel, T. Kanda, M. Keijsers, and S. Sabanovic, *Human–Robot Interaction—An Introduction*, Cambridge University Press, Cambridge, 2019.
- C. Breazeal, *Designing Sociable Robots*, MIT Press, Cambridge, MA, 2002.
- J. Cassell, J. Sullivan, and S. Prevost, *Embodied Conversational Agents*, MIT Press, Cambridge, MA, 2000.
- K. Dautenhahn, A. Bond, L. Canamero, and B. Edmonds, *Socially Intelligent Agents*, Springer, Boston, 2002.
- J. Gratch and S. Marsella, *Social Emotions in Nature and Artifact*, Oxford University Press, Oxford, 2014.
- N. Magnenat-Thalmann and D. Thalmann, *Handbook of Virtual Humans*, Wiley, West-Sussex (England), 2007.

In addition, *chapters* on SIAs are part of broader handbooks:

- R. Calvo, S. D'Mello, J. Gratch, and A. Kappas, *Handbook on Affective Computing*, Oxford University Press, Oxford, 2015.
    - Facial expressions of emotions for virtual characters, M. Ochs, R. Niewiadomski, and C. Pelachaud
    - Expressing emotion through posture and gesture, M. Lhommet and S. Marsella
    - Emotion modeling for social robots, A. Paiva, I. Leite, and T. Ribeiro
    - Preparing emotional agents for intercultural communication, E. André

  - Affect in human–robot interaction, R. C. Arkin and L. Moshkina
  - Relational agents in health applications: leveraging affective computing to Promote healing and wellness, T. W. Bickmore

- B. Siciliano and O. Khatib, *Springer Handbook on Robotics*, Springer International Publishing, Switzerland, 2016.
  - Cognitive human–robot interaction, B. Mutlu, N. Roy, and S. Sabanovic
  - Social robotics, C. Breazeal, K. Dautenhahn, and T. Kanda
  - Socially assistive robotics, M. Mataric and B. Scassellati

Likewise, a number of *conferences* addresses work related to SIAs: Please note that, as opposed to some other research domains, in the area of SIAs, and computer science in general, conferences and their proceedings are as important as (and sometimes even more important than) journal papers. High quality conferences have acceptance rates of 15% or lower. In the domain of SIAs, the following conferences are of relevance, albeit they strongly differ in their acceptance rates):

- International Conference on Autonomous Agents and Multiagent Systems (AAMAS) (http://www.ifaamas.org/), since 2002, proceedings by IFAAMAS, available by ACM Digital Library.
- International Conference on Computer Animation and Social Agents (CASA) (https://dl.acm.org/conference/casa), since 2004, proceedings by ACM.
- International Conference on Human–Agent Interaction (HAI) (http://hai-conference.net), since 2013, proceedings by ACM.
- International Conference on Human–Robot Interaction (HRI) (https://dl.acm.org/conference/hri), since 2006, proceedings by ACM.
- International Conference on Intelligent Virtual Agents (IVA) (https://dl.acm.org/conference/iva), since 1998, proceedings by ACM (Springer until 2017).
- International Conference on Robot and Human Interactive Communication (ROMAN) (https://www.ieee-ras.org/conferences-workshops/financially-co-sponsored/roman), since 1992, proceedings by IEEE.
- International Conference on Social Robotics (ICSR) (https://link.springer.com/conference/socrob), since 2010, proceedings by Springer.

A number of *journals* publish work related to SIAs:

- ACM Transactions on Human–Robot Interaction (https://dl.acm.org/journal/thri), since 2012, ACM.

- ACM Transactions on Interactive Intelligent Systems (https://dl.acm.org/journal/tiis), since 2011, ACM.
- Autonomous Agents and Multi-Agent Systems (https://link.springer.com/journal/10458), since 1998, Springer.
- Computers in Human Behavior (https://www.journals.elsevier.com/computers-in-human-behavior), since 1985, Elsevier.
- Frontiers in Robotics and AI, Section Human–Robot Interaction (https://www.frontiersin.org/journals/robotics-and-ai/sections/human-robot-interaction).
- IEEE Transactions on Affective Computing (https://www.computer.org/csdl/journal/ta), since 2010, IEEE.
- International Journal of Human–Computer Studies (https://www.journals.elsevier.com/international-journal-of-human-computer-studies), since 1994, Elsevier.
- International Journal of Social Robotics (https://link.springer.com/journal/12369), since 2009, Springer.
- Journal on Multimodal User Interfaces (https://link.springer.com/journal/12193), since 2007, Springer.

Acknowledgments

Many thanks to Catherine Pelachaud and Astrid Rosenthal-von der Pütten for their very helpful comments on this introductory chapter!

References

AAMAS. 2019. International Conference on Autonomous Agents and Multiagent Systems (AAMAS): Socially Interactive Agents Track. http://aamas2019.encs.concordia.ca/sia.html.

