

RESEARCH

Open Access



Behavioural Computer Science: an agenda for combining modelling of human and system behaviours

Tore Pedersen^{1,2}, Christian Johansen^{3*}  and Audun Jøsang³

*Correspondence:
cristi@ifi.uio.no

³ Department of Informatics,
University of Oslo, P.O.
Box 1080, 0316 Oslo,
Blindern, Norway
Full list of author information
is available at the end of the
article

Abstract

The rapidly increasing pervasiveness and integration of computers in human society calls for a broad discipline under which this development can be studied. We argue that to design and use technology one needs to develop and use models of humans and machines in all their aspects, including cognitive and memory models, but also social influence and (artificial) emotions. We call this wider discipline Behavioural Computer Science (BCS), and argue in this paper for why BCS models should unify (models of) the behaviour of humans and machines when designing information and communication technology systems. Thus, one main point to be addressed is the incorporation of empirical evidence for actual human behaviour, instead of making inferences about behaviour based on the rational agent model. Empirical studies can be one effective way to constantly update the behavioural models. We are motivated by the future advancements in artificial intelligence which will give machines capabilities that from many perspectives will be indistinguishable from those of humans. Such machine behaviour would be studied using BCS models, looking at questions about *machine trust* like “Can a self driving car trust its passengers?”, or *artificial influence* like “Can the user interface adapt to the user’s behaviour, and thus influence this behaviour?”. We provide a few directions for approaching BCS, focusing on modelling of human and machine behaviour, as well as their interaction.

Keywords: Behavioural sciences, Human aspect in technology, Internet of Things, Ubiquitous computing, Artificial intelligence and trust, Modelling, Security ceremonies

Introduction

The marriage of ubiquitous computing and AI opens up for an environment where humans will interact with autonomous systems that will be indistinguishable from other humans or other systems directly controlled by humans. Not only must humans relate to intelligent machines, the same machines must relate to humans and to other intelligent machines.

For humans, it could be a strange experience to interact with intelligent machines that might have contradicting traits of human behaviour. For example, a human normally reacts defensively (or aggressively) when physically attacked, but a human-looking robot might not react at all, or exhibit behaviour that seems strange to a human, but which is only intended at achieving the job that it was programmed for. The implications of this

disconnection between traits that are usually connected (i.e. human-looking and self-defence) is something humans will have to get used to.

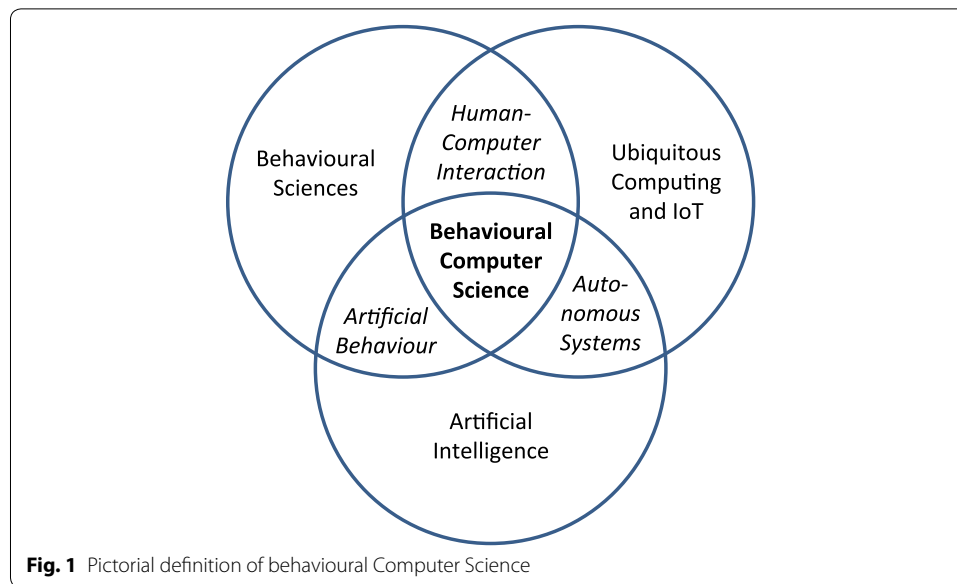
Our ethical compass should guide us to build intelligent machines that have desirable traits, whatever that might be. In order to achieve this goal it is essential that we understand how humans actually behave in interaction with intelligent machines, and this is a largely unexplored field. For example, what are the criteria for trusting an intelligent machine for which the intelligent behaviour *a priori* is unknown. Also, how can an intelligent machine trust humans with whom it interacts. Finally, how can intelligent machines trust each other. From a security point of view, the most serious vulnerabilities are no longer found in the systems but in the humans who operate the systems. In a sense, it is no longer a question of whether people can trust their systems, but whether systems can trust the humans operating them. These are daunting challenges in the envisaged intelligent ubiquitous computing and cyberphysical infrastructure.

Three important fields of scientific study are fundamental to understanding and designing the future smart-* infrastructures:

Behavioural sciences: working with systematic analysis and investigation of human behaviour through controlled and naturalistic observation and disciplined scientific experimentation. It attempts to accomplish legitimate, objective conclusions through rigorous formulations and observation. Examples of behavioural sciences include psychology, psychobiology, criminology and cognitive science. In contrast to traditional, rational and normative approaches to how people should ideally behave (we use behaviour as a general concept that includes the subcategories judgement and decision making), behavioural sciences give scientific, empirical, evidence-based, and descriptive approaches to how people actually make judgements and decisions. Thus, these two approaches are complementary: the rational agent model prescribes the ideal behaviour, whereas the behavioural model describes the actual behaviour.

Ubiquitous Computing and IoT: is a new paradigm in software engineering and computer science where computing is made to appear anytime and everywhere. In contrast to desktop computing, ubiquitous computing can occur using any device, in any location, and in any format. A user interacts with the computer, which can exist in many different forms, including laptop computers, tablets and terminals, but mostly everyday objects such as a fridge or a pair of glasses. The underlying technologies to support ubiquitous computing include Internet, advanced middleware, operating system, mobile code, sensors, microprocessors, new I/O and user interfaces, networks, mobile protocols, location and positioning and new materials. The IoT (Internet of Things) is the connected aspect of ubiquitous computing.

Artificial intelligence: abbreviated AI, is the intelligence exhibited by machines or software. It is also the name of the academic field which studies how to create computers and computer software that are capable of intelligent behaviour. Major AI researchers and textbooks [1] define this field as “the study and design of intelligent agents”, in which an intelligent agent is a system that perceives its environment and takes actions that maximise its chances of success according to some criteria.



We organize these areas under a new, broader concept which we call “*Behavioural Computer Science*” (abbreviated BCS). Any outcome of integrating models from these three areas would be called a *BCS-model*, which will always include behaviour aspects in some way or another. We would like to encourage research focus on the interactions between these three areas. This is illustrated in Fig. 1.

The intersections between any two of these areas represent existing or new research disciplines.

Human-Computer interaction (HCI) and interaction design [2] studies how a technology product should be developed having the user in focus at all stages. One focus of HCI is on the user interface (UI) of a technology, which allows the user to interact with the machine. Traditionally we have seen a two way communication, where the user voluntarily provides input through standard means such as button presses, and receives information e.g., through displays or lights. With the advent of ubiquitous computing, the Internet of Things (IoT) and advanced AI, the distinction and interface between computers and human becomes very blurred. In a smart-home environment the IoT constantly receives information from the human occupant of the house, without her even knowing it.

