


# Can Merging a Capability Approach with Effectual Processes Help Us Define a Permissible Action Range for AI Robotics Entrepreneurship?

Yuko Kamishima<sup>1</sup> · Bart Gremmen<sup>2</sup> ·  
Hikari Akizawa<sup>3</sup> 

Published online: 11 July 2017

© The Author(s) 2017. This article is an open access publication

**Abstract** In this paper, we first enumerate the problems that humans might face with a new type of technology such as robots with artificial intelligence (AI robots). Robotics entrepreneurs are calling for discussions about goals and values because AI robots, which are potentially more intelligent than humans, can no longer be fully understood and controlled by humans. AI robots could even develop into ethically “bad” agents and become very harmful. We consider these discussions as part of a process of developing responsible innovations in AI robotics in order to prevent catastrophic risks on a global scale. To deal with these issues, we propose the *capability-effectual approach*, drawing on two bodies of research: the capability approach from ethics, and the effectual process model from entrepreneurship research. The capability approach provides central human capabilities, guiding the effectual process through individual goals and aspirations in the collaborative design process of stakeholders. More precisely, by assuming and understanding correspondences between goals, purposes, desires, and aspirations in the languages of different disciplines, the capability-effectual approach clarifies both how a capability list working globally could affect the aspirations and end-goals of individuals, and how local aspirations and end-goals could either energise or limit effectual processes. Theoretically, the capability-effectual approach links the collaboration of stakeholders and the design process in responsible innovation

---

✉ Bart Gremmen  
bart.gremmen@wur.nl

Yuko Kamishima  
ykr09182@fc.ritsumei.ac.jp

Hikari Akizawa  
hikariakizawa@gmail.com

<sup>1</sup> Ritsumeikan University, Kyoto, Japan

<sup>2</sup> Wageningen University and Research, Wageningen, Netherlands

<sup>3</sup> Oikos Research, LLC, Tokyo, Japan

research. Practically, this approach could potentially contribute to the robust development of AI robots by providing robotics entrepreneurs with a tool for establishing a permissible action range within which to develop AI robotics.

**Keywords** Responsible innovation · Capability approach · Effectuation · Robotics · AI · Entrepreneurship

The recent rapid innovations in robotics have given rise to the development of Artificial Intelligence (AI) robots: a totally new kind of fully autonomous robot equipped with the latest developments in artificial intelligence. The effects of these so-called AI robots on human lives will exceed those that have resulted from other technologies because AI robots will be more intelligent than humans. This means that we will not be able to fully understand them, and, consequently may not be able to control them. Instead, we may run the risk that they will control us and indeed that they could even develop into ethically “bad” agents becoming very harmful for humans and their environment. From a business perspective, global competition for research into AI robots and the development of AI robots is increasing rapidly, making it difficult for us to pause and consider how to govern AI robots in the future. How should we deal with them? Should we let AI robots steer our lives in whatever direction they will? Or should we develop AI robots that think like a human engineer concerned about ethics? We have reformulated this last question into the main objective of our paper: to define a permissible action range for AI robotics entrepreneurship.

These imminent issues can be framed from the perspective of responsible innovation. Understanding the societal, legal and ethical issues related to the production and implementation of this technology will be essential if we are to ensure that it is firmly and fully embedded in society. Unless we learn how to embed AI robots in society, and to understand why a number of previous innovations have failed to do so, we run the risk that new AI innovations in robotics will generate further controversy, misunderstanding, and polemic (Stilgoe et al. 2013).

In the literature on responsible innovation, we find two main approaches: the first approach focuses on the collaboration process of stakeholders and discusses transparency, interaction, responsiveness, and co-responsibility (Blok et al. 2015; Von Schomberg 2013); and the second deals with the design process and moral issues (Van den Hove 2013). Since we want to define a permissible action range for AI robotics entrepreneurship, we need to combine both approaches. As a first step, we will link the design process to an ethical theory (the capability approach), and, then we will link the collaboration approach to a model of the logic of entrepreneurship—the effectual process model. As a third step, we will merge these into the capability-effectual approach. We propose this multi-disciplinary approach in order to define a permissible action range for AI robotics entrepreneurship.

This article is structured as follows: first, we briefly examine a central problematic of robotics, that is whether we should control this innovative technology by assigning purposes and ethics for responsible innovation; we then introduce the capability approach arguing that society will be minimally good if we share values concerning central human capabilities; we then explain the effectual process by which stakeholders collaboratively formulate new purposes and plans, and distinguish it from the causation process, which is controlling or optimising a plan with a purpose. Robotics is gradually permeating those two human processes through autonomous learning. Finally, we argue that central human capabilities and the

effectual process can influence each other. We conclude that the notion of central human capabilities not only reveals the permissible action range of robotics entrepreneurship, but also positively stimulates it.

## Robotics and Responsible Innovation

Historically, humans have invented powerful technologies such as fire, iron artifacts, steam engines, and computers, which have dramatically changed human life. Through the efforts of entrepreneurs, robotics will likewise have a huge impact on our lives. Self-driving cars can safely move us from place to place thereby eliminating our own dangerous driving. Robot caregivers can make caregiving more comfortable for both care recipients and caregivers. Soon, artificial intelligence will produce numerous hypotheses for biotech researchers. Unfortunately, however, the military will probably be able to hire robot soldiers for their dangerous duties before we eliminate all nuclear weapons. Technological innovations have always been a double-edged sword. What makes AI robots different is their distinctive features of autonomous action and swift evolution. In this section, we introduce and examine AI robotics as a case study for our capability-effectual thesis; outlining the basics of robotics and summarising the emergence of robotics.

### What is a Robot?

For clarity, we begin by briefly reviewing the basics of robots and robotics, i.e., the study of robots however it is beyond the scope of this paper to thoroughly review robotics given the constant and rapid developments in that field.

Rich Mahoney, then director of robotics at SRI International, explained the design of robots in a magazine interview (Anonymous 2015). A robot consists of four parts: the human interface, thought, sensory ability, and mechanics (e.g., actuators). The robotics field requires multiple disciplines: chemistry, physics, material science, cognitive science, and a wide range of engineering fields. At present, robots develop when one of the salient component technologies can be made in a cost-effective way.