C. Breazeal. 2002. *Designing Sociable Robots*. MIT Press, Cambridge, MA.

C. Breazeal, K. Dautenhahn, and T. Kanda. 2016. Social robotics. *Springer Handbook of Robotics*, Springer, 1935–1972.

J. Cassell, J. Sullivan, and S. Prevost. 2000. *Embodied Conversational Agents*. MIT Press, Cambridge, MA.

K. Dautenhahn. 1998. The art of designing socially intelligent agents: Science, fiction, and the human in the loop. *Appl. Artif. Intell.* 12, 7–8, 573–617. DOI: https://doi.org/10.1080/088395198117550.

K. Dautenhahn. 2007. Socially intelligent robots: Dimensions of human–robot interaction. *Philos. Trans. R. Soc. Lond. Series B Biol Sci.* 362, 1480, 679–704. DOI: https://doi.org/10.1098/rstb.2006.2004.

K. Dautenhahn, A. Bond, L. Canamero, and B. Edmonds (Eds.). 2002. *Socially Intelligent Agents*. Springer. DOI: https://doi.org/10.1007/b116424.

E. Deng, B. Mutlu, and M. J. Matarić. 2019. Embodiment in socially interactive robots. *Found. Trends® Robot.* 7, 4, 251–356. ISSN 1935-8253. DOI: http://dx.doi.org/10.1561/2300000056.

B. Eckstein, F. Niebling, and B. Lugrin. 2019. Reflected reality: A mixed reality knowledge representation for context-aware systems. In *11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*, 1–4. doi: https://doi.org/10.1109/VS-Games.2019.8864516.

D. Feil-Seifer and M. J. Matarić. 2005. Defining socially assistive robotics. In *IEEE 9th International Conference on Rehabilitation Robotics*, 465–468.

T. Fong, I. Nourbakhsh, and K. Dautenhahn. 2003. A survey of socially interactive robots. *Robot. Auton. Syst.* 42, 143–166. DOI: https://doi.org/10.1016/S0921-8890(02)00372-X.

L. Hoffmann, N. Bock, and A. M. R. v.d. Pütten. 2018. The peculiarities of robot embodiment: Development, validation and initial test of the embodiment and corporeality of artificial agents scale. In *ACM/IEEE International Conference on Human–Robot Interaction (HRI'18)*, 370–378. DOI: https://doi.org/10.1145/3171221.3171242.

IVA. 2019. *International Conference on Intelligent Virtual Agents (ACM IVA)*. https://iva2019.sciencesconf.org/.

D. Kromand. 2007. Avatar categorization. In *DiGRA '07 - Proceedings of the 2007 DiGRA International Conference: Situated Play.* The University of Tokyo, Tokyo, 4. ISBN/ISNN: ISSN 2342-9666.

J. Li. 2015. The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *Int. J. Hum.-Comput. Stud.* 77. DOI: http://dx.doi.org/10.1016/j.ijhcs.2015.01.001.

R. McDonnell, S. Jörg, J. McHugh, F. Newell, and C. O'Sullivan. 2008. Evaluating the emotional content of human motions on real and virtual characters. In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization (APGV 08)*. ACM, 67–74. DOI: https://doi.org/10.1145/1394281.1394294.

S. J. Russell and P. Norvig. 2009. *Artificial Intelligence: A Modern Approach*. Prentice Hall, Hoboken, NJ.

B. Spanlang, J.-M. Normand, D. Borland, K. Kilteni, E. Giannopoulos, A. Pomés, Et al. 2014. How to build an embodiment lab: Achieving body representation illusions in virtual reality. *Front. Robot. AI* 1, 9. DOI: https://doi.org/10.3389/frobt.2014.00009.

D. Traum. 2008. Talking to virtual humans: Dialogue models and methodologies for embodied conversational agents. In I. Wachsmuth and G. Knoblich (Eds.), *Modeling Communication with Robots and Virtual Humans*. Springer.

S. Trepte and L. Reinecke. 2010. Avatar creation and video game enjoyment: Effects of life-satisfaction, game competitiveness, and identification with the avatar. *J. Media Psychol.: Theor. Methods Appl.* 22, 4, 171–184. DOI: https://doi.org/10.1027/1864-1105/a000022.

R. Zallinger. 1965. The road to *Homo sapiens* (illustration). In F. C. Howel (Ed.), *Early Man*. TIME-LIFE Books, New York, 41–45.

PART I

ESTABLISHING SIA RESEARCH

2 Empirical Methods in the Social Science for Researching Socially Interactive Agents

Astrid Rosenthal-von der Pütten and Anna M. H. Abrams

2.1 Motivation

This introductory methods chapter is meant to be an informative overview for all non-social scientists who work with socially interactive agents (SIAs) and who would like to familiarize themselves with empirical methodologies in psychology and the social sciences. It is primarily written for young scholars, that is, undergraduate or graduate students who are new to this field of research and new to empirical methods in the social sciences. We will clarify the research process and explain methods for studying research questions surrounding human-centered development, testing, and distribution of SIAs. In particular, we will provide answers to the following questions:

- What do we mean by methods in empirical social sciences? (Section 2.1.2)
- Why do I need methodological knowledge in empirical social sciences? (Section 2.1.1)
- Which research questions are addressed in empirical social sciences? (Sections 2.1.2 and 2.2)
- Which empirical methods should I use to address my research question? (Section 2.2)
- How does the chosen method work, in principle, and what aspects are important to consider when constructing, conducting, and analyzing my study and its results? (Section 2.2)

- Where can I find additional resources about methods in the empirical social sciences? (Section 2.3)
- What are the hot topics discussed in the community concerning methods? What are the current challenges and future directions? (Sections 2.4 and 2.5)

This chapter will also be useful for established scholars in the field as we provide an overview of different methods that can serve as inspiration. Furthermore, we included helpful material such as lists of online-tools, questionnaires, and specialized methods books and will point you into the right direction for further reading.