Artificial behaviour is an emerging discipline which focuses on understanding how intelligent systems behave from a macro-perspective, and not from a computer-program perspective. When intelligent systems become intelligent enough they will have psychological traits that can be studied. As soon as the machine has actuators it can also change its surroundings, which when humans are involved it can turn into *artificial influence* on human behaviour. Smart home and smart infrastructures, as well as web services based on big data, are emerging technologies highly susceptible to artificial influence. *Computational trust* [3, 4] becomes an aspect of machine learning or heuristics, that in turn will be part of IoT systems and other (semi-)autonomous (self-*) systems. For such systems that can perform complex computations and communications, we need to study notions of trust, like trust of the user in the system, or of another interacting system or

component. *Machine ethics* is that part of the Ethics of Artificial Intelligence concerned with the moral behaviour of artificially intelligent beings. The field of *Roboethics* [5, 6] is concerned with the moral behaviour of humans as they design, construct, use and treat such beings.

Models in computer science should be understood in a general and inclusive manner, as any formally or mathematically grounded models useful in some way for building IT systems. The purpose of using models is to be able to tackle complexity; since we constantly see complexity becoming the norm for current day technology. Examples of established classes of models include:

Probabilistic models: some useful for machine learning others for describing cybe-physical systems,

Logical and formal models: often used for reasoning or for describing computation and machine behaviour,

Programming and their many types of models: used to precisely describe what a program does or to specify what it should do and check that it conforms to the specification.

Our proposal is that BCS also studies models of human behaviour and how humans and intelligent machines interact, and moreover, how such models can be combined with computer science models. In this way, the human part of the system can also be considered by the same (or enhanced) tools and methods that computer science uses to establish safety, security, reliability, and other properties that the human-machine system is supposed to satisfy.

Two motivating examples

To motivate the need for BCS modelling we will use examples from security ceremonies and IoT, but our results could also be used in other situations, from interactions in the private sphere, up to interactions concerning broader societal affairs, and even systems handling intelligence and national security. The following are examples of emerging fields which can be seen as part of BCS.

Security ceremonies have recently seen increased interest since they strive to involve the human aspect when designing and analysing security protocols [7, 8]. A few works have studied the human aspect of security breaches [9–13]. An example is spear-phishing attacks [14], where we see technology developers taking the attitude that “the breach will occur”, so they try to protect against it through e.g. network isolation of the infected system. This attitude in practice ignores existing research into e.g., learning techniques [15, 16], using visual queues [17], or that aim to teach the users against phishing [18]. We argue that cognitive models and models of social influence can give insights into how to build e-mail systems that can counter more effectively such targeted, well-crafted, malicious e-mails.

Home automation and ambient assisted living [19, 20] is one of the applications of IoT that is most closely interacting with humans occupying the house. Such systems need to learn patterns of behaviour, preferably distinguishing them among several occupants,

adapt to temporary changes in behaviour, as well as interact and take control requests from the humans.

Summarising the contents of this paper

We first argue in "[Behavioural computer interaction](#)" section why behavioural sciences are the desired place to look for models of human behaviour and social influences which impact the interaction of humans with the new technology. We identify important differences between a rational agent model and an empirical model of human behaviour, while providing a brief review of relevant literature related to the idea of complementing rational models with empirical data about human behaviour. Then in "[A reference model for BCS](#)" section we propose the Bella–Coles–Kemp model as an abstract model to anchor our discussions in BCS. "[Behavioural and social aspects of humans and technology](#)" section is the first of the two main sections of the paper where we identify several aspects of human behaviour that influence interaction with technology, and which we think are good to start modelling and using in automated reasoning and IT development tools. We briefly review relevant literature and provide an overview of psychological phenomena that affect human–technology interaction. In "[Modelling for behavioural computer science](#)" section we discuss various aspects of modelling and how these can best be applied to the notions identified in "[Behavioural and social aspects of humans and technology](#)" section while having the goal to be combined with computer science models. In "[Conclusion](#)" section we conclude and point to immediate future research, as well as reiterate the societal domains in which BCS may have the most impact. The concept of Behavioural Computer Science has already been presented, in a more restricted form, in the conference paper [21]. Here we provide more details, arguments, and examples, including more recent and matured views than what appeared in the short paper [21].

Behavioural computer interaction

In domains where humans interact with technology it is necessary to understand human behaviour in order to capture or foresee possible actions taken by humans in interaction with the technology. We refer here to an understanding that can be used by machines, thus through models that can be used in some forms of computations. If technology and their designers understand the typical tendencies of human cognition, emotion and action, it is easier for the resulting system to take into consideration how people actually behave, and adapt to this, instead of relying on assumptions about how they may behave.

As this implies, there are two primary approaches to including human models: one would follow the Rational Agent Model, prescribing the ideal behaviour (e.g., [22–24]) and another the Behavioural Model of Human Agency, describing the actual behaviour (e.g., [25, 26]). The rationalist approach to explaining human behaviour from a priori assumptions, is traditionally widespread—in academia, in professional practice, as well as in society in general. The rationalist tradition generally adheres to the view that people are rational agents that seek to maximize utility. Inherent in this approach lie the assumptions that (i) people know their ultimate goal, (ii) have the means to select the courses of action that are the most likely to lead to goal achievement, and (iii) have the capability to carry out the appropriate courses of action. To do arrive at this end-state,

people would need (i) to have unlimited access to all information, (ii) the ability to discriminate relevant from non-relevant information, (iii) the cognitive capability to handle and analyse the interaction between the relevant informational components inherent in the possible courses of action, (iv) to calculate how the courses of action would lead to the possible end-states, and (v) to foresee implications of the end-states. We shall soon see that these assumptions are seldom fulfilled, which leads us to focusing on the Behavioural Model of Human Agency.

One of the first proponents of the behavioural model was Herbert Simon, the 1978 Nobel Laureate in Economics. Simon found that people, when making real judgements and decisions, did not comply with the ideal that was assumed by the rationalist traditions. He was the first to coin the term Bounded Rationality [24] to describe the concept of non-ideal adherence to the rationalist assumptions and thus complemented the traditional rationalistic approach in the field of economics.

Prominent scholars in behavioural Science, after Herbert Simon, are Daniel Kahneman, the 2002 Nobel Laureate in Economics, and his colleague Amos Tversky. Notable findings in Kahneman and Tversky's research [26, 27] are that people often rely on intuitive thinking when making judgements under conditions of uncertainty. Intuitive thinking, when employed inappropriately in conditions when instead analytic thinking would have been the correct cognitive strategy, often leads to biased (and consequently incorrect) judgements. Although the rationalistic approach is valuable in explaining how people should ideally make judgements under conditions of certainty, the behavioural approach is the better in explaining how people actually make judgements in uncertain conditions, and also identifying the bias-inducing psychological mechanisms people employ (mostly without conscious awareness) [28].

