

# *AI Relationality and Personhood*

with Fraser Watts and Marius Dorobantu, “The Relational Turn in Understanding Personhood: Psychological, Theological, and Computational Perspectives”; William F. Clocksin, “Guidelines for Computational Modeling of Friendship”; Michael J. Reiss, “Is It Possible That Robots Will Not One Day Become Persons?”; and Léon Turner, “Will We Know Them When We Meet Them? Human Cyborg and Non-Human Personhood.”

## THE RELATIONAL TURN IN UNDERSTANDING PERSONHOOD: PSYCHOLOGICAL, THEOLOGICAL, AND COMPUTATIONAL PERSPECTIVES

by Fraser Watts and Marius Dorobantu 

*Abstract.* From the middle of the twentieth-century onwards, there has been a growing emphasis on the importance of relationality in what it means to be human, which we call a “relational turn.” This is found in various domains, including philosophical psychology, psychoanalysis, and theological anthropology. Many have seen a close connection between relationality and personhood. In the second half of the article, we consider the implications of this trend for artificial intelligence (AI) and robotics. So far, AI has largely neglected relational intelligence, though that could perhaps be about to change. Cybernetics was rendered more open to assumptions about the contextuality of intelligence by its rather different assumptions from AI. Social robotics increasingly requires relational intelligence, and promising steps might be found in computational modeling of human relationships. Questions about whether robots can achieve personhood are difficult to resolve, though the possibility should not be ruled out.

*Keywords:* AI; anthropology; cybernetics; intelligence; personhood; relationality; robotics

---

Could a robot be a “person”? Could a robot be “relational”? These questions are the focus of this special section. The questions are related, so much so that some people might say they are identical, as relationality

Fraser Watts was formerly Reader in Theology and Science at the University of Cambridge, and is now Visiting Professor of Psychology and Religion at the University of Lincoln, and Executive Secretary of the International Society for Science and Religion; e-mail: [fraser.watts@cantab.net](mailto:fraser.watts@cantab.net). Marius Dorobantu is a Postdoctoral Research Associate and Lecturer at the Faculty of Religion and Theology, Vrije Universiteit Amsterdam, and a Fellow of the International Society for Science and Religion; e-mail: [m.dorobantu@vu.nl](mailto:m.dorobantu@vu.nl).

is the essence of personhood. However, we prefer to keep some conceptual distinction between personhood and relationality, so the claim that persons are relational is a substantive claim, rather than one that is true by definition.

There has been a growing interest since the mid-twentieth century in relationality and personhood. It has been an important feature of Christian theological anthropology in this period, but interest in relationality has not been confined to religious thinking. It is important to point out that this shift, often referred to as the “relational turn,” is not unique to theology. Some theologians who write about the relationality of human persons don’t seem as aware as they might be that they are part of a cross-disciplinary movement in Western thinking. The relational turn has been evident in much philosophical and psychological thinking about the human person as well as in theological anthropology. In the first half of this article, we will review the relational turn in thinking about humans in psychology and theology.

We will then turn to the challenge that relationality and relational intelligence pose for artificial intelligence (AI) and robotics. We note that cybernetics, from which AI emerged, approached intelligence in much more systemic terms, with an emphasis on the context or overall system in which intelligence was operating. So far, AI has not made much progress in modeling or emulating relational intelligence, but there are some interesting trends on which to build. It is an important challenge for robotics, if robots are to interact effectively with humans.

#### CHANGING ASSUMPTIONS ABOUT HUMAN NATURE

One of the earliest sources of the personalist tradition is the Scottish philosopher, John Macmurray (1891–1976), particularly in his influential Gifford lectures delivered in 1953 and 1954: *The Self as Agent* (Macmurray 1957) and *Persons in Relation* (Macmurray 1961). He saw his philosophy as rejecting both mechanistic and organicist frameworks in favor of what he called “personalism.” His focus was on human nature, and he summed up his position in the introduction to *The Self as Agent* by saying: “The simplest expression that I can find for the thesis I have tried to maintain is this: All meaningful knowledge is for the sake of action, and all meaningful action for the sake of friendship” (Macmurray 1957, 15). He was also ahead of his time in emphasizing the importance of emotion rather than cognition in influencing motivation (Macmurray 1935).