An introductory textbook on robotics (Mataric 2007) defines a robot as “an autonomous system which exists in the physical world, can sense its environment, and can act on it to achieve some goals” (p. 3). The goals can be simple (“Don’t get stuck”) or complex (“Do whatever it takes to keep your owner safe”). The rapidly developing field of artificial intelligence is used in robotics to develop the thinking part of robots. Artificial intelligence controls robots mainly by goal achievement; it can optimally plan how a robot should act to achieve a desired goal and can enable the acquisition of knowledge and skills (p. 255). Among the many learning methods, two are particularly popular: *Reinforcement Learning* (the trial-and-error method) and *Neural Network Learning* (the external teacher method). Even though robot design is inspired by biology, robots are *not* models of natural systems, such as human brains or circus animals.

More recently, *Deep Learning*, or Representational Learning, has appeared on the AI horizon due to a research paper from the University of Toronto in 2006 (Matsuo 2015). In 2015, AI installed in computers had already outperformed humans in terms of accuracy and speed with regard to facial recognition. Yutaka Matsuo, an AI researcher, describes Deep Learning as a breakthrough technology for this half century. This means that, even if humans

and AI robots share the same concept (e.g., a cat), we are unable to understand how AI robots *think*. AI robots might acquire knowledge quite differently because their learning process is not the same as that of the human brain, and they can also use quite different information for example ultra-violet rays and supersonic waves (Matsuo 2015, Chapter 6, Section 2).

### Critical Questions about AI Posed by Robotics Researchers

The AI issue only recently became widely recognised in Japan when the media rushed to report warnings from well-known figures, such as Stephen Hawking and Elon Musk, who signed an open letter (Future of Life Institute n.d.). In Europe, the Global Challenge Foundation (GCF) (GCF 2016) reported that disruption from AI is one of the major “global catastrophic risks” that should especially command our attention, along with nuclear war, natural pandemics, engineered pandemics from biotechnology, catastrophic climate change, and failures of geo-engineering. A global catastrophic risk is defined as “a possible event or process that, were it to occur, would end the lives of approximately 10% or more of the global population, or do comparable damage” (GCF 2016, p. 22). For comparison, the Spanish influenza pandemic of 1918–1920 killed between 2.5% and 5% of the world’s population. The report advocates the importance of research on “how to give AI systems desirable goals” (GCF 2016, p. 58) in order to control their speedy and autonomous development. We look at this risk more precisely in Dhar (2016) and Bostrom (2014).

Dhar (2016) reported on a symposium entitled “The Future of Artificial Intelligence,” which New York University hosted in January 2016. He summarised the symposium’s discussions among AI researchers from academia and industry by posing five fundamental questions:

1. Why is it different this time? The history of AI has seen several “boom-bust” cycles, where the optimism was driven by some perceived significant advance, followed by disappointment due to unrealistic expectations. We are currently in another boom. Is it different this time? If so, why?
2. How should we control systems that are potentially more intelligent than humans, whose working we don’t fully understand? A version of this question that is more pressing is how should we control systems that are extremely complex and potentially more accurate than humans but can’t directly explain their own behavior?
3. Should there be an objective function for AI systems or is diversity more appropriate?
4. Are we likely to see ourselves replaced by robots for most tasks or augmented by machine intelligence? In the process, will AI create more jobs than it will destroy, or the other way around?
5. Is our current regulatory framework for governing the rights and actions of humans adequate for dealing with robots? (p. 5)

In this paper, we focus on the first, second, and third questions because they allude to unprecedented problems related to this technology. The latter two are also important but somewhat familiar in the history of innovations (e.g., the steam engine and Luddite movement). The first question helps us to grasp the novelty of current AI developments in comparison to past booms. AI is continuously developing, and Dhar’s point is that the field of AI has entered the next stage of *deep learning* and *big data* (Dhar 2016). AI systems can learn—i.e., reason, plan, explain, and learn to learn—by directly interacting with the world

through big data that is generated rapidly and massively from various sources, such as social media and video surveillance cameras, among others. Before the era of big data, humans could handle and input only so much data, which represented limited and fragmented parts of the world.

The second question suggests that because we do not understand the reasoning of AI systems, we should consider how we can control them. AI systems are now able to make scientific discoveries by creating hypotheses (Dhar 2016, p. 6). For example, they can discover new knowledge—predictions and models—by *reviewing* numerous papers in biotechnology fields. Although AI researchers are inspired and influenced by the human brain, AI is not an exact model of it, as was mentioned in the previous section. Even if humans provide a clear purpose for AI systems, we will still have to contend with their unforeseen behaviours. Thus, Dhar argues that we need to address the “ethics of inference” (p. 6).

The third question indicates the difficulty of defining optimal reasoning or morality. We cannot escape this fundamental problem. Researchers in this field used to design AI systems by focusing on optimising their objective functions, assuming that these functions are “exogenously specified—that it is someone else’s job to get it right” (Dhar 2016, p. 7).

Nick Bostrom (2014), who has a background in both philosophy and computational neuroscience, regards the risks and problems surrounding the AI issue as being a consequence of what he calls “superintelligence”—the explosion of the intelligent capacities of AI systems due to their rapid evolution. AI systems can plan some actions by themselves, and they do not necessarily need to follow external directions. Bostrom has questioned how or if we can control such intelligence. He has proposed several methods of control to prevent undesirable outcomes, classifying them into two categories: capability control and motivation selection. Whereas the former method aims to control what AI systems *can* do, the latter aims to control what it *wants* to do by engineering the motivation systems and their final goals. For example, one of the latter methods, the specification of normativity, tries to set up a system that learns *appropriate* values from indirectly formulated criterion.

## Responsible Innovation

Humans face an unusual situation as the field of AI develops to the next stage: AI systems can discover knowledge by deep learning and interact with the world through their ability to process “big data” that is vast amounts of information. It is certain that AI systems have begun to evolve swiftly and endlessly. Even if we know that their answers are *right*, we can neither understand their reasoning nor follow their thinking speed. AI researchers are proposing potential methods for controlling AI systems by limiting their abilities or making them select their motivations. This means, unfortunately, that they have not discovered how humans can control novel AI systems.