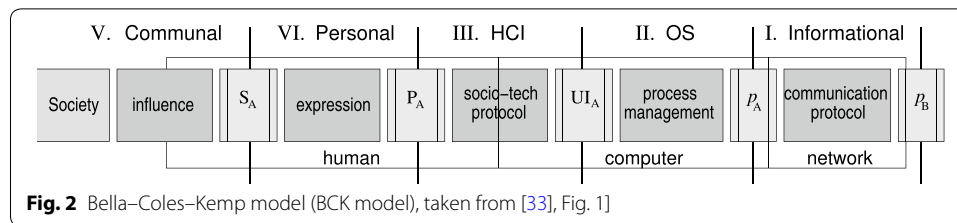
Consider three examples where the behavioural approach to explaining human judgement has successfully enriched an existing academic discipline:

Behavioural economics: focusing on how people actually behave in economic contexts, as opposed to how they should ideally behave (e.g., [26, 29]), has been a fruitful addition to Economics;

Behavioural game theory: focusing on how people actually behave in formal games, as opposed to how they should ideally behave (e.g., [30]), has enriched traditional Game Theory; and

Behavioural transportation research: focusing on how people actually make choices in transportation and travel contexts, as opposed to how they are assumed to behave (e.g., [31, 32]), has been a fruitful addition to the traditionally rationalistic field of Transportation Research.

Behavioural Computer Science can be one more fruitful collaboration between behavioural sciences and computer science, and this paper gives a few venues of exploration. In particular, such collaborations could have a good influence on the emerging field of IoT, as well as the human aspect in Security.



A reference model for BCS

To anchor our thoughts we will use a model introduced in [33], which we call “*the Bella-Coles-Kemp model*” and abbreviate as *BCK model*. This is a rather general and abstract model for any forms of human involvement in computer systems, thus providing a good common basis for defining Behavioural Computer Science concepts.

The BCK model, pictured in Fig. 2, provides abstractions and separations, while allowing for more details to be given for each of the vertical layers individually. We can see the behaviour sciences (like psychology and cognitive sciences) as a good ground for finding specific models for the layers V and IV; whereas for layer III we already have a good start in works from Human-Automation Interaction (HAI). Traditional computer science studies layers I, II, and also partly layer III.

When explaining the BCK model it is good to make correlations with existing established concepts; and we choose here models for security protocols, usually based on the Dolev-Yao assumptions [34], and using specification languages like the applied pi-calculus [35, 36]. Usually, security protocols are formed of the parties (or players) and the interaction medium they use for communication (or for any other exchange of information). The parties are usually honest, whereas the intruder (attacker) controls the interaction medium. More than two parties can be involved in a protocol, but for our example purposes here let us consider only two honest parties, Alice and Bob. Third parties, usually dishonest, appear due to the ability of the intruder to disguise as a party in any number of protocol runs. The Dolev-Yao model defines the powers that the attacker has over the interaction medium, like power to delete, change, or insert messages, to and from any other party.

In the BCK model the parties form the light boxes, whereas the interaction medium forms the dark boxes. The parties appear at different layers of the BCK model and in different abstractions; i.e., the light boxes represent the players in the respective layer, which are abstractions of the parties or are controlled by the parties. In the layer I (also called “Informational”) we encounter the processes p_A and p_B controlled by Alice respectively Bob, which are running the computers of Alice and Bob, communicating through the network, i.e., the dark box. Layer I would thus be studied by communicating systems and networking researchers, and for security it could be subject to the Dolev-Yao assumptions. However, this layer has also aspects not related to security, like properties of the transmission medium (e.g., messages can be lost or not, or delayed and for how long).

In BCK other players appear at the other layers: at layer II (also called “Operating System”) *the user interface* UI_A associated to Alice, which interacts with the computer process p_A , e.g., by sending information taken from the user required by the protocol

run by p_A , like a password or biometrics. The same UI_A interacts at layer III (also called “Human–Computer Interaction”) with a “*persona*” P_A of Alice for some particular system. We adopt the notion of *persona* to characterize a user in a socio-technical interaction situation. We define a *persona* to be a set of social and cognitive attributes of a human, including emotions, senses, or memory [37]. The *persona* has interaction at layer IV (also called “Personal”) with *the self* S_A of Alice, which in turn is influenced by the Society through various social interaction protocols at layer V.

Players may interact only as part of a layer, and one layer may involve players pertaining to different users. Important to note is that in BCK one player usually is involved in two adjacent layers.

Research in computer science has mostly focused on layers I and II, and largely ignored layers III–V. We see layer II as pertaining also to the technological community, whereas layer V would pertain to the social sciences. Layer IV would be investigated more by the psychology researchers. Layer III on the other hand is at the interaction between technological and social sciences, with a rapidly evolving field, having terminology s.a.: HAI [38], user-centred design [39], interaction design [2].

One usefulness of BCK is to make explicit the need for collaboration between the fields of sociology, psychology, and technology, in order to tackle the complexities of current systems s.a. security ceremonies or IoT. One can very well focus on individual layers, but the BCK model brings the isolated results into the general picture which eventually needs to be handled in order to claim results and designs of practical use.

We expect BCS research to detail all the layers III–V, the same as has been done until now in CS with the layers I and II. The interaction medium, the dark box, can be split into more fine-grained divisions, and each division would have its interaction protocol and assumptions. For layer II it is easy for computer scientists to bring their knowledge of operating systems design and see that a UI could consist of a screen and its driver, a display client, like a browser displaying an input form, a keyboard with its drivers, and the many other components that transport the information between these many UI components and the end process p_A . However, for the social protocols of layer V, completely different concepts and models should be used. One could study various means of social manipulation, and quantitative and qualitative measures could be devised for analysing their usefulness in terms of power to influence, e.g., depending on the social scale or training level of the users, i.e., the self S_A .

The BCK model has been applied to study security ceremonies in [40] where the focus is on layer III and introduces probabilities as opposed to classic (rationalist view) non-deterministic models. The concepts discussed in this paper are concentrating on layers V and IV, and how these could be modelled and combined with the methods used in CS in layers I and II. Works on layer III constitute a good middle ground [41–43]. We particularly wish to focus our modelling efforts for V and IV on behavioural research, and depart from the rational view and assumptions on human.

Behavioural and social aspects of humans and technology

In many domains, academic as well as professional and political, it is a generally held view that people make rational judgements and thus are assumed to think, act, decide and behave according to the rational agent model [24]. The rational agent model implies that people always strive to maximize utility.

With regard to the concept of *utility*, this is generally understood as the satisfaction people derive from the consumption of services and goods [44]. From this perspective, it is an overarching assumption that every individual knows his or her ultimate goals and also how to fulfil their goals. If one looks at utility from a psychology perspective, a problem arises because there is more than one definition of utility.

Experienced utility: is the satisfaction one derives in the consumption moment, and thus the most valid and objective measure of in-situ utility.

But there are two other types of utility that are different from this:

Predicted utility: (or, alternatively, *expected utility*) is the utility one predicts (and expects) beforehand that one will experience in the future consumption moment.

Remembered utility: is the utility one remembers having experienced in a consumption moment some time ago.

The problem with these three different aspects of utility is that the rational agent model implicitly assumes them to be equal, whereas empirical psychology research [45, 46] has found that these aspects reflect different utilities; the utility that one actually experiences may be different from both what one beforehand predicted and what one later will remember. The rational agent model does not take this into consideration when it generally regards predictions, experiences and memory of utility as representing the same type of utility.

Another problem with the rational agent model regards the concept of *rationality*. In this view it is assumed that people act strictly logical and rational in the pursuit of maximized utility. Inherent in this view, conditions are assumed to be certain, meaning that every individual is assumed to have unlimited access to all information and is also capable of analysing the relevant information needed to make a judgement, as well as calculating the outcome of every combination of informational components, so that the best decision can be made. Of course, no single individual is able to adhere strictly to this model of rational behaviour, but the point here is that the rational agent model assumes rationality as a general principle, and does not concern itself with empirical evidence about actual behaviour (i.e., as opposed to inferred behaviour) as to whether this assumption is actually valid.