2.1.1 Why Do I Need Methodological Knowledge in the Empirical Social Sciences?

Depending on your discipline, you have a specific understanding of the term "methods." An engineer might understand methods as different systematic approaches that can be followed in order to reach the desired (technical) solution to a problem. The engineering method consists of stages such as idea, concept, planning, design, and then development of the former into a working prototype that demonstrates the solution to the problem [Ertas and Jones 1996]. The solution may be a tangible working prototype or an intangible working simulation. This prototype is being tested and debugged before launch. In computer science, depending on the problem statement, you might use theoretical, experimental, or simulation computer science methods. For instance, the experimental computer science approach [Zelkowitz and Wallace 1998] serves to identify concepts that facilitate solutions to a problem and then evaluate these solutions. One example for this evaluation process would be simulation studies with which researchers can evaluate a technology by executing the product using a model of the real environment, testing whether their hypothesis of the environment's reaction to the technology is supported. These are examples of methods that need no human in the loop (except for the engineer or computer scientist). In contrast to engineering and computer science, in psychology and the social sciences the human being and its relation to other human beings is the central focus of the research endeavor. Psychology is a scientific attempt to understand and explain human mental processes and behavior. Psychological science includes fields such as perception, cognition, attention, emotion, intelligence, subjective experiences, motivation, brain functioning, and personality. In social psychology this extends to interaction between people, such as interpersonal relationships. The social sciences are concerned with the scientific study of human society and social relationships.

The term SIA already implies why you will need to gain at least some knowledge about social science concepts and methods. SIAs are meant to be "socially" interactive, drawing on social psychological principles of interaction. Moreover, SIAs are developed to be deployed in social settings (rather than caged robot arms in production lines). Thus, their development and deployment involves an additional problem space than the technical questions that we have discussed above. For this additional problem it will be useful to know about empirical methods in psychology and the social sciences.

Consider that you followed a systematic approach to develop a social robot that helps to gather supplies in a hospital and assists nurses. You have run simulations to test whether it moves correctly and whether speech input is processed as intended. You have bench-marked two different navigation systems and two different natural language processing units and identified which one performs better on your training data. Now, you are ready to give the social robot the go to interact with humans. Will the human, let's say his name is Ben, find the robot useful? Is the interaction smooth? Does Ben understand the functionality of the robot? Does he like working with it? Does Ben consider the robot a team member? Does the social robot change the way how the human team members work with each other, and if yes, in what way? When you want to answer these questions, you need to know about the process of studying human perception, human behavior, and human attitude building. Ideally, engineers, computer scientists, and researchers in the field of psychology and social sciences work together in an interdisciplinary team from the start until the end of a development process following a human-centered design approach.

2.1.2 What Are Methods in the Empirical Social Sciences?

There are different methods for the acquisition of knowledge. We consider ourselves as social *scientists* and will therefore apply an empirical approach to acquiring knowledge instead of knowing because we have a "gut feeling," because it has always been like that, or because an authority said so. We will apply the *empirical method* that uses observation or direct sensory experience to obtain knowledge and uses evidence for verification of information [Gravetter and Forzano 2012, pp. 13–15]. Within the empirical method, we follow either the *hypothetico-deductive model* of the scientific method and engage in an "approach to acquiring knowledge that involves formulating specific questions and then systematically finding answers." [Gravetter and Forzano 2012, p. 16]. In contrast, there are also systematic methodologies based on empirical data but use *inductive reasoning*, for example, focusing on the construction of (new) theories through methodical gathering and analysis of data, such as grounded theory. This approach will only be briefly covered in this chapter (see Sections 2.2.3.1 and 2.4.2), but you will find recommendations for further reading in Section 2.3.

Once you have specified your research question or hypothesis, you have to think about your research strategy. In Section 2.2 you will learn more about different research strategies. Most commonly, in the field of SIAs researchers conduct evaluation studies. Evaluation is the process of developing and implementing a plan to assess something (e.g., your SIA) against the background of a specific research question or hypothesis using a systematic approach to assessment through previously defined measures (see Section 2.2.1). These measures can be quantitative and qualitative 2.4.2. Evaluations serve to determine the merit, worth, or value of something to inform judgements about the relative strengths and weaknesses, and the impact of variables. Since they are so prevalent in SIA research, we will put a focus on evaluation studies that can be realized differently, see Section 2.2.3.