Though Macmurray was a philosopher, he did not engage much in his later years with the philosophy of his day, and philosophy has largely neglected him. Philosophically, he may have been shaped by British idealism more than by anything else (McIntosh 2011). Though the relational approach to personhood espoused by John Macmurray has been

widely accepted, he seems to have had rather little direct influence. After a period of neglect, the first full length study of his work was edited by Philip Conford, an advocate of organicist farming, in association with Tony Blair (Conford and Blair 1997). There is also a volume about his work edited by David Fergusson and Nigel Dower (2002).

A later thinker in the personalist tradition was Rom Harré, though he does not refer to Macmurray. Harré was mainly a philosopher who taught philosophy of science at Oxford for 35 years. He was also a brilliant, prolific, and charismatic polymath, who worked in mathematics, psychology, social science, and chemistry. A particularly important book in the present context is *Personal Being* (Harré 1983), though it was not his first venture into persons and relationality. That began with a book written with Paul Secord (Harré and Secord 1973) that built on philosophical work on human action and argued that human social behavior should be understood in terms of intentions rather than causes. It is a position similar to that set out by Macmurray in *Self as Agent*, though there is no reference to Macmurray. *Personal Being* was the middle volume of a trilogy that had started with *Social Being*, and which ended with *Physical Being*.

We will come shortly to the developing emphasis on relationality in theology, and to some extent in AI, but first we note two other examples of the broad shift toward a more relational way of thinking about humans. One is in psychoanalysis, which made a significant shift toward relationality in the mid-twentieth century. The pressure to move away from explanations framed in terms of drive, and toward relational explanations, seems to have come largely from internal pressures rather than external influences. However, one of the first British “object relations” theorists was Ronald Fairburn, based in Edinburgh, who was aware of Macmurray’s work (Clarke 2006). Surprisingly, the move toward a more relational approach in psychoanalysis seems to have taken place independently in the United Kingdom and the United States. It was termed “object relations” in the United Kingdom, and is often called “relational psychoanalysis” in the United States, though these two independent movements were subsequently synthesized.

Another field in which the relational turn is being felt is in the evolution of religion. The so-called “cognitive science of religion” (CSR), which has been dominant for some time, makes largely individualistic assumptions about how early humans understood the world at the dawn of religion. As Leon Turner has argued, it is still largely trapped within the constraining assumptions of abstract individualism (Turner 2020). In contrast, Robin Dunbar’s approach to the evolution of religion, framed in terms of the social brain, moves beyond abstract individualism and takes a more radically relational approach to the evolution of religious cognition (Dunbar 2022).

It is an approach that provides a helpful lens through which to examine current changes in religion (Watts and Dorobantu 2023). Dunbar’s

approach makes use of the classic distinction between shamanic and doctrinal religion. Watts has argued that whereas shamanic religion involves a mode of cognition that is intuitive, embodied, and socially embedded, doctrinal religion uses a different mode of cognition that is more conceptual and propositional (Watts 2020). Religious experience in the doctrinal phase of religion seems to be more individualistic than the radically collectivist experience of participation in trance dancing.

We endorse the relational turn in thinking about human nature. Humans are socially embedded and interactive, and are indeed shaped by their social context. They are influenced by that context in a constant two-way, systemic interaction between individuals and their interpersonal worlds. More than that, we endorse the more radical view that humans are *constituted* by their relationships (Turner 2012). To be in relationship with others is a core and essential part of what it is to be human. That is not contradicted by the lives of those who choose to be hermits, as the human relations from which they are choosing to be absent are still significant in constituting their humanity; hermits have often had a good deal of interaction with visitors and those who support them, and they position themselves in cross-temporal spiritual relationships with predecessors from their traditions. Neither is the centrality of relationships for human nature contradicted by those with special modes of cognition such as neurodiversity, that gives a particular quality to the human interactions (Leidenhag 2020)

### Theological Anthropology

Theology seems to have reached relational assumptions largely through its own internal resources, drawing implications from Trinitarian theology. There is a large recent literature in theological anthropology emphasizing relationality and personhood, dating from John Zizioulas' *Being as Communion* (1985) and Alistair McFadyen's *The Call to Personhood* (1990). The relationalist thinker from outside theology with whom Alistair McFadden engaged most was Rom Harré, especially his book, *Personal Being* (1983). Leon Turner, who contributes to this special section, has also engaged with Harré in his work on relationist theological anthropology (2008).