We believe that AI researchers have three problems: (1) they tend to separate AI research from robotics research; (2) they do not consider other social factors, including entrepreneurship; and (3) they tend to deal with problems through control.

Regarding the first problem, we propose that the AI issue is considered as part of the robotics issue. We understand the importance of studying the field of AI separately, and that AI systems can be insulated from robotics (i.e., avoiding interaction in the physical world by actuators), however, we also recognise that the current advancement of the field of AI at present will lead the thinking part of the development of a new kind of robots. Robotics is under pressure from a practical standpoint to use AI systems as the thinking part of robots in

order to add economic value. We already have the example of Google's self-driving car, which has stimulated interest in robotics across other industries worldwide. Thus, in our study, we will look at robotics issues overall rather than just AI issues, since we posit that the field of AI affects human life more spontaneously and uncontrollably via robots.

Regarding the second problem, we recognise that robotics issues are thoroughly enmeshed with social factors. Robotics has already become part of the economic and political strategies of many countries and regions. For example, robotics will be indispensable in helping humans with caregiving and in saving lives in situations such as natural disasters and nuclear plant failures. Global competition in robotics entrepreneurship is already apparent. In a small but symbolic newspaper article (Ogawa 2014), a director of DARPA (the Defense Advanced Research Projects Agency in the U.S. Department of Defense) stated that the best way to defend against *technological surprises* is to develop our own innovations; and, for this, robotics is one area of focus.

The third problem is that AI researchers tend to think that the best way to defend humans is to control robots. However, robots with AI systems can plan their own actions for optimising certain goals, and, being equipped with deep learning, they can even formulate new plans. We believe that humans need to find different ways to survive other than through controlling or optimising robots. A prevailing optimistic idea posits that humans have innate reward functions (values), which can be used in robotics (Dhar 2016; Bostrom 2014; Matsuo 2015). However, given that, so far, we have been unable to avoid waging war, it is difficult to determine how well-tuned human functions really are. Further, we know that many species have disappeared from the earth. Ultimately, robots may evolve to a different species from humans and could thus threaten humans with extinction.

Blok and Lemmens (2015) warn of uncritically setting goals for the grand challenges for responsible innovation. When we try to control AI robots through their abilities, data, goals, or values, these items may contain the biases of those who designed them. Even if we can control robots, we cannot avoid unintended social consequences. The *Collingridge Dilemma* (Blok and Lemmens 2015, p. 25) frequently emerges - we cannot return to the past when we notice mistakes. Moreover, the current speed of AI development is unprecedented. Following Bostrom and Yudkowsky (2011), we posit that constructing a trustworthy AI robot will require that the robot thinks like a human engineer concerned about ethics, and is not merely a product of ethical engineering. They conclude that the discipline of AI ethics is likely to differ fundamentally from the ethical discipline of noncognitive technologies, in that:

- The local, specific behavior of the AI may not be predictable apart from its safety, even if the programmers do everything right;
- Verifying the safety of the system becomes a greater challenge because we must verify what the system is trying to do, rather than being able to verify the system's safe behavior in all operating contexts;
- Ethical cognition itself must be taken as a subject matter of engineering (Bostrom and Yudkowsky 2011, p. 5).

In the following sections, we will combine the two categories, capability and motivation, that Bostrom (2014) has proposed as control methods to prevent undesirable outcomes. This represents an alternative to the "control approach" for dealing with the problems of AI robots in that it links the capability approach to the effectual process model, resulting in the capability-effectual framework. The capability approach offers a perspective on how we can look at and deal with the various goals and values in global societies, while the effectual

process model offers a perspective on collaborative design processes that stakeholders might employ in robotics entrepreneurship.

## The Capability Approach and Human Well-Being

We consider the capability approach as a promising way to determine what kinds of goals and purposes are permissible for entrepreneurs to assign to AI robots.

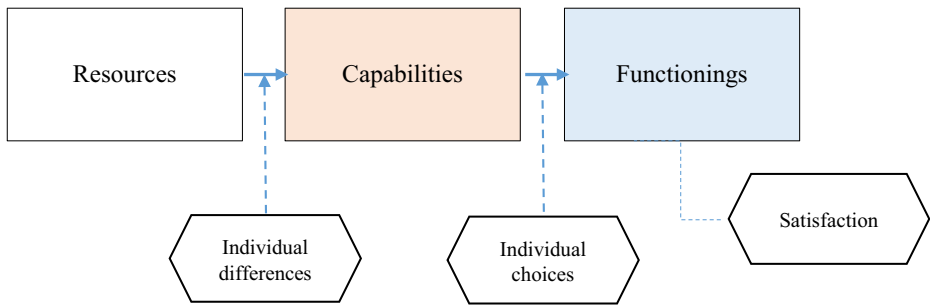
### A Brief Overview of the Capability Approach

Since the late 1980s, Indian-born economist Amartya Sen and American philosopher Martha Nussbaum have forged what is known as the “capability approach.” This theory seeks to objectively evaluate a person’s well-being. It assesses what individuals can do and what they can be. Weights are assigned based on a person’s real freedom (capabilities), not on their actual achievements (functioning). Hence, human development should mean the enlargement of human capabilities, not the expansion of their resources or the excellence of their functioning, and also an increase in their satisfaction, as shown in Fig. 1. Various types of freedoms, including social, economic, and political freedoms, are indispensable.

Nussbaum’s capability approach differs somewhat from Sen’s. Based on her adherence to New Aristotelianism, Nussbaum has identified ten capabilities that are prerequisites for a good human life. According to her, this list delivers “a bare minimum of what respect for human dignity requires” (Nussbaum 2000, p. 5). She terms these ten capabilities as “central human capabilities” that set a minimum limit or threshold level for a good life. They are: 1) capabilities for life, 2) bodily health, 3) bodily integrity, 4) senses/ imagination/ thought, 5) emotions, 6) practical reasoning, 7) affiliation, 8) other species, 9) play, and 10) control over one’s environment (Nussbaum 2000, p. 78–80).<sup>1</sup>

Nussbaum’s capability approach focuses on humans generally, so is her approach insensitive to differences among individuals? For example, elderly people have different needs than younger people, so their valued capabilities will differ from those of younger people. Although this is true among humans, when we want to distinguish humans and robots, we need to specify what it means to lead a good life. However, as long as we retain the idea that human beings are not robots, are different from them, we may as well retain the idea that human beings share certain distinctive traits, regardless of their local environment. This would help to draw the line between humans and robots.