In summary, *some errors in human behaviour* often stem from the differences between predicted, experienced and remembered utility; e.g. when making judgements at time t_0 about some consumption related moment in the future at time t_1 , one often disregards the fact that their current experiences will be different from their expectations. Errors may also occur as a consequence of making judgements in conditions under uncertainty, i.e., when the requirements of the rational agent model cannot be fulfilled.

Although many proponents of the rational agent model acknowledge that the model is ideal and that it not necessarily realistically reflects actual human behaviour, they nevertheless assume that individuals' errors of judgement or of informational processing (i.e., their failure to comply with the model and thus make perfect judgements), are non-systematic. Non-systematic errors mean that each single individual does not exhibit the same types of judgement errors as other individuals and, furthermore, these differ over time, i.e., errors made today are different than those made yesterday. Although it is acknowledged that people are not perfect judgement machines and thus make errors, it is assumed that mistakes are non-systematic and thus random [28, 29] (classically modelled through non-determinism).

Kahneman [29], and other behavioural scientists, questioned the explanatory powers of the rational agent model, because they could not make their empirical data fit the rational agent model. As psychologists—or behavioural scientists—they studied how people actually behave, as opposed to how they are assumed to behave according to the rational agent model. Thereby they provided empirical data that supported a new view—namely that people's judgement errors were not at all non-systematic and thus random, but in fact systematic; people tended to make the same kinds of misjudgments as others did, and misjudgments made today are the same as those made yesterday. Thus, people's mistakes were more or less universal. Findings like these paved the way for an extension of Bounded Rationality as a concept of understanding human behaviour [24].

One major and universal finding in this new avenue of research is that there are two fundamentally different systems of cognitive processing [29, 47]:

System 1: Intuitive Thinking, is associative, effortless, emotion-influenced, automatic, and thus often operating without conscious awareness;

System 2: Analytic Thinking, is analytic, effortful, not influenced by emotions, sequential, controlled and thus operating with conscious awareness.

Because Intuitive Thinking is effortless and automatic, people have a tendency to rely heavily on this cognition mode in most everyday activities—where we automatically know how to judge, behave and decide—and it works fine. The problem is that we sometimes employ this automatic mode of thinking also in situations where we have less knowledge or experience. A failure to activate Analytic Thinking thus results in what is now commonly labelled as *biased judgements*.

Another major finding from behavioural sciences that is relevant to BCS is the discovery of four psychological mechanisms (also called *heuristics*) that are mostly responsible for the human tendency to make unwarranted swift judgements [27]. These four mechanisms – leading to biases in situations where we are uncertain – are inherent in Intuitive Thinking (which is thus sometimes called Heuristic Thinking). When we are making judgements under conditions of uncertainty, we are known to employ one or more of these heuristics, which often fail to make correct judgements. Let us now take a look at each of these heuristics and define their major characteristics.

The availability heuristic: Captures how people make judgements based on what is easily retrievable from memory, or simply what comes easily to mind. An example is when

asked to list as many English words as possible that begin with the letter A. This is a simple task, because words beginning with the letter A are fairly easily retrievable from memory. However, consider listing as many words as possible that have the letter A as the third letter in the word. This is a much more difficult task, because words with the letter A as the third letter are less easily retrievable. As a consequence, one may conclude that there are more English words beginning with the letter A, than words having the letter A as the third letter, simply because instances of the former is more easily accessible.

The representativeness heuristic: Describes how people make a judgement based on how much the instance or the problem in front of them is perceived as similar to another known instance or problem. If the degree of perceived similarity is large enough, people will easily make incorrect judgements. An example is when asked to give answers to questions under strict time constraints, and to reply as fast as possible. If one is shown a picture of a whale and she incorrectly labels the whale as a fish, an incorrect judgement is made based on the representativeness heuristic.

The anchoring and adjustment heuristic: Implies that people—under conditions of uncertainty—without conscious awareness will establish an “anchor”, and from this anchor adjust their judgement, often in the “right” direction, although not to the point of accuracy. In a condition of total uncertainty, even non-relevant information that one has been primed with, or that is easily accessible from memory, can serve as an anchor.

The affect heuristic: Explains how the current affective state may influence human judgements. For example, when in a positive mood, one may be more easily susceptible to deception and manipulation because of a tendency to making hasty and possibly incorrect judgements, whereas one may be less inclined to do so when in a negative mood.

To *counteract* the tendency towards the Intuitive Thinking, in order to make people less susceptible to the heuristics that may generate incorrect judgements, one possible intervention could be to “slow” people’s actions down, inviting them to be consciously aware of their actions, and thereby make them employ System 2-thinking. The message that we get when trying to delete a file, saying “Are you sure you want to delete this file?” is an example of such an intervention.

For the spear phishing example, where one receives a malicious email from an address that resembles that of a known colleague. This is an attack that is difficult to counter because it *activates both the availability heuristic and the representative heuristic*; the user may or may not have no easily accessible information stored in the mind that may suggest that this is an hostile attack (susceptibility to the availability heuristic) and, furthermore, the user recognizes the email address as being from a near colleague (the representative heuristic). Additionally, when considering that malicious attackers could also employ mechanisms of social influence [48], such as the six principles of persuasion, wherein e.g., the concept of Authority (people have a tendency to obey instructions from authority figures) or Liking (people have a tendency to be more easily persuaded by people they like), they have access to a versatile tool-kit of psychological manipulations and deceptions, which they could use with malevolent intent. Thus, no alertness or caution is prompted.

Human choices and human prediction power are very important for interactions with computer systems, e.g. security can be influenced by poor predictions about the possibilities of attacks and attack surface can be wrongly diminished in the mind of the human, whereas wrong choices can incur safety problems. In [26] it is argued that it is difficult for a human to make accurate predictions about a situation or an experience (e.g., sentiment, preference, disposition) when the future forecasting time point t_0 is rather distant from the current time point t_1 on which the same experience is evaluated. The more distant this time point is, the more inaccurate the prediction (and thus the choice) will be.

Modelling for Behavioural Computer Science

We are interested in how behavioural concepts could be mathematically modelled, and more importantly, how these behavioural models can be coupled and integrated with existing models from computer science. Thus, our study here pertains to the layers V and IV of BCK. We start discussing a very simple model, building on the work done in [40] for layer III, based on works from HCI [49–51] or from cognitive theories [52, 53].

One point made by Kahneman and Thaler [26, 29] is that the circumstances (i.e., the context of the human and of the system) vary between the present t_0 and future t_1 time points. Four large areas of such *varying circumstances* can be identified:

The emotional state of the human, or the **motivational state** of the human might vary when t_0 and t_1 are distant from each other.

The aspects of the choice, of the product, of the experience, that are considered as important or are made salient/observable at t_0 , might not be present at t_1 or may be difficult to experience or observe at this later time point.

Memory of similar choices or experiences is important. If the memory is biased then the current choice and prediction for the future will be biased. Tests of memory manipulation have been made [45] and one observation is summarized as the *Peak/End Rule*, as opposed to the common belief that the monotonicity of the experience counts. Humans recall more easily the experiences of the peak emotions or emotions during the end of the episode, which wrongfully become representative of the total experience. *Affective forecasting* [32, 46] is a concept introduced to explain that when focusing on some aspect for making a decision, this aspect will inappropriately be perceived as more important at the time of (prediction and) decision than it normally will be at the time of experience, thus leading people to predict that future incidents will have a higher emotional impact – negative as well as positive, depending on whether the future incident itself is perceived as negative or positive – than what will usually be the case.