2.2 Models and Approaches

How do you proceed once you have made up your mind that you want to do a study? In the following, we will guide you through the research process step by step. This section includes the research process in eight steps (see Section 2.2.1). Please note: the elaborations regarding the steps and important concepts and factors for each step are limited. In this book chapter, we can only provide a glimpse into the broad topic of empirical social science methods. In addition, you will find recommendations for further reading throughout this section and in Section 2.3. We provide two scenarios to exemplify how researchers derive a study design and which methodological choices they make considering the appropriateness of different methodological options. The following two scenarios are meant to give you concrete examples for methodological options when explaining the research process steps in Section 2.2.1; however, we also go through the full procedure of how to plan, conduct, analyze, and report a study using the two examples in Sections 2.2.2.1 and 2.2.2.2 to provide a more "hands-on" guide.

Example 1—Evaluating a Learning Robot: Imagine there is a competition within your class on social robotics. Using the Keepon robot platform, the students in your class form two teams, each building a social robot to assist children with vocabulary learning for Spanish. The robots differ with regard to the social roles they take on in interaction (see Figure 2.1). The robot of the T-Team acts like a tutor, while the robot of the P-Team acts like a peer. You want to know which robot is better in helping children and which team has won the competition.

Example 2—Developing an Agent for Assisted Living: You are working at a research lab in a third party-funded project with the aim of developing a virtual assistant for

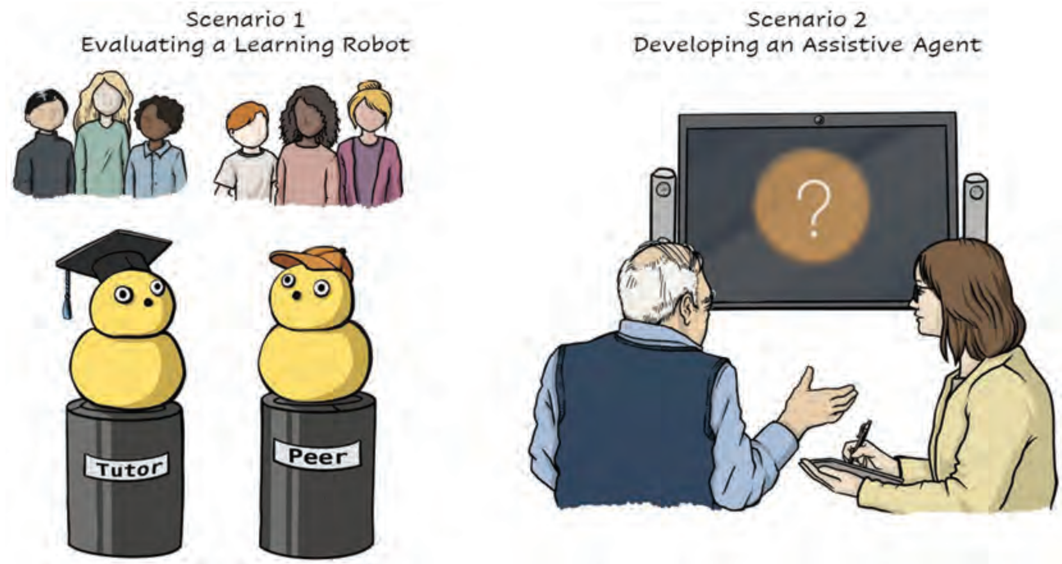

Figure 2.1 Example scenarios for study design.

older adults to be installed in their homes (see Figure 2.1). You are at the beginning of the project and want to know who exactly the target group is for this technology, what the virtual assistant should be capable of, and how it should look like. At the end of the project there should be a prototype and an estimate of whether this might be a successful product on the market.

2.2.1 The Research Process

The textbooks on empirical methods agree on the nature of the research process as involving at least eight steps [e.g., Gravetter and Forzano 2012]:

- Step 1—Find a research topic
- Step 2—Form a research question or hypothesis
- Step 3—Define the research strategy and experiment design
- Step 4—Operationalization of variables
- Step 5—Define and select sample
- Step 6—Conduct the study/data collection
- Step 7—Data processing and data analysis
- Step 8—Report results

For each step, we will provide an overview about (i) what has to be considered in the step, (ii) which methodological decisions a researcher has to take, and (iii) what the methodological alternatives are for the respective decision.

2.2.1.1 Step 1—Find a Research Topic

The research process begins with identifying the research topic. In our examples, the research topic is given because the lecturer decided to run a competition, or the funding agency provided money to develop an assistant to bring to market. However, when you are about to do a Bachelor's, Master's, or PhD thesis, you will have to define your own research topic. You might want to identify a human need and develop a SIA addressing human needs. You might be inspired by one of the topics covered in this book and want to contribute to this area, or you might have observed a social phenomenon in interactions with SIAs that in your opinion deserves further investigation. All of these exemplary approaches are valid methods to identify and define a research topic.