This theological literature never mentions John Macmurray, which is surprising, especially as McIntosh (2011) considers Macmurray to be a religious thinker. There are at least two possible routes by which John Macmurray might have influenced theology, but neither seems to have been particularly significant. One is through David Fergusson, an eminent Scottish theologian who was very aware of Macmurray's work (Fergusson and Dower 2002). Also, at least some of the early object relations theorists were religious and were engaged with theology to some extent, including

Donald Winnicott, a high church Anglican, and Harry Guntrip, a former Methodist minister. Object relations theory was sometimes associated with a more positive psychoanalytic view of religion than in Freud (Parsons 2021).

Rowan Williams (2018) attributes the personalist turn in theological anthropology to an essay by Vladimir Lossky on the “theological notion of the human person.” Though most readily available in Lossky’s book, *In The Image and Likeness of God* (1974), it was originally published in 1955, so there is almost exact synchronicity between his essay and Macmurray’s Gifford Lectures, though it is most unlikely that they were aware of each other. Rooted in the Eastern theology of the person of Christ, Lossky emphasizes the importance of distinguishing something being a unique instance of its kind, and the quality that makes each one “irreducible to its nature” (Williams 2018, 29).

For the most part, it is the relationality of the three persons of the Trinity that has been most influential in the development of personalism in theological anthropology. In brief, the argument is that relationality is a defining feature of the Trinitarian God, and that God’s relationality is reflected in the relationality of humans. It is the approach taken by Alistair McFadyen (1990), Stanley Grenz (2001), and many others. It is a fair point, as far as it goes. However, it should be recognized that *person* in Trinitarian theology is a technical term, and should not be understood in the same way as a human *person*. Human relationality is clearly very different from Trinitarian relationality. Humans have sufficient independence from one another that there can be divisions between them. In contrast, though the persons of the Trinity can be distinguished, they cannot be divided from one another, as the Athanasian creed asserts at some length.

Some of the most sophisticated work in theological anthropology is to be found in reflection on how to define the image of God (Cortez 2010; Dorobantu in press a). Much classical work has focused on the substantive interpretation of *imago Dei*, and especially on the special nature of human rationality. However, there has been a growing consensus that, theologically, this is not a satisfactory approach. Various alternative, more theological, approaches have been advanced in recent decades, including the functional interpretation (focusing on what humans do rather than what they are), the eschatological interpretation (focusing on what humans are called to become), and the relational interpretation (focusing on the distinctive features of the relationships humans have with others and with God).

Marius Dorobantu (2022 in press a) argues that the relational interpretation of *imago Dei* is the most convincing, on two distinct grounds. One is that all other theological interpretations rely to some extent on the relational interpretation. In that sense, relationality seems to be widely acknowledged as being at the heart of any theological interpretation of

*imago Dei*. The other is that the relational interpretation provides the most convincing way of distinguishing between humans and intelligent machines. This is not an argument against nonrelational interpretations, just a claim that the relational interpretation is central. Dorobantu also develops the relational interpretation of *imago Dei* in a distinctive way, suggesting that the key feature is the sense of being a person, and of other humans as also being persons. That, in turn, depends on the distinctive conceptual, reflective consciousness of humans, as well as their metacognitive abilities. Humans don't only have the ability to know things; but also to know and be able to conceptualize what they know.

Though relationships are a *necessary* feature of personhood, as well as a staple of *imago Dei*, we suggest that relationality does not in itself provide a *sufficient* account. Following Dorobantu (2021), we propose a multilevel approach in which there is an account both of human *capacities*, and also of *relationality*. The two cannot really be separated, as human capacities are manifested in the distinctive relationality of humans. Much as we endorse the view that the human person is constituted by relationality, we believe it is also important to be explicit about the human capacities that are manifest in relationships and that make possible the kind of authentic personal relationships that humans can engage in (Visala 2014, 118).

### The Shift from Individualism to Relationality

Whenever a cross-disciplinary shift in thinking of this kind occurs, as is the case with the relational turn, it is always interesting to enquire why several disciplines are making the same move almost simultaneously. It may be that the move happens first in one discipline and then quickly spreads out to others. There is some truth in that in this case, but it is not clear that it is the whole explanation of this multidisciplinary change. There seems to be something in the *zeitgeist* that makes the relational turn obvious and attractive, or which makes the alternatives seem limited and unattractive. It is rather like evolutionary "convergences," in which the same evolutionary developments arise independently in various different contexts (Conway Morris 2003). There seems to be a somewhat similar cultural convergence in the relational turn in thinking about human nature at the present time.