<sup>1</sup> Among these items, imagination is perhaps one of the most important human capabilities in the context of stakeholder dialogue. See Blok (2014, pp. 5–6). In political philosophy, Nussbaum’s capability approach suggests national governments should constitutionally guarantee these ten capabilities to all their citizens so that they would be able to live a normal range of life (capability 1), have a healthy life (capability 2), enjoy their own bodily freedom (capability 3), use the senses, imagination, and thinking in a truly human way (capability 4), develop their own good emotions towards others such as love and grief (capability 5), form a critical conception of their own good life (capability 6), engage in a dignified social interaction with others (capability 7), care for non-human animals and nature (capability 8), laugh and enjoy recreational activities (capability 9), and make political as well as material choices (capability 10).



**Fig. 1** The relationship between resources, capabilities, and functionings

## The Capability Approach and Its Application to AI Robots

How does technology influence human capabilities? This question has already been posed and tackled by several authors in their efforts to bring the capability approach into the discussion of the ethical implications of AI technologies, as robots can both enhance and diminish human capabilities.

As Oosterlaken shows in her extensive literature review (Oosterlaken 2012), the capability approach has been applied to the field of technologies, most frequently in ICT, and in particular, for development in ICT.<sup>2</sup> However, it has also been applied to the field of robotics for evaluating the quality of healthcare provision by robots (Coeckelbergh 2010; Borenstein and Pearson 2010).

For example, in arguing for replacing human care with AI assistive technologies, Mark Coeckelbergh points out that we need to develop comprehensive criteria for good care, and that Nussbaum’s capabilities approach, which has the principle of human dignity at its core, is useful (Coeckelbergh 2010, p. 184). In considering good care, he argues that Nussbaum’s list of capabilities “can be used as a list of criteria to evaluate health care and the use of AI technology in health care” (p. 185). The list, according to him, is a “particularly useful instrument to articulate what is at stake in health care ethics” (p. 188).

Borenstein and Pearson (2010) also view Nussbaum’s list of capabilities as a way to promote human flourishing in care work. Emphasising the importance of autonomy and individual choices for a meaningful life, they state that “the capabilities approach would clearly require designing and using robots in a way that expands opportunities for human flourishing for all human beings, including those whose rational faculties are limited or those whose protection by a social contract may be in question” (p. 279). What can be clearly drawn from these studies, and easily agreed with, is that, according to the capability approach, robotics design and use must resemble expanding human capabilities in order to flourish—that is, to be able to make choices about one’s life.

Although humanoid AI robots are not human beings, they are expected to behave like humans, at least to some extent. Moreover, stakeholders also need to behave like humans in robotics entrepreneurship; otherwise, we will face a dystopia. Therefore, we need to shape an idea of what it means to be human, and Nussbaum’s (1990) “thick vague theory of the good”

<sup>2</sup> Oosterlaken speculates that there are two reasons for this. One is the popularity of ICT as a “weapon against poverty” in the last decade or so, and the other is ICT’s indeterminate character “in the sense that it can directly and simultaneously contribute to the expansion of human capabilities in very different areas: health, education, recreation, livelihoods, democracy, etc.” (Oosterlaken 2012, p. 12).



may be useful in this regard. According to her, we who uphold the idea of human rights inscribed in the UN Declaration of Universal Human Rights share the belief that certain characteristics, such as reasoning and compassion, which are very often encouraged in public education, as well as in public culture, in most parts of the world—are ethically essential to being human. Those characteristics are therefore evaluated objectively as being good.

We believe that, in dealing with the issues of robotics entrepreneurship, it is better for us to take the stance that Nussbaum (1992) calls “internalist essentialism,” which understands the content of the human essence from *within* our human experiences and from which she derived her list of human capabilities. This capability approach is inherently an ethical framework that focuses our understanding about what it means to be human. Appealing to the capability list is perhaps one way to draw a line between what is permissible and what is not for entrepreneurs to assign to AI robots. For, even though we humans are driven by desires, we have been able to maintain some communal living because of local ethics. If an ethic can arise and be shared by a community, then it can be shared by a community on a global scale. The content of the capability list, which is subject to perpetual updating, can be referred to as a shared good at the core of humanity.

It is important to note that some claim a certain pragmatist resonance in the capability approach (McReynolds 2002; Zimmermann 2006). For example, McReynolds (2002) points out that “[b]oth Dewey’s and Nussbaum’s approaches view ongoing human activity as the central concept in ethical inquiry. Moreover, both are concerned with taking seriously the interests and desires of actual humans beings without wanting to foreclose possibility of reform, as would a superficial relativism” (p. 143). The capability approach with its internalist essentialism and open-endedness looks almost pragmatically at human activity, at what persons with aspirations can do and be. On the other hand, as we will remark in the next section (3–3), Dewey’s ideas of desire and purpose may correspond to what Sarasvathy (2001, 2008) calls aspirations and goals. As Dewey says:

[t]he widening of the area of shared concerns, and the liberation of a greater diversity of personal capacities which characterize a democracy, are not of course the product of deliberation and conscious effort. On the contrary, they were caused by the development of modes of manufacture and commerce, travel, migration, and intercommunication, which flowed from the command of science over natural energy. But after greater individualization on one hand, and a broader community of interest on the other have come into existence, it is a matter of deliberate effort to sustain and extend them. (Dewey 1916, p. 101)

We live in a world where AI issues are no longer negligible. Responsible innovation seems to be the most compatible with the idea of a permissible range of robotics entrepreneurship that is based on a shareable ethic and focused on individual capabilities.

We focus on the capability approach among several ethical frameworks precisely because we believe that unlimited profit-seeking robotics entrepreneurship could be socially detrimental and that, in order to avoid such a potential disaster, we need stakeholders with *phronesis* i.e. a person’s practical intelligence, particularly in discerning how or why to act with virtue, that Aristotle advocated in his *Nicomachean Ethics*. With this virtue, one can be both rational and moderate regarding one’s end. This is exactly the kind of practical intelligence that we believe the stakeholders in robotics entrepreneurship should have; and arguably, some of the conditions for acquiring such intelligence can be found in the capability list. In the next section, we

discuss the effectuation process, which is a design process that can show us how robotics entrepreneurship could link the design of AI robots with phronesis.