2.2.1.2 Step 2—Form a Research Question or Hypothesis

Once you identify a research topic, you will have to review the literature in that field and find the specific research question(s) you want to address. If applicable (and in most cases it is applicable) you should consult theories that are relevant to your research topic. The literature review will help you to *define your central concepts* and get an overview of which research questions have already been addressed, what *empirical evidence* is available, and where the *research gaps* are. This allows you to formulate research questions or derive hypotheses that are based on prior findings regarding that research question (see Section 2.2.2.1 on how to do this based on our examples).

2.2.1.3 Step 3—Define the Research Strategy and Experiment Design

There are many different ways to design a study. Your research strategy and study design depend on the type of research question or hypothesis you have proposed. Remember that a "research strategy is a general approach to research determined by the kind of question that the research study hopes to answer." [Gravetter and Forzano 2012, p. 159].

We will review different *types of research strategies* and explain when they are applicable: (i) the descriptive research strategy, (ii) the correlational research strategy, and (iii) experimental or quasi-experimental research strategies.

Moreover, regarding the latter research strategy, experiments, we provide additional information on how to design experiments and explain three *types of experiment designs*: (i) the within-subjects design, (ii) the between-subjects design, and (iii) the factorial design.

Research Strategies. The *descriptive research strategy* is "intended to answer questions about the current state of individual variables for a specific group of individuals" [Gravetter and Forzano 2012, p. 160] and is not concerned with relationships between variables. For instance, you could assess how much money people would be willing to spend on a virtual assistant or which functions they would want to have incorporated into the system. If you're going to examine the relationship between variables, there are two different ways to do so.

One approach includes simple observation of variables of interest as they exist naturally for a set of individuals. This is called *correlational research strategy*. If you are interested in the amount of money older adults are willing to spend depending on their income, you would choose a correlational strategy. You would assess people's willingness for investment and their income and run a statistical test on this data to discover a correlation. You would continue to examine whether there is any pattern of relationship between the variables and how strong this relationship is. This strategy can only *describe* a relationship but cannot *explain* the relationship because *correlation is not causation*.

Another approach to examining relationships follows an *experimental or quasi-experimental research strategy*. The *experimental research strategy* is "intended to answer cause-and-effect questions about the relationship between two variables" [Gravetter and Forzano 2012, p. 163]. You can answer questions such as "does interacting with a robot peer lead to longer attention in a learning task compared to interacting with a robot tutor?" To answer cause-and-effect questions, you manipulate one variable (the *independent variable*) to create so-called treatment conditions (robot peer vs. robot tutor). In addition, you prepare for measurement of a second variable (the *dependent variable*) to obtain a set of scores within each treatment condition (attention span while learning). It is of great importance that all other potentially influencing variables are controlled, as far possible. By *controlling all other variables*, you can conclude that differences in the scores of your dependent variable between treatment conditions is due to your manipulation of the independent variable. With regard to our example, you would design an experiment in which one group of children is interacting with the peer robot and one group of children is interacting with the tutor robot (two treatment conditions, independent variable: social role of robot) and you measure how long they focus their attention on the task (dependent variable: attention) and compare the scores between the two groups. It is crucial that you *randomly assign participants* to one of the groups. In the context of our example, this can easily be done by inviting children into your lab and randomly assigning them to treatment conditions.

Sometimes, however, it is not so easy to assign participants randomly to the experimental groups. Imagine, you are conducting the experiment in a school. For a whole week, you put the peer robot into one classroom and the tutor robot into another classroom. You cannot resolve the class structure for a week and randomly assign children, thus, you use two naturally existing groups: the classrooms. At the end of the experiment you are comparing learning gains in each class by administering a vocabulary test. This is called a *quasi-experimental research strategy*. Although, quasi-experimental settings use some of the rigor and control of true experiments, they are always flawed to a certain extend and cannot obtain an absolute cause-and-effect answer because there might exist other group factors systematically influencing the outcome. For instance, the human teacher in one classroom might motivate children more to use the social robot for learning vocabulary, thereby generating more learning time and greater learning gain (thus, in this example, the possible *confounding variable* is: motivation by teacher).

Experiment design. There are, however, many more methodological decisions required when planning an experiment. For instance, you have to decide whether to use a within-subjects design, a between-subjects design, or a factorial design. As for the *within-subjects design,* you would use a single group of participants who receive or experience all of the treatment conditions. Thus, a within-subjects design looks for differences between treatment conditions within the same group of participants. In this case you would have one group of children who interact with both robots, the peer and the tutor. In contrast, the *between-subjects design* requires separate independent groups of participants for each condition. In this case you would use two groups of children. Each group interacts with only one version of the robot. Sometimes, researchers want to investigate more than one independent variable. This would require designing a *factorial study design*. In the context of the current example, you might be interested in the question of whether girls and boys react differently to peer or tutor robots, thereby introducing a second independent variable (in this case a so-called *quasi-independent variable* because you cannot actively manipulate the gender of your participants, but there are naturally existing groups). When two or more independent variables are combined in a single study, they are called factors. Our example would be a two-factor design in which both factors have two values, resulting in a 2×2 factorial design with the factors gender (values: boy or girl) and the robot's social role (values: peer or tutor). You can design this study as a complete between-subjects design or as a so-called *mixed design* in which one factor is a between factor (gender) and one is a within factor (robot's social role).