What was it that led to the relational turn? Part of the answer may be that the highly individualistic theorizing that had preceded the relational turn was proving inadequate and constraining in many areas, and leading to pressures for some kind of paradigm shift. There is more work to be done on the history of individualism in Western thought. Some of the seeds were probably sewn in the Reformation period, with its new emphasis on personal faith. However, that personal faith was still assumed to be socially embedded in the community of faith. The more extreme individualism, sometimes termed "abstract individualism" (Turner 2020), that

tried to take the individual out of social and relational context, came later, in “late modernity,” that is, the latter part of the nineteenth century.

There is a need for balance here, and for a position that avoids extreme individualism, but which also respects the needs of persons for distance and individuation. It is understandable that theological objections have been advanced against extreme individualism, but there are also problems with the other extreme, in which individual identity is submerged in a radical collectivism. Stephen Verney (1976) sets out such a balanced position, framed in theological terms, in which learning interdependence can hold the balance between extremes of individualism and collectivism. He looks toward the development of a “true self” in which each person discovers a true relatedness to themselves, to the other, and to a deeper spiritual reality.

The late nineteenth century, in which radical individualism emerged most strongly, was a period with many bright new ideas that seemed promising at the time, but which turned out to be less serviceable than had been hoped. It was a period of “optimistic simplification,” similar in mood to what occurred in the 1960s (Watts 2016a). There was a new scientism that assumed that progress could be made by bringing into the scientific domain things that had previously been outside it, creating psychological and social science. There was also a new physicalism that prioritized explanations of human phenomena that were framed in terms of the new evolutionary thinking, and which were reductionist in their attempt to provide bottom-up explanations of human phenomena in terms of the body, and especially the brain. Third, there was also a new individualism that assumed that isolated individuals were the building blocks out of which society emerged. This extreme individualism reached its apotheosis in Margaret Thatcher’s infamous remark that “there is no such thing as society.”

Though these new assumptions seemed promising at the time, they did not work out as well as had been hoped. One example that has been studied carefully is the new approach to “emotion” of late modernity, an approach that was scientific, physicalist, and individualistic, replacing the more subtle previous discourse framed in terms of “passions and affections” (Dixon 2003). A classic formulation of the bold new approach to emotion was laid out in a paper by William James (1894). However, as Dixon recounts, James’ new approach to emotion quickly proved too simplistic, and ten years later, he had to salvage his position by publishing such a major revision that it amounted almost to a recantation. Something akin to the earlier distinction between passions and affections eventually had to be reintroduced as a distinction between primary and secondary emotions (Watts 2016b).

The new approach to intelligence of late modernity was highly individualistic. The psychometric intelligence testing that was introduced in the late nineteenth century saw intelligence as a property of individuals,

something that could be taken out of social context and measured. A little later, intelligence was seen as something that an individual inherited, and which was genetically based. As with emotion, this replaced earlier assumptions of intelligence as something collective and transpersonal, in which a person could participate (Barfield 1953). Barfield describes an evolution of human consciousness that has moved away from an initial form of participation toward a more detached kind of thinking, though he hoped to see a return to a different and more self-conscious kind of “final participation” (Barfield 1957).

These new assumptions about intelligence lasted quite some time, but are now breaking down. There has been an increasing recognition of more interpersonal intelligences, such as emotional intelligence, that don't have much to do with general intelligence (Goleman 1996). A new approach to cognition (“4E”) sees cognition as embodied, embedded, enacted, and extended (Newen et al. 2018). The recognition that cognition is socially embedded has far-reaching implications for how intelligence should be conceptualized. AI has largely inherited a set of individualistic assumptions; though there have been challenges to them even in AI, as we will see shortly.

On many fronts, there is thus a retreat from individualism toward a more relational understanding of human nature. That includes both religion and theology, and also cognition and intelligence more generally.

#### AI, ROBOTICS, AND RELATIONALITY

AI is inevitably influenced by the personal and cultural prejudices of the people that create, monitor, and develop it. For the most part, AI has made very individualistic assumptions about intelligence, and largely ignored relational intelligence and social interactions. It has reflected the prevailing modern assumption that intelligence is a property of particular individuals, or devices. The individualistic assumptions about intelligence in the field of AI are defensible as a matter of scientific strategy. Science often has to simplify things, and to limit the scope of its enquiries, in order to make progress at all. Otherwise, scientists find themselves tackling problems that are so diffuse that no progress is made at all. However, once progress has been made on a manageable front, it is possible to shed some of the early restrictions, and move toward wider horizons. There are indications that AI is currently moving in a relational direction.