## Effectuation and Robotics Entrepreneurship

In this section, we will discuss the collaborative design process among stakeholders using robotics. This can be modelled as an effectual process, which has been proposed in recent entrepreneurship research.

### Entrepreneurship is Not Causal Reasoning

Entrepreneurship researchers have long discussed whether business opportunities exist objectively or not (Shane and Venkataraman 2000; Harmeling et al. 2009). Many researchers, assuming that entrepreneurs use causal reasoning, describe how they explore and exploit objective opportunities provided from new sources (causes), such as new technologies (Shane 2000; Drucker 1993; Kirzner 1997). Few researchers grasp the actual process of emergence, which has never been fully explained causally as though it were a natural science. Sarasvathy (2001, 2008) has studied this social process more closely, determining that opportunities are subjectively made. That is, expert entrepreneurs transform their personal and available means into new opportunities in unknowable environments. Indeed, opportunities do not merely exist objectively; rather, they are designed subjectively. Sarasvathy called this “effectuation” and described it as the inverse logic of causation. Effectuation has been rigorously derived by a method of cognitive science and profoundly guided by pragmatism. It might also support the recent practical discussion that no market (opportunity) initially exists before entrepreneurs; rather, entrepreneurs gradually create and organise it, starting by cultivating would-be customers (Blank 2013). We will now review Sarasvathy’s seminal work.

### Effectuation

Sarasvathy (2008) collected and analysed information in experimental settings and found several heuristics—that is, empirical rules humans develop for solving unknown problems—that expert entrepreneurs tend to use when they face typical decision-making issues during the business start-up phase. She distilled these into five principles, which we summarise as follows after Sarasvathy’s (2008, pp. 73–95) metaphors (in parentheses):

- P1: (Bird-in-hand) Starting with means and creating new effects
- P2: (Affordable loss) Beginning with a determination of affordable loss
- P3: (Crazy quilt) Alliances with and pre-commitments from stakeholders
- P4: (Lemonade) Exploiting contingencies
- P5: (Pilot-in-the-plane) To the extent that we can control the future, we do not need to predict it.

Sarasvathy (2008) examined how entrepreneurs use these heuristics in actual situations, using a process model as a thought experiment. Initially, entrepreneurs have some means that arise from “who I am, what I know, and whom I know” (p. 101) and set tentative goals (P1) to the extent of their affordable loss (P2). With those goals, they interact with other people, some of whom make

commitments (P3) within their own affordable loss structures (P2), and those self-selected people form a team to provide new means and goals. The team iterates this process (means – goals – commitments) and gradually identifies viable artifacts (e.g., markets and opportunities) (P5). Sarasvathy sometimes uses the word “effects” instead of “artifacts,” depending on the context. In this paper, we use “artifacts” to connote something designed artificially.

At the outset, effectuation is moved by “varied imaginations and diverse aspirations” (Sarasvathy 2008, p. 73), rather than by a specific and clear goal, as in causation logic. Practically, entrepreneurs use both logics, effectuation and causation, but, in the early stage of a new venture, entrepreneurs prefer effectual over causal logic. The theory of effectuation has been developed in a variety of conceptual and empirical papers (for example Dew 2009; Read et al. 2009a, b; Wiltbank et al. 2006, 2009).

Before we draw on this process model for discussing robotics entrepreneurship, we will briefly discuss three problematic issues. Arend, Sarooghi, and Burkemper (2015, p. 639) formally evaluated and critiqued this model as being insufficient to be used as a theory. However, it may not be useful to evaluate effectuation as a theory. The heuristics in effectuation are inherently context dependent; therefore, this model comes close to Weber’s ideal type for describing overall meanings, but not for offering causal explanations as a theory does. Indeed, the means-driven logic of P1 is the reverse of causal explanation (Sarasvathy 2008, p. 16).

We have two other criticisms that relate to P1 reasoning concerning goals, effects (or artifacts), and aspirations. First, the difference between goals and effects is unclear. Sometimes, they are used interchangeably. Causation is “goal-driven” (p. 15); it chooses “among means to create a particular effect” (p. 75). In contrast, effectuation is “means-driven” (p. 15) and consists of “designing possible effects using a particular set of means” (Sarasvathy 2008, p. 75). The relationship between goals and effects (or artifacts) is also unclear. Sarasvathy asserts that effectuation “begins with a given set of means and allows goals to emerge contingently over time from the varied imaginations and aspirations” (Sarasvathy 2008, p. 73). However, she has also argued that the commitments of diverse stakeholders “constrain future sub-goals and goals that get embodied into particular features of artifact” (Sarasvathy 2008, p. 109). We will clarify the definition of “goal” later on.

Second, aspiration seems to be regarded as the initiator of effectuation, but this concept is not explained sufficiently. The aspiration (e.g., to cook a meal) is that “the generalized end goal ... remains the same in both causation and effectuation” (p. 74), but “an effect is the operationalization of an abstract human aspiration” (Sarasvathy 2008, p. 75). In this explanation, effectuation resembles causation in the broad sense of attaining the goal (see also Sarasvathy 2001).

## Goals, Artifacts, and Aspirations

Almost 80 years ago, John Dewey (1938) stated that the formation of a purpose, or end-view, always starts with desire, but purpose is different from desire because it is translated into a plan based on foresight concerning the outcome of an action. *Desire* and *purpose* may correspond, respectively, to what Sarasvathy has referred to as *aspirations* and *goals*.

The formation of purposes is, then, a rather complex intellectual operation. It involves (1) observation of surrounding conditions; (2) knowledge of what has happened in similar situations in the past, a knowledge obtained partly by recollection and partly

from the information, advice, and warning of those who have had a wider experience; and (3) judgment which puts together what is observed and what is recalled to see what they signify. A purpose differs from an original impulse and desire through its translation into a plan and method of action based upon foresight of the consequences of acting under given observed conditions in a certain way. (Dewey 1938, p. 68–69)

Dewey's observations may provide insights into a more detailed process of effectuation. At least, we can assume that humans have accumulated goal-pursuing experience in past situations. If Dewey is right, humans refer to such experiences and may create both new goals and plans and methods of action (artifacts) for the current situation. This also sheds light on the collaborative nature of effectuation in which, gradually accumulating many experiences through an interaction with stakeholders, new goals and plans (artifacts) are created for a current situation. Two further questions arise: are aspirations (desires) only the spark for forming goals (purposes)? do goals define artifacts (plans) after their formation?