When planning an experiment, please note that you might want to include a *control condition* (or a control group). A control condition refers to a non-treatment condition in an experiment where participants do not receive the treatment being evaluated. Here, a reference classroom group that does not interact with a robot but has normal class and is also measured in the dependent variable.

2.2.1.4 Step 4—Operationalization of Variables

The next important step in planning your study is the operationalization of your variables.

> In this step, we explain (i) what *operational definitions* are, (ii) why it is important to consider different *modalities of measurement*, and (iii) what *scales of measurement* exist.

In step 2 of the research process, the task was to identify theories relevant to your research question and to define appropriate constructs. The "problem" with *constructs* is that they are *hypothetical attributes* or mechanisms that help explain and predict behavior in a theory. Examples of constructs are motivation, knowledge, intelligence, or cognitive load. These constructs cannot be observed or measured directly, but it is possible to *observe and measure the external factors and the external behaviors* associated with the construct. Constructs can be influenced by external stimuli and in turn can influence external behavior. For instance, the theory of similarity attraction suggests that people are like others who they perceive as being similar to themselves, rather than dissimilar. Attraction is the relevant construct here. Attraction is hard to measure directly because it is a mental process. However, we can manipulate external factors such as similarity of the other person (e.g., similar = same gender/attitude/similar appearance; dissimilar = opposite gender/diverging attitude/diverging appearance). Moreover, we can observe and measure external behavior that might be affected by attraction such as a rating for how much we like that other person. What is needed is an *operational definition* that "specifies a measurement procedure (a set of operations) for measuring an external, observable behavior, and uses the resulting measurements as a definition and a measurement of the hypothetical construct" [Gravetter and Forzano 2012, p. 105]. This process is also referred to as *operationalization*. In our example, the construct similarity can be operationally defined in a variety of ways. For instance, for our group of participants evaluating the assistive agent, we created an agent more similar (matching gender) or dissimilar to them (opposite gender). Hence, we are comparing two different levels of similarity that in this case is defined by whether or not the agent has the same gender.

A simple way to come to the operational definition for the variables of interest is to consult previous research that made use of the same variable, because this research should report in detail how the variables have been defined and measured. By adopting these definitions and measurements in your study your results will be directly comparable to the results obtained in previous research.

Usually, there are different options for measuring any particular construct and variable. For example, when you want to assess acquired knowledge in a specific language you could use self-report and ask people how much they think they have learned, you could administer language tests (e.g., vocabulary) or observe whether the verbal behavior in that language has changed and is more fluent, more verbose, and contains less grammatical errors than before a treatment. In this example we would use different *modalities of measurement*: *self-report measures* such as interviews and questionnaires, and *behavioral measures* such as performance tests or behavior in interactions. A third modality are *physiological measures* (e.g., galvanic skin response, heart rate, or brain imaging techniques). All three modalities have certain advantages and disadvantages that can influence the quality of the measurement. There are two criteria for the evaluation of quality of operationalizations of variables and these are *validity* and *reliability*. A valid measurement has been demonstrated to actually measure what it claims to be measuring and a reliable measurement is able to produce identical results when it is used repeatedly to measure the same individual under the same conditions (see Gravetter and Forzano [2012, pp. 107–119]). If participants deliberately lie in a self-report this poses a threat to the validity of your measurement. In case you decide to use increased heart rate as a measure for similarity attraction you also might face a validity problem. Heart rate can increase due to a number of causes such as fear, anxiety, arousal, or embarrassment. The question is how can you be sure that measurement of heart rate is in fact a measurement for fear? To determine the validity and reliability of measures you should learn and read more about different types of validity and reliability in a methods book (see Section 2.3 for suggestions, e.g., some types of reliability can be tested for with statistical tests) and consult more closely the discussions in previous work using the variables you are using.

Once you have chosen the measures that you want to use in your study, you should be aware of the scale of measurement. Traditionally, there are four types of measurement scales: nominal scales, ordinal scales, and interval and ratio scales. *Nominal scales* represent qualitative (not quantitative) differences in the variable measured (some are female or male; being female is not superior nor inferior to being male). Categories on an *ordinal scale* are organized sequentially and consists of a series of ranks (e.g., first, second, third; small, medium, large). With an ordinal scale, you can determine not only differences but also the direction of differences (not the magnitude of differences). Interval and ratio scales are organized sequentially, and all categories have the same size [e.g., degrees in Celsius, each interval (degree) has the same size]. Hence, interval and ratio scales allow the determination of difference as well as its direction and magnitude. Interval scales have an arbitrary zero point (e.g., Celsius or Fahrenheit have an arbitrary zero point in addition to positive and negative values) while ratio scales have a meaningful zero point. For ratio-scaled variables, zero is the complete absence of something. The scale of measurement of your variables also determines which statistical test you can use when describing your data and when trying to discover relationships between variables. In this regard, please note that so-called Likert scales (explanation can be found below in the examples) that are most frequently used in self-assessments are ordinal scaled but given the robustness of many parametric tests can be used as interval scales in statistical testing (see Norman [2010]).