#### Cybernetics

However, before we come to that, it is worth noting that cybernetics, one of the most significant precursors of AI, was much more contextual in its assumptions about intelligence than AI. Indeed, cybernetics is essentially



a form of systems theory. Rather than building demarcations around the intelligence of a particular individual or device, it was always inclined to consider intelligence in context, and to see a particular individual or device as embedded in a wider context or system. In that sense it makes ecological assumptions about intelligence. As Andrew Pickering puts it, cybernetics “eroded the modern understanding of the bounded, self-contained and self-moving individual,” and moved toward an image of the self as “constitutively bound up with the world and engaged in processes of coupled becoming” (2010, 386).

It is interesting, historically, to note that cybernetics developed at the same time as John Macmurray was developing his relationalist philosophical assumptions in Edinburgh. Norbert Wiener, one of the originators of cybernetics, published his classic text on cybernetics in 1948 (Wiener 1948). Also in 1948, the *Daily Herald* carried a front-page article with the headline, “The clicking brain is cleverer than Man’s,” describing a homeostatic machine built by W. Ross Ashby (Pickering 2010, 1), and in 1952, Ashby published his classic book, *Design for a Brain* (Ashby 1952). The following year, William Grey Walter, another of the first generation cyberneticians, published an influential book on *The Living Brain* (Walter 1953).

It is also worth noting that Cybernetics was much more open to religion and spirituality than AI has so far been. Wiener came from the Jewish tradition and later published a short book on *God and Golem* (1964). Pickering (2010) has drawn attention to the significant interest in eastern spirituality and altered states of consciousness among several of the early cyberneticians, including Ross Ashby, Grey Walter, and Stafford Beer (who began as a high-church Anglican, converted to Catholicism, and later became a teacher of tantric yoga). Beer ventured more explicitly into theology than most cyberneticians, especially in his essay on “Cybernetics and the knowledge of God” (1965). An Anglican Churchman who picked up on that essay and developed it was H. C. N. Williams, Provost of Coventry Cathedral, especially in a series of public lectures given in Illinois in 1966, one of which was on “A Cybernetic Approach to the Christian Faith.” In the Preface, he acknowledges his indebtedness to Norbert Wiener and Stafford Beer (Williams 1967).

Why was there such interest in religion and spirituality, especially eastern spirituality, among these early cyberneticians? Pickering (2010, 385–87) suggests several factors. The most interesting from the present point of view is that cybernetics had a relational understanding of the brain and the self. It was also fascinated by altered states of consciousness, and the possibilities of the brain. Stafford Beer, in particular, had a strong view of the unknowability of the world, which engendered a sense of mystery and connected with his spirituality. In contrast, AI has often espoused rather reductionist positions, such as that the human mind is

“nothing but” a computer program. Cybernetics eroded the distinction between science and religion (Pickering 2010; Modern 2021).

### Relational Trends in AI

Various factors are now leading AI to rediscover some of the contextual approach to intelligence of cybernetics. One key factor is the development of social robotics (e.g., Sheridan 2020). Many robots are created solely to perform mechanical tasks. However, there is increasing interest in developing robots designed for social interaction. That requires relational intelligence of some kind, including knowledge of another intelligent being, and an ability to make use of that knowledge in interacting appropriately with another intelligence. Social robots need something that is functionally equivalent to human empathy, and some kind of equivalent to affect. Rosalind Picard’s work on affective computing is a significant step toward a more relational intelligence (Picard 1995).

It might be imagined that the ability of chatbots such as ChatGPT to chat with people, simulating some kind of interaction with them, is exhibiting relational intelligence. However, that would be a mistake. GPT has no knowledge of its conversation partner. It is just responding to utterances, and is not developing any model of who it is interacting with. There may be an appearance of ChatGPT being interactional, but there is no interactional intelligence involved. Nevertheless, there may be ways in which a computer could engage in interactions with other agents in a way that really did exhibit relational intelligence.

One of the most practically useful forms of relational intelligence may be an automated companion, such as the companions for the elderly developed by Yorick Wilks (2010). He thought it was relatively unimportant whether or not such a companion had a humanlike appearance. An automated companion could be implemented on a phone, which could always be carried by the person concerned, and be constantly available. Wilks’ vision was of a personalized automated companion that would have expert knowledge of the individual concerned, and which would enable it to support the person’s memory, and be a conversation partner.