The Effectual process is congruent with the formation of a purpose, and, as in Dewey's explanation, it can be viewed on both the individual and the societal level. According to Huang and Bargh (2014), a goal represents the desired end state of a plan (artifact). In the effectual process, a goal may arise only after a new plan has been invented from means. This leads to the notion that once a new plan becomes sufficiently clear, we can control actions by setting any goal, whereby the plan can be modified and optimised to attain that goal. We have assumed that the concept of aspirations corresponds to desires (Dewey 1938) and unconscious goals (Huang and Bargh 2014). Huang and Barge argue that unconscious goals (aspirations) have a strong influence on actions, sometimes more than conscious goals. Further research is needed on the dynamic relationship between goals and artifacts.

## **Toward Robotics Entrepreneurship**

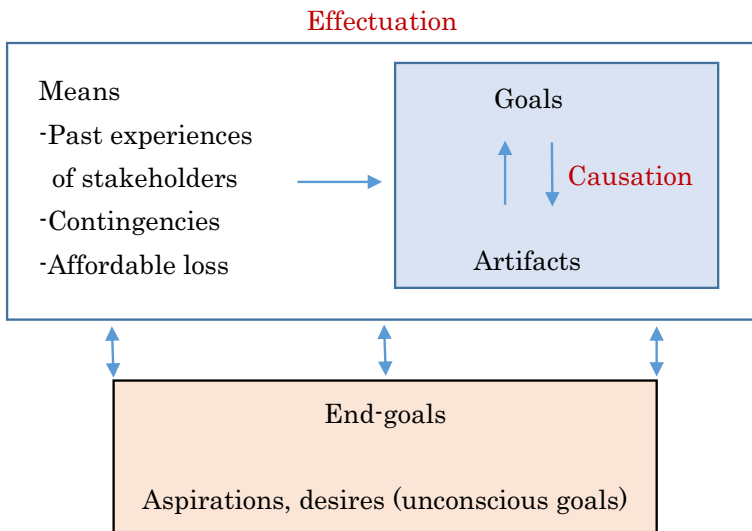
Whereas effectuation is a process that designs artifacts and goals from available means, causation is a process that sets a goal for a certain artifact or optimises an artifact for some goal. Figure 2 extends Sarasvathy's (2008) effectual process in three respects. First, the means include contingencies (P4) and affordable losses (P2) for simplicity, given that they are selective conditions that are inherently inclusive in the variety of means. Second, we have clarified that the effectuation process comprises the causation process. Third, it explicitly shows how unconscious aspirations influence the whole process of effectuation.

AI Robots can be not only a means like conventional technologies but also a part of the entire effectual process through their learning mechanisms. Robotics entrepreneurs must develop and use AI robots with deliberation from the outset because the thinking part can evolve rapidly and autonomously, without aspirations or towards unlimited profit-seeking.

## **Using the Capability-Effectual Approach to Understand Robotics Entrepreneurship**

### **The Capability-Effectual Approach**

Figure 3 shows our capability-effectual approach. Here, people's aspirations are based on their present capability space. They may lack some capabilities—for example, bodily

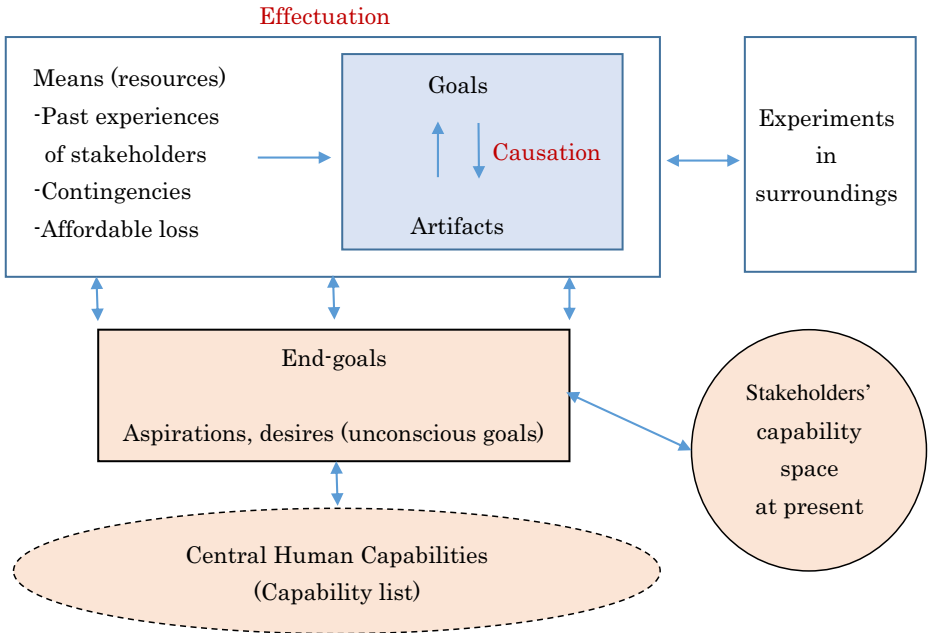


**Fig. 2** Effectual process extended

integrity, owing to a car accident and a consequent loss of legs. In this case, even with the help of a wheelchair, they may not be able to move around sufficiently to run a house, carry out such tasks as turning lights on/off, sweeping the floors, opening/closing doors and windows, etc. Such people may aspire to expand their ability to run their houses with the help of AI robots, and their end goals will be achieved once their capability space reaches the level of their expected or desired capability space with the help of robotics entrepreneurship. In addition, the content of such capability space must be within the permissible range of reasonable empiricism for obvious reasons: even if a person wishes to murder the driver who ran over them, AI robots should not have such a capability. The capability list that drives effectuation, therefore, must represent the values and ethics of the contemporary human community, controlling the unconscious goals (aspirations and desires) of individuals.