These four areas need to be modelled in a unitary way in order to capture formally (with mathematical structures) the context of the human behaviour and its variability. Moreover, the resulting mathematical structures should be made such that it is easy to combine with models of the systems.

We will thus work with a notion of “State” and changes between states (which we usually call “Transitions”). How exactly to model an *emotional or motivational state* is not trivial, and we discuss these in more detail later. When we model *emotions* (as needed

for *affective forecasting*, as well as for many aspect of the Self) we can start from the following concepts related to the *impact bias* [46]: the *strength* (or *intensity*) of an emotion and the *duration* [54]. Both of these can be quantified and included in a *quantified model of emotions*.

Changes between states can be, as we have already mentioned, “*temporal changes*”, i.e., changes that happen because of passage of time. These we can consider in two fashions:

Gradual/continuous change of emotion or motivation happens over time (e.g., modelled with time derivatives, in the style of physics); or

Discrete changes where we jump suddenly from one value to a completely different value (e.g., think of motivation which can gradually decrease until it reaches a threshold where it is suddenly completely forgotten).

Other temporal notions different than durations could be needed like *futures* or order *before/after*, for which there are well established models in computer science, e.g. temporal logics [55–57].

Events is another concept that we identify as influencing the Self, in the sense that *emotions are relative to events*. Events can be considered instantaneous and are sometimes modelled as *transitions* labelled by the respective event. The reason is that the event changes the state in some way, e.g., changes the memory of the Self, or attributes of various variables of the context, as well as of the Self.

A cognitive explanation for people’s biased retrieval of past experiences appears when we relate them to emotions. Whereas currently experienced emotions (related to a currently experienced event) are stored in the episodic memory, past experienced emotions (related to previously experienced events) are stored in the semantic memory. The semantic memory is largely susceptible to biases due to the influence of current beliefs about previously experienced emotions on the retrieval of memories [58]. Thus, memory may make people behave in ways unexplainable by the rational model. These concepts contribute to defining *models for the predicted and the remembered utilities*, as well as how these models correlate with that of the experienced utility.

For modelling a State we can start by including the *aspects* that are of interest for the situation under study. Aspects could be modelled as logical variable that are true or false in some state, because they are either considered or not considered (i.e., observable/salient or not). The expressiveness of the logic to be used would be dependent on what aspects we are interested in; but we can start by working with predicate logic. Depending on the system being developed, we encourage to choose the most suited logic, e.g.: the SAL languages and tools which have been nicely used to describe the cognitive architecture of [10], Sec.2]; or one can use higher-order dynamic logic [59], Chap.3] and the tools around it like the KeY system [60].

Modelling memory and especially how can memory be manipulated and how the memory influences choices and thus transitions between states is not easy. Quite a few studies can be considered [61, 62], some of which are more close to models and to logics [63–65]. We can also use models from computer science and logics like dynamic logic [59], used to talk about programming data structures, but also logics of knowledge and

belief [66] which have well developed models for how beliefs can be updated over time due to various kinds of changes [67].

Modelling the preferences and priorities Another important concept is that of *focusing illusion* [68], which is the illusion that an attribute/emotion that the human focuses on (since it is relevant for the respective emotion or activity or situation etc.) is more important than it actually is. The question is how to model the fact that an attribute is important?

- One alternative is to use *weights* and weighted models [69, 70]. We would then need empirical methods for automatically learning the weights as well as for measuring the importance/weight of the respective aspect.
- To fully capture focusing illusion we need to also include in the model a measure of how much *overrated* is the respective weight of the attribute in the current situation.
- Another question then is: How are these weights affecting the UI, persona, or the properties of the whole BCS-system?

These concepts cumulate in a model for the *Self*, involved in layers V and IV. Now the question is how does this model relate to the model of the *Persona* (that is involved in the HAI at layer III) and with the *Society* (at the outermost border of the BCK model)?

The relation between the Self and the Persona can be seen as a simplification relation (called projection in more formal terms). The projection operation is done on a subset of the variables that make up the State of the Self, thus resulting in the state of the *Persona*. This projection would retain only those aspects that are relevant in the respective context, i.e., in the context of the computer system being studied. This means that the projection operation should also be related to the model of the UI (i.e., the one between layers II and III).

The simplification relation is however not enough. We need to understand the interactions between the *Self* and the *Persona*. We can see two interaction directions:

From the Persona to the Self i.e., to the user with all the experiences, sensors, memory, thinking systems, heuristics, etc.; and

From the Self to the Persona i.e., to a simplified view of the user, specifically made for the UI and the system being studied.

Since a *Persona* is an abstraction of the human relevant for the interaction with a specific UI, then through the *Persona* we can see stimuli from the UI going to the *Self*, and influencing it. Therefore, the first communication direction can be seen as communications coming from the UI but *filtered through the Persona*.

For the second direction we see more the actions of *expression* (e.g., described by [10, 50]) that the *Self* makes out of the thoughts, reasoning, intuition, past experiences and memory models, into something relevant to this BCS-system and to the UI that the human interacts with. In consequence, we may say that the *Self* interacting with the UI is filtered by the *Persona* we designated. But this *Self* is aware of more than just the UI: maybe she is aware of computer networking aspects (which pertain to layer I) or operating systems aspects of layer II (like how browsers work or how the operating system can

be protected from bugs and viruses, whether an antivirus is installed or a firewall, etc.). All these examples are outside the direct visual interaction of the Persona with the UI, which is captured through the layer III, thus part of layer IV. The few studies on modelling human interacting with a user interface using formal methods [33, 71] are based on a non-deterministic approach, whereas [40] develops a probabilistic model in a more general setting.

Interactions at layers V and IV would be studied empirically, including also models of the Self and of Personas. A model would *start from general assumptions*, incorporated as prior information/probabilities. For a specific system, with a specific Persona defined, the model would need to be updated, in a continuous fashion, by gradually learning from each empirical study and new evidences, thus *updating the priors*.

Because we use empirical evidences we need to introduce a notion of *uncertainty about the probabilities* that the studies reveal. Therefore, models of *subjective logic* [72] could be useful for expressing things like: “The level of uncertainty about this value given by this empirical study is the following.”

One would then be interested in applying standard analysis techniques like model-checking [57] over these new models with uncertainty. This would allow to:

- Find ways how to protect the Self from malicious inputs and manipulation from the UI through the Persona.
- Find ways to protect the Self from the Social interaction in layer V, commonly called social-engineering attacks.

One type of such protective methods are known as *debiasing techniques* [73] which are useful for tackling focusing illusion. BCS would study how these could be integrated in the designed system, in the sense that the UI or the security protocol could implement features meant to manipulate the User in such a way that she would be prepared for a possible attack; or better, in such a way that they alert the user to the security aspects. Such features could involve: recollections, so that the same aspects of t_1 (now) are as in t_0 (the time point when the User has probably been trained to use the system).