Wilks used a hybrid methodology, combining both chatbot and programming using discourse theory. Scripts were derived from interactions between two human beings, which were converted to pseudocode, and programmed into the computer. The computer thus had some kind of knowledge of the person it was interacting with, and could deploy that knowledge in a reasonably appropriate way in the interaction. Admittedly, the computer did not fully *understand* its knowledge of the person to whom it was a companion. Nevertheless, in some limited sense, it had interactional knowledge at its disposal. Wilks also considered the possibility of developing an automated spiritual companion (Wilks in press).

Other features would be needed in cognitive modeling of human relational intelligence. It would be important in modeling relational intelligence to have a narrative model with a time-extended structure in which one thing happens after another. Another challenge would be to implement a simulation of the contribution of embodiment to relationality. As Malcolm Jeeves and others have argued, mirror neurons seem to be an important part of the neurological basis for imitation, empathy, and relationality (Jeeves 2011). There is also an increasing understanding of the endocrinological basis of social bonding, in which endorphins, triggered by collective synchronized movement, seem to play an important part (Dunbar 2022). These are complex matters, and not yet well understood; there is more work to be done on how embodiment and interaction with the environment contribute to relational cognition.

Dorobantu (in press b) has identified trends in contemporary AI that point toward an increasing emphasis on relationality. For example, William Clocksin has emphasized the importance of relational intelligence in AI (2003), a view that was probably influenced by his interest in theological anthropology. He claims that the most significant limitations of AI are philosophical rather than technological. There has been much discussion in AI of Moravec's paradox, that is, that it is more difficult to program computers to do mundane tasks than difficult ones. Clocksin suggests that the reason for this is that mundane tasks require relational intelligence, and an ability to relate to other things or persons, which is something that current AI conspicuously lacks.

As Clocksin sees it, AI has been too much influenced by the Aristotelian assumption that human distinctiveness lies in *rationality*. On the contrary, he suggests that human distinctiveness lies more in *relationality* than rationality. If computers are to develop humanlike intelligence, it will be necessary for them to develop not just rationality, but relationality. As a step toward relational intelligence in AI, Clocksin has developed a computational model of friendship (Clocksin, this volume; in press).

Neil Lawrence (2017; forthcoming) has considered the contrast between human intelligence and AI, and reached a related conclusion about the importance of communication in human intelligence. In his view, the fundamental difference between humans and computers lies in communication. He formulates that in terms of a metric called the "embodiment factor," the ratio of an agent's computation to its communication. He claims that humans compute more than machines, but that machines are able to communicate more of what they compute. Lawrence sees communication as one of the key challenges facing humans. Humans know a lot, but what they know is more "locked in" than with computers.

The result is that humans have devoted huge resources to communication, and this has created a pressure for relationality. Machines, in contrast, communicate much more quickly and easily, and so have not faced

this relational pressure. However, as Lawrence sees it, if a machine is to have humanlike intelligence, it will need to incorporate, in some way or rather, the essentially relational intelligence of humans. The communication problems of humans have also led them to construct models of other people (theory of mind), and models of the self. Many aspects of how humans understand themselves may arise from the fact that humans communicate less easily than they compute.

### Human Relational Intelligence

Though AI has generally taken an individualistic approach to machine intelligence some, such as Clocksin (2003), regard this as one of the deficiencies of AI, and wish to see AI develop a more relational approach to intelligence. What would that involve? There is an important distinction between (i) intelligence *about* relationships, where the machine itself is not engaged in interaction with another agent, but is displaying understanding of how agents interact with each other, and (ii) intelligence *in* relationships, where the machine is engaged in interactions, especially in human-computer interactions. The machine can then exhibit knowledge of the person it is interacting with, and use that knowledge to guide its interaction. We will consider each of these in turn.

AI has long experience of modeling interactions. One such area is modeling how weather systems interact. There is an obvious practical value in being able to model that, and meteorological modeling has now become quite advanced and capable of fairly accurate predictions. However, interacting weather systems are not agents, and the modeling of such interactions does not involve anything that can be regarded as social intelligence.