2.2.1.5 Step 5—Define and Select Sample

Once you have established your study design and measures, you should invest some thought into defining and selecting your sample.

> In this step we explain (i) what is a *population*, a *target population*, and a *sample*; (ii) different *sampling procedures* and when to use them; and (iii) how to determine the adequate *sample size* for your study by using *power analysis*. We therefore briefly explain *statistical hypothesis testing*.

First, we have to distinguish between the *population*, being the large group of interest to a researcher, and the *sample*, the small set of individuals who participate in the study. Very often, you will have a so-called *target population* that is defined by the researcher's specific interests. By target population, researchers address a group of individuals in the target population that shares one specific characteristic. For instance, a target population could constitute all German children in fourth grade or all individuals over 70 years living alone in an independent home. Usually, researchers do not have the means to draw a sample from the whole target population (all children in second grade), but from an *accessible population* (e.g., all children in second grade in one city). However, the goal is always to *generalize study results* of the sample to the population. Therefore, researchers seek to find a *representative sample* that closely mirrors or resembles the population and its defined characteristics. When the sample does not closely resemble the population but has different characteristics from those of the population, this is called a *biased sample*. Researchers have to be careful which sampling procedures they use in order to avoid sampling bias.

The likelihood of the sample being representative or biased depends on the procedure that is used to select participants for your study. There are two types of *sampling procedures*: probability sampling methods and non-probability sampling methods. *Probability sampling* methods require that the odds of selecting a particular individual are known and can be calculated. In order to do so, you must (i) know the exact size of the population and all its members, (ii) each individual in the population must have a specified probability of selection, and (iii) selection of individuals must be a random process. For *non-probability sampling* methods, the odds of selection are not known, the researcher does not know the population size and cannot list all members of the population. In this case, you do not use an unbiased method of selection. Thus, non-probability sampling methods have a greater risk of producing a biased sample. For the research field of SIA, not all population parameters are understood and can be identified. It is, thus, unlikely that you will be able to perform probability sampling methods. You will more likely perform non-probability sampling methods, such as convenience sampling. *Convenience sampling* means that you will be using those individuals who you have easy access to. Availability and individuals' willingness to participate are the decisive factors here. These are, for instance, students who are enrolled in one of your classes, or the children of the elementary school where you know teachers who are willing to help you in doing a study, or those people in the mall that happen to be there when you are conducting a field trial with your new social robot. Although convenience samples are obviously convenient, that is, less expensive and easier to get, they are also more prone to be biased. There are, however, ways to handle potential bias. You can ensure that your sample is reasonably representative and not strongly biased; for instance, you can work with schools from different districts of the city and be careful to select a broad cross-section of children (males and females, with siblings and only child, with and without immigration background). Moreover, you should describe your sample in detail in your research report and thus allow other researchers to evaluate how representative or biased your sample might have been and take this into consideration when evaluating the results of your study.

Once you know how you want to select your sample you have to determine the required *sample size* for your study—how large should the sample be in order to be representative? A general principle from statistics is the law of large numbers: the larger the sample size, the more representative the sample. There are, however, also practical limits to the sample size (e.g., time and expenses). Thus, most often you will have to compromise between the benefits and advantages of a large sample size and the costs of running a study with many participants. A *rule of thumb* is that you need about 25–30 individuals in every group you are testing [Gravetter and

Forzano 2012, p. 142] because accuracy of the sample mean in relation to population mean increases with sample size, but the improvement of accuracy slows dramatically once the sample size is around 30 (per experimental condition!). Because of this limited added accuracy, researchers often opt for a sample size of 25–30 per condition.

The sample size is also determined by other statistical factors that can be taken into account in a so-called *power analysis*, which is a statistical procedure to determine the required sample size for detecting an effect of a given size with a given degree of confidence.

Power stands for the probability of finding an existing effect and is influenced by the significance level, the sample size, and the effect size (high power diminishes the risk of false negatives). Given any three of these four components, we can estimate the fourth. Hence, when we know the significance level (e.g., $p < 0.05$), the assumed effect size of the effect we are looking for (e.g., $d = 0.5$, which would constitute a medium sized effect in a *t*-test), and the power we want to use in our study (e.g., 80%), we can calculate the required sample size for a *t*-test (e.g., 102 participants, 51 in each group). In a *t*-test, you determine the differences in means of two groups (children interacting with tutor or with peer robot). On the other hand, if you have a given maximal sample size (e.g., you only have access to 40 people with a very specific characteristic and no chance to get access to more individuals of that target population), the power analysis can determine the probability of detecting an effect of a given size with a given level of confidence. If you plan an experiment with two groups, trying to find an effect of medium size with 40 participants, the probability of determining this effect will be extremely low (power = 46%). This means that your study would have a 46% chance of finding a statistically significant effect of treatment condition given there really is an important difference between the treatment conditions. This might lead you to overthink and revise this experiment design. Statistics books often feature lists with examples of power analyses. There are also freely available software tools that help with performing power analyses (e.g., G*Power3; http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/).