Modelling frameworks of use for BCS exist in various maturity degrees, and BCS research can consider these as a starting point. One line of research follows the so called *Actor Network Procedures* (ANP) introduced in [74] and the alternative works of [75, 76]. These have the intention of modelling security ceremonies [7, 8] by drawing inspiration from the Actor Network Theory in sociology, where the book [77] gives a good overview. ANP is meant to be a graphical modelling language, yet with formal logical reasoning underpinnings. For BCS modelling we encourage the use of graphical languages, where prime examples that are also formally grounded include the statecharts [78, 79] and the live sequence charts [80], which were intended for describing concurrent and reactive systems.

Another line of work comes from human factors engineering [81, 82] that use task analytic models [83–87] to capture the human behaviour involved in (often for controlling) the system. In this direction, efforts into modelling human behaviour combined with models of the system have been used for doing model checking in [88, 89]. Modelling

has also been related to the development [90] and evaluation [91] of user interfaces. In particular, the area of modelling user experience is largely unexplored [92].

Conclusion

We have argued that concepts and findings from behavioural sciences can be translated into models useful for computer science. Such models could be used for analysing the BCS-systems using techniques such as automated model checking [57]. Moreover, behavioural models and related modelling languages can be used by system developers when making new BCS-systems to also consider the human interacting with the system. We can already see promising results in this direction from using formal methods to analyse HAI systems [38] or human related security breaches [10].

As we have shown, there is now an abundance of research arguing that people behave and act in other ways than those assumed under the rationality paradigm. In consequence we have proposed that computer science incorporates in the design of systems that interact with humans knowledge about actual behaviour. Psychology and Behavioural Sciences have by now provided a large amount of empirical evidence showing how human behavioural tendencies in many instances depart from strict rational assumptions.

Knowledge about human behaviour and tendencies have often been employed by private interests for commercial purposes, for example with the aim of influencing or convincing people to purchase a specific product or service [93]. Such knowledge has also been employed by political interests in order to convince people to endorse a particular political view. Even if some would argue that such approaches may have been employed in people's own interests, in order to have people make choices that are actually good for them, others would argue that the main purpose of such approaches is to serve the initiator—whether this is a private company interest or a political agenda.

Thus, some may fear that if, by following our proposal, models of human behaviour are made such that machines can work with them, then more easy it can be for a totalitarian regime to control people by using computers for mass-surveillance/-manipulation. One could even fear such models also in a democratic society because large corporations that control information, like those involved in search engines, social media, network corporations, or device and software producers, could be tempted to use such models in a negative manner, trying to gain control over their users. Indeed, any large corporation could possibly gain access to data from the previously mentioned providers and use behaviour models to extort their users, but this is a matter that should be regulatory controlled in a context of legislation.

Contrary to such possible uses of *behavioural-machine models* (BCS-models), our main intention with the BCS proposal is to take the perspective of the individual when building HAI-interfaces that take into consideration human behavioural tendencies (in the interest of humans instead of political or commercial interests) with the aim of designing systems that empower individuals to make more correct judgements when interacting with an automated system.

Our proposal would serve the individual and thereby the society, as well as system designers and owners, due an increased knowledge about human behavioural tendencies. Additionally, when taking into consideration that society and all its functions rely

both on situational awareness and on resilience –in infrastructure as well as in individual citizens– in order to maintain structure and daily life in all societal domains, it is imperative that disruptions of vital societal functions can be avoided or countered [94]. It is thus not an overstatement to claim that the new conceptual domain of Behavioural Computer Science would also have valuable and positive implications for institutions such as National Security Agencies (seeking to secure society and to make society resilient) and for National Intelligence Agencies (seeking to identify threats to society and prompt situational awareness). Future research would explore implications of the concept of Behavioural Computer Science in contexts where individuals and systems interact, ranging from smart living and IoT to national security and intelligence domains.

Authors' contributions

Authors contributed in various important aspects. Tore Pedersen has driven the long process that this work went through. All authors read and approved the final manuscript.

Authors' information

Tore Pedersen is Associate Professor, Dean of Academics, and Director of Center for Intelligence Studies at the Norwegian Defence Intelligence School (NORDIS). He is also affiliated Associate Professor at the Department of Psychology, Bjørknes University College. Tore Pedersen received his PhD from Karlstad University and has been working since in various aspects of psychological science. He is currently engaged in empirical research on cognitive and methodological aspects of the National Intelligence and Security domain.

Christian Johansen is a researcher in the Precise Modelling and Analysis group of the University of Oslo (UiO) and member of the ConSeRNS interdisciplinary group "Concurrent Security and Resilience for Networked Systems". He received PhD in 2010 from UiO and has worked since in different areas of computer science, including modelling of security protocols; modelling languages and verification of complex systems; models and tools for parallel programming and concurrent systems; (legal) electronic contracts. His work, more than 30 international articles, has been published in highly ranked conferences s.a. CONCUR, ATVA, or LFCS, and top journals s.a. JLAP.

Audun Jøsang joined Oslo University as Professor in 2008. Prior to that he was Associate Professor at QUT, research leader of Security at DSTC in Brisbane, worked in the telecommunications industry for Alcatel in Belgium and for Telenor in Norway. He was also Associate Professor at the Norwegian University of Science and Technology (NTNU). He has a Master's in Information Security from Royal Holloway College, University of London, and a PhD from NTNU in Norway. He is a CISSP and CISM with broad knowledge and experience in information security.

Author details

¹ Center for Intelligence Studies, Norwegian Defence Intelligence School, Oslo, Norway. ² Department of Psychology, Bjørknes University College, Oslo, Norway. ³ Department of Informatics, University of Oslo, P.O. Box 1080, 0316 Oslo, Blindern, Norway.

Acknowledgements

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

Tore Pedersen and Audun Jøsang were partially supported by the project Oslo Analytics funded by the IKTPLUSS program of the Norwegian Research Council. Christian Johansen was partially supported by the project IoTSec – Security in IoT for Smart Grids, with Number 248113/O70 part of the IKTPLUSS program funded by the [Norwegian Research Council](#).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 15 February 2018 Accepted: 9 March 2018