More promising is “multi-agent AI” (MAAI), which models the interactions between agents in a way that is intended to be psychologically realistic. The focus is on the system that arises from multiple interacting agents, and on what the system is capable of. In MAAI, the agents are sometimes intended to be humans, though they might equally well be passive agents without goals. In the study of religion, there is now a considerable body of work on modeling of religious networks, formed from the interaction of agents who form religious networks (e.g., Lane 2021; Shults and Wildman in press).

Clocksin’s computational simulation of the quasi-friendships between agents is also a simulation of interaction, but it uses Affinity, a computer program that simulates interaction and friendship formation between agents, and the focus is on multiple interacting dyads (Clocksin this issue; in press). His system simulates how individual agents engage in relationships with other agents in a way that reflects their needs and values. It models individual differences between agents, and includes certain mechanisms (e.g., narrative structure) and information (e.g., remembering

encounters) that need to be present in order for there to be a simulation of a “relationship.” Each agent has a cognitive architecture and operates as a cognitively sophisticated individual.

Developing a computer simulation of the interactions between agents requires assumptions to be made explicit that may otherwise not be specified, and that is in itself a contribution to knowledge. Experiments can be run in which different values are given to certain variables to test which set of values produce patterns of interactions that are most like those that are observed between humans. The value of such modeling lies in making a theoretical contribution to the social sciences. It provides a precise way of specifying different kinds of interaction.

For example, the particular characteristics of spiritual friendships could be simulated through choosing appropriate values in the interactional program; that would make a theoretical contribution to the understanding of spiritual friendships. In simulating spiritual friendships, the model assumes that certain needs and values, such as universalism and belongingness, are particularly relevant to spiritual friendship. These are variations in how agents interact, not a completely different kind of interaction.

Though the computer simulation incorporates knowledge of how interactions work, it is questionable whether the computer is engaged *in* interactions. It might be argued that the program represents intelligence in relationships, because each agent uses its knowledge to engage in relationships. Agents are not engaging in relationships with humans, but they are engaging in relationships with a whole population of other agents, so there are societal effects that can be observed. Clocksin discusses his approach to relational intelligence in the following article in this section.

### Robots and Personhood

The other two articles in this section focus on the question of whether robots could be persons. The problems in finding a definitive approach to this question leads Michael Reiss to turn the question on its head, and to ask if there are any reasons why robots should *not* one day become persons.

He approaches that from various perspectives. From an evolutionary perspective, he points out that biologists dislike oversharpe binaries, such as the assumption that humans are persons, but other species are not persons in any way at all. From a developmental perspective, he points out that neonates initially lack the capacities associated with personhood, but may nevertheless be accepted as persons. From a chemical perspective, he questions whether there are any good reasons for insisting that persons must necessarily be carbon based. From a historical point of view, he points out that it is not so long since some humans were considered to be properties rather than persons.

Reiss draws attention to the automaton in the form of a monk, dating from the sixteenth century, that can raise a cross and rosary to his lips, and move his lips as though uttering silent prayers, which may have been considered to be efficacious. He also draws attention to the philosophical position of panpsychism, which challenges the assumption that mechanisms cannot have minds.

Leon Turner considers further the complex issues raised by the question of whether robots are persons. One approach would start from human beings as the only example of persons that we currently have. The question then becomes whether robots are like humans. That is not as straightforward as might be imagined. There are already quite a number of gray areas and, as technology advances, there are likely to be even more. There are also conceptual issues about what kind of resemblance to a human being is required. Is it having a human body? Or having the functional capacities that humans have? And, if so, which capacities are particularly critical for a robot to qualify as persons?

Then, following the relational turn in theological anthropology, Turner suggests that we might take a relational approach to the question of whether a robot counts as a person. Rather than deciding the question on the basis of whether a robot resembles a human, we might decide it on the basis on whether or not it has relationality, and a narrative sense of identity. This could potentially expand the range of what (or who) could be counted as a person. However, there will still be various borderline cases, which will not be easily settled, as it is not clear what the decisive tests ought to be.

Digging deeper, Turner points out that in theological anthropology, there are various different theoretical views about the nature of personhood and where it comes from, and he considers the different views of John Zizioulas, David Kelsey, and Noreen Herzfeld. However, as each of these views is still the focus of discussion, and is not universally accepted, none of them really helps in settling the question of who should count as a person. In the end, Turner suggests, we may have to fall back on the hope that we will know a person when we meet one.