By merging the capability approach with effectual process in this way, we believe that we can determine a permissible range for goals and purposes/activities under robotics entrepreneurship. If AI robots are to live with human beings, their artificial intelligence should be ethical in terms of this permissible range. Despite ongoing efforts to implement ethics for AI robots, it is unlikely that AI robots will become moral agents in the way that humans are. Nonetheless, they can be given the end goals of behaving as limited moral agents by human beings.

Humans share end goals in the range created by the listed capabilities, which is never established once and for all but is continuously being translated into goals and purposes in local settings. In robotics entrepreneurship, this process is dynamic and never-ending, engaged in by entrepreneurs in collaboration with philosophers. Entrepreneurs themselves need to have a general idea of what it means to be human. More importantly, they need to have certain capabilities to engage in that process. Researchers classify entrepreneurial motivations into two types—necessity and aspiration—which closely relate to the capabilities in Nussbaum’s list. The lack of some capabilities, such as bodily health



**Fig. 3** The capability-effectual approach

or bodily integrity, leads directly to the necessity for a basic life. Others, such as those relating to other species, play, and control over one’s environment, link to aspirations for a better life. Entrepreneurial aspirations, not only ideals but also necessities in a broad sense, trigger and continuously energise an effectual process, which begins with the aspirations that arise from a person’s present capabilities. With regard to end goals, an entrepreneur continuously reinvents goals and purposes from usable means. Without end goals or aspirations, the effectual process is unlikely to continue. Such goals must not only be within a permissible range, but must also serve as a necessary precondition for an effectual process. A capability-effectual approach holds that robotics entrepreneurship can invent and reinvent goals and purposes in order to use robots by having end goals as aspirations within a permissible action range.

**Conclusion**

In this paper, we started by stating that with AI robots humans are faced with an unprecedented type of technology. Robotics researchers are calling for discussions about goals and values so that they may control AI robots. This is a matter of responsible innovation for “global catastrophic risk” (GCF 2016). To address this, we have proposed the capability-effectual approach drawing on two levels of research: the capability approach in ethics, and the effectual process model in entrepreneurship research. An effectual process approach is based on the collaborative design process of stakeholders. A capabilities approach provides central human capabilities from a “thick vague theory of goods,” which guides the effectual process through individual goals, values, purposes, and aspirations. More precisely, by assuming a correspondence between goals, desires, purposes, and aspirations in the languages of different

disciplines, the capability-effectual approach clarifies both how a capability list working globally could affect the aspirations and end-goals of individuals, and how local aspirations and end-goals could either stimulate or limit effectual processes.

Theoretically, the capability-effectual approach contributes to previous discussions about the collaboration of stakeholders and the design process in responsible innovation research by showing how they are linked. Practically, the capability-effectual approach might contribute to the sound development of human communities. Moreover, from now on, it may be indispensable for humble human experiments because robotics and artificial intelligence have developed to a stage of autonomous deep learning and have begun to permeate human effectual processes.

Therefore, we argue that stakeholders in robotics entrepreneurship must not only use some kind of capability-effectuation approach themselves, but, by equipping AI robots with this capability, they should also develop AI robots that think like a human entrepreneur concerned about ethics. We can reformulate this as the definition of the permissible action range for AI robotics entrepreneurship. Although the scope of this paper is limited in arguing that we need to set a permissible action range for robotics entrepreneurship, and that we can do so by merging the capability approach with effectual processes, we believe this is indeed a sound, albeit small, step.

We suggest there are three key items that should be placed on the future ethical research agenda of robotics entrepreneurship. The first item should be the concretisation of the as yet general and abstract capability list. In stakeholder workshops, this list could be interpreted in terms of the characteristics of AI robotics. A second important item for further research is how robotics entrepreneurs can use the capability-effectuation approach to develop ethically “good” AI robots that become more helpful for humans and their environment. Our suggestion is to first develop the ethical skills of the robots before training them to do other kinds of things. A third important item, which we did not tackle in this paper: who decides this permissible action range in practice? On this point we suggest that much may be learned from the EU framework “HORIZON 2020,” which states that “the grand challenges of our time”—including, of course, robotics—“require the active involvement of public actors and stakeholders in research and innovation processes (European Commission 2011)” (Blok 2014, p. 1). This EU framework appears to have a global reach and is worthy of further examination in a subsequent paper.

**Acknowledgements** The authors would like to thank the anonymous reviewers of this journal for their helpful comments and also the participants at the Philosophy of Management Conference 2016 held at St Anne’s College, Oxford from 14 to 17 July 2016, who provided insightful comments on an earlier draft. They would also like to thank Vincent Blok for his very helpful comments and Hideki Hashimoto for his inspiring discussion on our primary idea behind this paper.