Published online: 23 March 2018

References

- Poole D, Mackworth A (2010) *Artificial Intelligence: foundations of computational agents*. Cambridge University Press, Cambridge
- Rogers Y, Sharp H, Preece J (2011) *Interaction Design: beyond Human–Computer interaction*, 3rd edn. Wiley, New York
- Jøsang A, Ismail R, Boyd C (2007) A survey of trust and reputation systems for online service provision. *Decis Support Syst* 43(2):618–644. <https://doi.org/10.1016/j.dss.2005.05.019>
- Pinyol I, Sabater-Mir J (2013) Computational trust and reputation models for open multi-agent systems: a review. *Artif Intell Rev* 40(1):1–25. <https://doi.org/10.1007/s10462-011-9277-z>
- Veruggio G, Operto F (2008) Roboethics: social and ethical implications of robotics. In: Siciliano B, Khatib O (eds) *Handbook of Robotics*. Springer, Berlin, pp 1499–1524. https://doi.org/10.1007/978-3-540-30301-5_65
- Tzafestas SG (2016) *Roboethics: A Navigating Overview*. Intelligent systems, control and automation: science and engineering, vol 79. Springer, Berlin. <https://doi.org/10.1007/978-3-319-21714-7>
- Ellison C (2007) Ceremony design and analysis. *Cryptol ePrint Archiv Rep* 2007:399
- Radke K, Boyd C, Nieto JMG, Brereton M (2011) Ceremony analysis: strengths and weaknesses. *Inform. Security and Privacy. IFIP AICT*, vol 354. Springer, Berlin, pp 104–115
- Yee KP (2002) User interaction design for secure systems. *Information and communications security. LNCS*, vol 2513. Springer, Berlin, pp 278–290. https://doi.org/10.1007/3-540-36159-6_24
- Ruksenas R, Curzon P, Blandford A (2008) Modelling and analysing cognitive causes of security breaches. *Innov Syst Softw Eng* 4(2):143–160
- Adams A, Sasse MA (1999) Users are not the enemy. *Commun ACM* 42(12):40–46
- West R (2008) The psychology of security. *Commun ACM* 51(4):34–40
- Lee JK, Moon SY, Park JH (2016) Hb-dipm: human behavior analysis-based malware detection and intrusion prevention model in the future internet. *J Inf Process Syst* 12(3):489–501
- Hong J (2012) The state of phishing attacks. *Commun ACM* 55(1):74–81. <https://doi.org/10.1145/2063176.2063197>
- Fette I, Sadeh N, Tomasic A (2007) Learning to detect phishing emails. In: *Proceedings of the 16th international conference on World Wide Web. WWW '07*. ACM, New York, pp 649–656. <https://doi.org/10.1145/1242572.1242660>
- Abu-Nimeh S, Nappa D, Wang X, Nair S (2007) A comparison of machine learning techniques for phishing detection. In: *Proceedings of the anti-phishing working groups 2Nd annual eCrime researchers summit. eCrime '07*. ACM, pp 60–69. <https://doi.org/10.1145/1299015.1299021>
- Egelman S, Cranor LF, Hong J (2008) You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In: *Proceedings of the SIGCHI conference on human factors in computing systems. CHI '08*. ACM, New York, pp 1065–1074. <https://doi.org/10.1145/1357054.1357219>
- Kumaraguru P, Sheng S, Acquisti A, Cranor LF, Hong J (2010) Teaching johnny not to fall for phish. *ACM Trans Internet Technol* 10(2):7–1731. <https://doi.org/10.1145/1754393.1754396>
- Augusto JC, Huch M, Kameas A, Maitland J, McCullagh P, Roberts J, Sixsmith A, Wichert R (2012) *Handbook of Ambient Assisted Living*. IOS Press, Amsterdam
- Brush, A.J.B., Lee, B., Mahajan, R., Agarwal, S., Saroiu, S., Dixon, C.: Home automation in the wild: Challenges and opportunities. In: *SIGCHI*, pp. 2115–2124. ACM, (2011). <https://doi.org/10.1145/1978942.1979249>
- Johansen C, Pedersen T, Jøsang A (2016) Towards behavioural computer science. In: *10th IFIP WG 11.11 international conference on trust management (IFIPTM)*. IFIP AICT advances in information and communication technology, vol. 473. Springer, Berlin, pp 154–163. https://doi.org/10.1007/978-3-319-41354-9_12
- Simon HA (1979) Rational decision making in business organizations. *Am Econ Rev* 69(4):493–513
- Simon HA (1983) *Reason in human affairs*. Stanford University Press, Palo Alto
- Simon HA (1997) *Models of bounded rationality: empirically grounded economic reason*. MIT Press, Cambridge
- Thaler RH, Sunstein CR (2008) *Nudge: improving decisions about health, wealth, and happiness*. Yale University Press, New Haven
- Kahneman D, Thaler RH (2006) Anomalies: utility maximization and experienced utility. *J Econ Persp* 20(1):221–234. <https://doi.org/10.1257/089533006776526076>
- Gilovich T, Griffin D, Kahneman D (2002) *Heuristics and Biases: the psychology of intuitive judgment*. Cambridge University Press, Cambridge
- Tversky A, Kahneman D (1974) Judgment under uncertainty: Heuristics and biases. *Science* 185(4157):1124–1131
- Kahneman D (2003) A perspective on judgment and choice: mapping bounded rationality. *Am Psychol* 58:697–720
- (2003) *Behavioral game theory: experiments in strategic interaction*. Princeton University Press, Princeton
- Gärling T, Ettema D, Friman M (2014) *Handbook of sustainable travel*. Springer, Berlin
- Pedersen T, Friman M, Kristensson P (2011) Affective forecasting: predicting and experiencing satisfaction with public transportation. *J Appl Soc Psychol* 41(8):1926–1946. <https://doi.org/10.1111/j.1559-1816.2011.00789.x>
- Bella G, Coles-Kemp L (2012) Layered analysis of security ceremonies. *Information security and privacy. IFIP AICT*, vol 376. Springer, Berlin, pp 273–286
- Dolev D, Yao AC (1983) On the security of public key protocols. *IEEE Trans Inf Theory* 29(2):198–207. <https://doi.org/10.1109/TIT.1983.1056650>
- Abadi M (2001) Mobile values, new names, and secure communication. In: Hankin C, Schmidt D et al (eds) *POPL*. ACM, New York, pp 104–115
- Abadi M, Blanchet B, Fournet C (2016) The applied pi calculus: mobile values, new names, and secure communication. *J ACM* 65:1
- Semančík R (2007) Basic properties of the persona model. *Comput Inf* 26(2):105–121
- Bolton M, Bass E, Siminiceanu R (2013) Using formal verification to evaluate human-automation interaction: a review. *IEEE Trans Syst Man Cyb* 43(3):488–503. <https://doi.org/10.1109/TSMCA.2012.2210406>
- Bevan N (2001) International standards for HCI and usability. *Int J Hum Comput Studies* 55(4):533–552. <https://doi.org/10.1006/ijhc.2001.0483>