2.2.1.6 Step 6—Conduct the Study/Data Collection

Before conducting a study, the last step should be to critically review everything that you have prepared and decided so far from an ethical viewpoint (see Chapter 3).

In this step we explain (i) *research ethics*, (ii) *informed consent* and *debriefing* of participants, and (iii) provide *useful tips* for conducting a study.

Considering *research ethics* is very important and in many countries it is institutionalized with so-called *institutional review boards* (IRBs) or *ethics committees*. A common process involves obligatory notifications to the IRB about every study involving human subjects. These reports should include detailed descriptions of the entire study, discussion of potential ethical concerns, and specification of measures to reduce potential harm to human subjects. The IRB commonly reviews study proposals and judges them upon ethical innocuousness. A positive evaluation of the ethics board is the official permission for conducting the study. IRBs often perform a risk–benefit analysis and assess the individual risks a participant is subjected to in a study and the benefits your research provides for society. One might tend to think that most research in SIA does not involve high risk for participants because people will not be physically harmed. Unfortunately, this is a fallacy because psychological harm can result from some studies. You might think that administering an IQ test in your study is a low-risk endeavor. However, when a person participates in this test and receives a low IQ score, this can seriously threaten the person's self-concept. IRBs usually provide guidelines on how to conduct studies that follow ethical standards. IRBs are governed by Title 45 Code of Federal Regulations Part 46 of the United States. Many other countries have similar rules for the establishment and working processes of ethical review boards. If there is no official regulation of the state, then, very often, universities and academic associations have committed themselves to establish an ethical review board. Even when there is no institution requiring you to do an ethical review, your research integrity should tell you to follow ethical guidelines and seek for guidance in this matter. Most academic journals and conferences will ask you to state whether your research has been IRB reviewed and might reject research that has not. Sometimes this can be avoided when you can explain in detail what measures you have undertaken in order to ensure ethical standards.

The American Psychological Association (APA) provides their "*Ethical Principles of Psychologists and Code of Conduct*" online for your reference (https://www.apa.org). The absolute basics for research are *informed consent* and *debriefing* (see section 8 in APA Ethics Code).

Informed consent means that you inform the participant that he or she is about to take part in a study and get permission to collect data [.20 2010]. This is especially important when you are collecting data that cannot be anonymized such as audio or video data. In this case, IRBs often require a process for data handling and data protection. The APA Ethics Code describes informed consent as follows: "psychologists inform participants about (1) the purpose of the research, expected duration and procedures; (2) their right to decline to participate and to withdraw from the research once participation has begun; (3) the foreseeable consequences of declining or withdrawing; (4) reasonably foreseeable factors that may be expected to influence their willingness to participate such as potential risks, discomfort or adverse effects; (5) any prospective research benefits; (6) limits of confidentiality; (7) incentives for participation; and (8) whom to contact for questions about the research and research participants' rights. They provide opportunity for the prospective participants to ask questions and receive answers." (APA Ethics Code, Section 8.02). Some universities or their IRBs provide examples or guidelines on how to construct an appropriate informed consent form.

Moreover, you should *debrief participants* properly, which means that "psychologists provide a prompt opportunity for participants to obtain appropriate information about the nature, results, and conclusions of the research, and they take reasonable steps to correct any misconceptions that participants may have of which the psychologists are aware. If scientific or humane values justify delaying or withholding this information, psychologists take reasonable measures to reduce the risk of harm. When psychologists become aware that research procedures have harmed a participant, they take reasonable steps to minimize the harm." (APA Ethics Code, Section 8.08). One specialty that frequently occurs in studies with SIAs is that researchers use a so-called Wizard-of-Oz (WoZ) scenario, [Dahlbäck et al. 1993]. This means that participants ostensibly interact with an autonomous system, but actually the social robot or virtual agent is controlled by a so-called "wizard," a hidden confederate of the experimenter controlling the actions of the robot or virtual agent. In this setup, participants are deceived about the true nature of the SIA (for a review on WoZ in HRI, see Riek [2012]). Deception is to be avoided unless the researcher has determined that "the use of deceptive techniques is justified by the study's significant prospective scientific, educational, or applied value and that effective non-deceptive alternative procedures are not feasible" (APA Ethics Code, Section 8.08). If your study setup includes any kind of deception, you are obliged to debrief participants as early as possible about the deception, preferably at the conclusion of their participation but no later than at the conclusion of the data collection, and permit participants to withdraw their data. For further discussion on deception in research, see, for instance, Christensen [1988] and Tai [2012].

When you have received the IRB approval, you can start recruiting participants, conducting the experiment, and collecting your data. Here are some *useful tips* that you usually do not find in a textbook but are based on experience. When recruiting participants, *always recruit more participants than you need*. There is always someone who does not show up or your technology is on strike on one day. You will experience that not all test trials produce suitable data to be included in your dataset. Thus, plan to recruit more participants than you need in order to cope for any dropouts. Clearly specify the *inclusion and exclusion criteria* for study participation.