### Compliance with Ethical Standards

**Conflict of Interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Anonymous. 2015. What forms will robots take in the future: an interview with Rich Mahoney, director of robotics at SRI international. *AXIS*. April, 32–35.
- Blank, S. 2013. *The four steps to the epiphany*. Menlo Park: K&S Ranch.
- Blok, V. 2014. Look who's talking: responsible innovation, the paradox of dialogue and the voice of the other in communication and negotiation processes. *Journal of Responsible Innovation* 1 (2): 171–190.
- Blok, V., L. Hoffmans, and E.F.M. Wubben. 2015. Stakeholder engagement for responsible innovation in the private sector: Critical issues and management practices. *Journal on Chain and Network Science* 15 (2): 147–164.
- Blok, V., and P. Lemmens. 2015. The emerging concept of responsible innovation. Three reasons why it is questionable and calls for a radical transformation of the concept of innovation. In *Responsible innovation 2: Concepts, approaches, and applications*, ed. B.J. Koops et al., 19–35. Dordrecht: Springer International Publishing.
- Borenstein, J., and Y. Pearson. 2010. Robot caregivers: Harbingers of expanded freedom for all? *Ethics and Information Technologies* 12 (4): 277–288.
- Bostrom, N. 2014. *Superintelligence: Paths, dangers, strategies*. Oxford: OUP Oxford.
- Bostrom, N., and E. Yudkowsky. 2011. *The ethics of artificial intelligence*, retrieved on Jan. 20, 2017 from <http://www.nickbostrom.com/ethics/artificial-intelligence.pdf>.
- Coeckelbergh, M. 2010. Health care, capabilities and AI assistive technologies. *Ethical Theory and Moral Practice* 13 (2): 181–190.
- Dew, N. 2009. Serendipity in entrepreneurship. *Organization Studies* 30 (7): 735–753.
- Dewey, J. 1916. *Democracy and education: an introduction to the philosophy of education*. New York: Macmillan.
- Dewey, J. 1938. *Experience and education*. New York: Simon & Schuster.
- Dhar, V. 2016. The future of artificial intelligence. *Big Data* 4 (1): 5–9.
- Drucker, P. 1993. *Innovation and entrepreneurship*. London: Harper Collins.
- Future of life institute. n.d. *An open letter: research priorities for robust and beneficial artificial intelligence*. Retrieved on May 16, 2016 from <http://futureoflife.org/ai-open-letter/>.
- Global challenge foundation. 2016. *Global catastrophic risks 2016*. Retrieved on May 16, 2016 from <http://globalprioritiesproject.org/2016/04/global-catastrophic-risks-2016/>.
- Harmeling, S.S., S.D. Sarasvathy, and R.E. Freeman. 2009. Related debates in ethics and entrepreneurship: Values, opportunities, and contingency. *Journal of Business Ethics* 84 (3): 341–365.
- Huang, J.Y., and J.A. Bargh. 2014. The selfish goal: autonomously operating motivational structures as the proximate cause of human judgment and behavior. *Behavioral and Brain Sciences* 37 (02): 121–135.
- Kirzner, I.M. 1997. Entrepreneurial discovery and the competitive market process: an Austrian approach. *Journal of Economic Literature* 35 (1): 60–85.
- Mataric, M.J. 2007. *The robotics primer*. Cambridge: The MIT Press.
- Matsuo, Y. 2015. 人工知能は人間を超えるか: ディープラーニングの先にあるもの. [Do artificial intelligences outreach human? Thinking of deep learning]. Tokyo: KADOKAWA.
- McReynolds, P. 2002. Nussbaum's capabilities approach: a pragmatist critique. *The Journal of Speculative Philosophy* 16 (2): 142–150.
- Nussbaum, M.C. 1990. Aristotelian social democracy. In *Liberalism and the good*, ed. R.B. Douglass et al., 203–252. New York: Routledge.
- Nussbaum, M.C. 1992. Human functioning and social justice: In defense of Aristotelian essentialism. *Political Theory* 20 (2): 202–246.
- Nussbaum, M.C. 2000. *Women and human development: the capabilities approach*. Cambridge: Cambridge University Press.
- Ogawa, Y. 2014. 革新こそ最大の防御. [*Innovation is the best defense*], 3. Nikkei Shimbun.
- Oosterlaken, I. 2012. The capability approach. Technology and design: taking stock and looking ahead. In *The capability approach, technology and design*, ed. I. Oosterlaken and J. van den Hoven. Dordrecht: Springer.
- Read, S., N. Dew, S.D. Sarasvathy, M. Song, and R. Wiltbank. 2009a. Marketing under uncertainty: the logic of an effectual approach. *Journal of Marketing* 73 (3): 1–18.
- Read, S., M. Song, and W. Smit. 2009b. A meta-analytic review of effectuation and venture performance. *Journal of Business Venturing* 24: 573–587.
- Sarasvathy, S.D. 2001. Causation and effectuation: toward a theoretical shift from economic inevitability to entrepreneurial contingency. *Academy of Management Review* 26 (2): 243–263.
- Sarasvathy, S.D. 2008. *Effectuation: elements of entrepreneurial expertise*. Northampton: Edward Elgar Publishing.



- Shane, S. 2000. Prior knowledge and the discovery of entrepreneurial opportunities. *Organization Science* 11 (4): 448–469.
- Shane, S., and S. Venkataraman. 2000. The promise of entrepreneurship as a field of research. *Academy of Management Review* 25 (1): 217–226.
- Stilgoe, J., Owen, R. and P. Macnaghten (2013). Developing a framework of responsible innovation, *Research Policy*, 42: 1568–1580.
- Van den Hove, J. 2013. Value sensitive design and responsible innovation. In *Responsible innovation: managing the responsible emergence of science and innovation in society*, ed. R. Owen, J. Bessant, and M. Heintz, 75–84. London: John Wiley & Sons.
- Von Schomberg, R. 2013. A vision of responsible research and innovation. In *Responsible innovation: managing the responsible emergence of science and innovation in society*, ed. R. Owen, J. Bessant, and M. Heintz, 52–74. London: John Wiley & Sons.
- Wiltbank, R., N. Dew, S. Read, and S.D. Sarasvathy. 2006. What to do next? The case for nonpredictive strategy. *Strategic Management Journal* 27 (10): 981–998.
- Wiltbank, R., S. Read, N. Dew, and S.D. Sarasvathy. 2009. Prediction and control under uncertainty: strategy in new venture investing. *Journal of Business Venturing* 24 (2): 116–133.
- Zimmermann, B. 2006. Pragmatism and the capability approach: challenges in social theory and empirical research. *European Journal of Social Theory* 9 (4): 467–484.

**Yuko Kamishima** is Professor of Philosophy at College of Psychology of Ritsumeikan University. She received her doctoral degree from University of Tokyo. Her published work includes *Post-Rawlsian Theory of Justice: Sen, Pogge and Nussbaum* (Mineruva Shobo, 2015, in Japanese). She has recently published a translation of Onora O'Neill's *Bounds of Justice* (2000) (Misuzu Shobo, 2016, in Japanese).

**Bart Gremmen** is professor of Ethics in Life Sciences at the Philosophy Group of Wageningen University. He received his PhD from the University of Twente. His current research is about environmental ethics, animal ethics philosophy of technology, and hermeneutics of science and technology.

**Hikari Akizawa** is an independent researcher, and her current research interests include entrepreneurship, family business, and corporate governance. She is a president of both Oikos Research and Japan Academy of Family Business. Until 2015, she was a full professor of management at Chuo University for fifteen years. She received a Ph.D. from Department of Value and Decision Science at Tokyo Institute of Technology in 1999.