40. Johansen C, Jøsang A (2014) Probabilistic modeling of humans in security ceremonies. In: Aldini A, Martinelli F, Suri N (eds) Workshop on quantitative aspects in security assurance (QASA). LNCS, vol 8872. Springer, Berlin, pp 277–292. https://doi.org/10.1007/978-3-319-17016-9_18
41. Newell A, Card SK (1985) The prospects for psychological science in Human–Computer interaction. *Hum Comput Interact* 1(3):209–242. https://doi.org/10.1207/s15327051hci0103_1
42. Harrison M, Thimbleby H (eds) (1990) Formal methods in Human–Computer interaction. Cambridge Univ Press, Cambridge
43. Carroll JM (2003) HCI models, theories, and frameworks: toward a multidisciplinary science. Morgan Kaufmann. Elsevier, Amsterdam
44. Oliver RL (2010) Satisfaction: a behavioral perspective on the consumer. M.E. Sharpe, New York
45. Kahneman D (2000) Evaluation by moments, past and future. In: Kahneman D, Tversky a (eds) Choices, values and frames. Cambridge University Press, Cambridge, p 693
46. Wilson TD, Gilbert DT (2003) Affective forecasting. *Advances in experimental social psychology*, vol 35. Academic Press, Cambridge, pp 345–411. [https://doi.org/10.1016/S0065-2601\(03\)01006-2](https://doi.org/10.1016/S0065-2601(03)01006-2)
47. Slovic SA (2002) Two systems of reasoning. In: Gilovich T, Griffin D, Kahneman D (eds) Heuristics and biases. Cambridge University Press, Cambridge, pp 379–396
48. Cialdini RB (2001) The science of persuasion. *Scientific American*. Springer, Berlin
49. Butterworth R, Blandford A, Duke DJ (2000) Demonstrating the cognitive plausibility of interactive system specifications. *Formal Asp Comput* 12(4):237–259. <https://doi.org/10.1007/s001650070021>
50. Curzon P, Ruksenas R, Blandford A (2007) An approach to formal verification of Human–Computer interaction. *Formal Asp Comput* 19(4):513–550. <https://doi.org/10.1007/s00165-007-0035-6>
51. Ruksenas R, Back J, Curzon P, Blandford A (2009) Verification-guided modelling of salience and cognitive load. *Formal Asp Comput* 21(6):541–569. <https://doi.org/10.1007/s00165-008-0102-7>
52. Newell A (1990) Unified theories of cognition. Harvard University Press, Cambridge
53. Hollnagel E, Woods DD (2005) Joint cognitive systems: foundations of cognitive systems engineering. CRC Press, Boca Raton
54. Buehler R, McFarland C (2001) Intensity bias in affective forecasting: the role of temporal focus. *Personal Soc Psychol Bull* 27(11):1480–1493
55. Manna Z, Pnueli A (1992) The temporal logic of reactive and concurrent systems: specification. Springer, Berlin
56. Stirling C (2001) Modal and temporal properties of processes. Springer, Berlin
57. Baier C, Katoen JP (2008) Principles of model checking. MIT Press, Cambridge
58. Robinson MD, Clore GL (2002) Episodic and semantic knowledge in emotional self-report. *J Personal Soc Psychol* 83(1):198–215
59. Harel D, Tiuryn J, Kozen D (2000) Dynamic logic. MIT Press, Cambridge
60. Beckert B, Hähnle R, Schmitt PH (2007) Verification of object-oriented software: the keY approach. Springer, Berlin
61. Conway MA (1997) Cognitive models of memory. MIT Press, Cambridge
62. Neath I, Surprenant AM (2003) Human memory: an introduction to research, data, and theory, 2nd edn. Thomson/Wadsworth, Belmont
63. Byrne MD, Bovair S (1997) A working memory model of a common procedural error. *Cogn Sci* 21(1):31–61. https://doi.org/10.1207/s15516709cog2101_2
64. Oberauer K, Kliegl R (2006) A formal model of capacity limits in working memory. *J Memory Lang* 55(4):601–626. <https://doi.org/10.1016/j.jml.2006.08.009>
65. Wang Y (2013) Cognitive informatics for revealing human cognition: knowledge manipulations in natural intelligence. IGI Global, Hershey
66. Fagin R, Halpern JY, Moses Y, Vardi MY (1995) Reasoning about knowledge. MIT Press, Cambridge
67. Baltag A, Moss LS (2004) Logics for epistemic programs. *Synthese* 139(2):165–224
68. Wilson TD, Wheatley T, Meyers JM, Gilbert DT, Axsom D (2000) Focalism: a source of durability bias in affective forecasting. *J Person Soc Psychol* 78(5):821–836
69. Droste M, Kuich W, Vogler H (2009) Handbook of weighted automata. Springer, Berlin
70. Kwiatkowska MZ (2007) Quantitative verification: models techniques and tools. In: Foundations of software engineering ACM SIGSOFT. ACM, pp 449–458. <https://doi.org/10.1145/1287624.1287688>
71. Ruksenas R, Curzon P, Back J, Blandford A (2007) Formal Modelling of Cognitive Interpretation. In: 13th International workshop on design, specification, and verification of interactive systems (DSVIS'06). LNCS, vol. 4323. Springer, Berlin, pp 123–136
72. Jøsang A (2001) A logic for uncertain probabilities. *Int J Uncertain Fuzziness Knowl Based Syst* 9(3):279–311. <https://doi.org/10.1142/S0218488501000831>
73. Pedersen T, Kristensson P, Friman M (2012) Counteracting the focusing illusion: effects of defocusing on car users predicted satisfaction with public transport. *J Environ Psychol* 32(1):30–36. <https://doi.org/10.1016/j.jenvp.2011.10.004>
74. Pavlovic D, Meadows C (2011) Actor-network procedures: modeling multi-factor authentication, device pairing, social interactions. [arXiv.org](https://arxiv.org)
75. Pieters W (2011) Representing humans in system security models: an actor-network approach. *J Wirel Mobile Netw Ubiquitous Comput Dependable Appl* 2(1):75–92
76. Prisacariu C (2014) Actor network procedures as psi-calculi for security ceremonies. In: International workshop on graphical models for security. Electronic proceedings in theoretical computer science, vol. 148. Open Publishing Association., pp 63–77. <https://doi.org/10.4204/EPTCS.148.5>
77. Latour B (2005) Reassembling the social—an introduction to actor-network-theory. Oxford Univ Press, Oxford
78. Harel D (1987) Statecharts: a visual formalism for complex systems. *Sci Comput Program* 8(3):231–274
79. Harel D, Naamad A (1996) The statemate semantics of statecharts. *ACM TOSEM* 5(4):293–333
80. Harel D, Marely R (2003) Come, Let's play: scenario-based programming using LSCs and the play-engine, vol 1. Springer, Berlin

81. Wickens CD, Hollands JG, Banbury S, Parasuraman R (2015) *Engineering psychology and human performance*, 4th edn. Taylor & Francis, Routledge
82. Salvendy G (2012) *Handbook of human factors and ergonomics*. Wiley, New Jersey
83. Kirwan B, Ainsworth LK (1992) *A guide to task analysis: the task analysis working group*. Taylor & Francis, Routledge
84. Crandall B, Klein GA, Hoffman RR (2006) *Working minds: a practitioner's guide to cognitive task analysis*. MIT Press, Cambridge
85. Paterno F, Mancini C, Meniconi S (1997) In: Howard S, Hammond J, Lindgaard G (eds) *ConcurTask-Trees: a diagrammatic notation for specifying task models*. Springer, Boston, pp 362–369. https://doi.org/10.1007/978-0-387-35175-9_58
86. Nóbrega L, Nunes NJ, Coelho H (2006) Mapping concurtasktrees into uml 2.0. In: Gilroy SW, Harrison MD (eds) *Interactive systems. Design, specification, and verification*. Springer, Berlin, pp 237–248
87. Mitchell CM, Miller RA (1986) A discrete control model of operator function: a methodology for information display design. *IEEE Trans Syst Man Cybernet* 16(3):343–357
88. Bolton ML, Siminiceanu RI, Bass EJ (2011) A systematic approach to model checking Human–Automation interaction using task analytic models. *IEEE Trans Syst Man Cybern Part A Syst Hum* 41(5):961–976. <https://doi.org/10.1109/TSMCA.2011.2109709>
89. Bolton ML, Bass EJ (2010) Formally verifying human-automation interaction as part of a system model: limitations and tradeoffs. *Innov Syst Softw Eng* 6(3):219–231
90. Pérez-Medina J-L, Dupuy-Chessa S, Front A (2007) A survey of model driven engineering tools for user interface design. In: Winckler M, Johnson H, Palanque P (eds) *Task models and diagrams for user interface design*. Springer, Berlin, pp 84–97
91. Ivory MY, Hearst MA (2001) The state of the art in automating usability evaluation of user interfaces. *ACM Comput Surv* 33(4):470–516
92. Law EL-C, Van Schaik P (2010) Modelling user experience-an agenda for research and practice. *Interact comput* 22(5):313–322
93. Turner JC (1991) *Social influence*. Thomson Brooks/Cole Publishing Co, Pacific Grove
94. Omand D (2012) *Securing the State*. C Hurst Publishers, London

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com