## REFERENCES

- Ashby, W. Ross. 1952. *Design for a Brain*. New York: Wiley.
- Barfield, Owen. 1953. *History in English Words*. London: Faber and Faber.
- . 1957. *Saving the Appearances: A Study in Idolatry*. London: Faber and Faber.
- Beer, Stafford. 1965. "Cybernetics and the Knowledge of God." *Month* 34:291–303.
- Clarke, Graham S. 2006. *Personal Relations Theory: Fairbairn, Macmurray and Suttie*. Abingdon, UK: Routledge.
- Clocksins, William F. 2003. "Artificial Intelligence and the Future." *Philosophical Transactions of The Royal Society* 361 (1809): 1721–48.
- . In press. *Computational Modelling of Robot Personhood and Relationality*. New York: Springer.

- Conford, Philip, and Tony Blair. 1997. *The Personal World: John Macmurray on Self and Society*. Edinburgh: Floris Books.
- Conway Morris, Simon. 2003. *Life's Solution. Inevitable Humans in a Lonely Universe*. Oxford: Oxford University Press.
- Cortez, Marc. 2010. *Theological Anthropology: A Guide for the Perplexed*. Edinburgh: T&T Clark.
- Dixon, Thomas. 2003. *From Passions to Emotions: The Creation of a Secular Psychological Category*. Cambridge: Cambridge University Press.
- Dorobantu, Marius. 2021. "Human-Level, but Non-Humanlike: Artificial Intelligence and a Multi-Level Relational Interpretation of the Imago Dei." *Philosophy, Theology and the Sciences* 8 (1): 81–107.
- . 2022. "Imago Dei in the Age of Artificial Intelligence: Challenges and Opportunities for a Science-Engaged Theology." *Christian Perspectives on Science and Technology (New Series)* 1:175–96.
- . In press a. *Artificial Intelligence and the Image of God: Are We More Than Intelligent Machines?* Cambridge: Cambridge University Press.
- . In press b. "AI and Christianity: Friends or Foes?" In *Cambridge Companion to Religion and AI*, edited by Beth Singler and Fraser Watts. Cambridge: Cambridge University Press.
- Dunbar, Robin. 2022. *How Religion Evolved: And Why It Endures*. London: Pelican Books.
- Fergusson, David, and Nigel Dower, eds. 2002. *John Macmurray: Critical Perspectives*. Lausanne: Peter Lang Publishing.
- Goleman, Daniel. P. 1996. *Emotional Intelligence: Why It Can Matter More than IQ*. New York: Bantam Books.
- Grenz, Stanley. 2001. *The Social God and the Relational Self: A Trinitarian Theology of the Imago Dei*. London: John Knox Press.
- Harré, Rom. 1983. *Personal Being: A Theory for Individual Psychology*. Oxford: Basil Blackwell.
- Harré, Rom, and Paul F. Secord. 1973. *The Explanation of Social Behaviour*. Oxford: Basil Blackwell.
- James, William. 1894. "What Is an Emotion?" *Mind* 9 (34): 188–205.
- Jeeves, Malcolm. 2011. *Rethinking Human Nature: A Multidisciplinary Approach*. New York: Wm. B. Eerdmans.
- Lane, Justin E. 2021. *Understanding Religion through Artificial Intelligence: Bonding and Belief*. London: Bloomsbury.
- Lawrence, Neil. 2017. "Living Together: Mind and Machine Intelligence." arXiv:1705.07996.
- . Forthcoming. *The Atomic Human*. London: Penguin Books.
- Leidenhag, Joanna. 2020. "The Challenge of Autism for Relational Approaches to Theological Anthropology." *International Journal for Systematic Theology* 23 (1): 109–134.
- Lossky, Vladimir. 1974. *In the Image and Likeness of God*. New York: St Vladimir's Seminary Press.
- Macmurray, John. 1935. *Reason and Emotion*. London: Faber and Faber.
- . 1957. *The Self as Agent*. London: Faber and Faber.
- . 1961. *Persons in Relation*. London: Faber and Faber.
- McFadyen, Alistair. 1990. *The Call to Personhood*. Cambridge: Cambridge University Press.
- McIntosh, Esther. 2011. *John Macmurray's Religious Philosophy: What It Means to Be a Person*. Abingdon, UK: Routledge.
- Modern, John L. 2021. *Neuromatic: Or, A Particular History of Religion and the Brain*. Chicago: University of Chicago Press.
- Newen, Albert, Leon de Bruin, and Shaun Gallagher, eds. 2018. *Oxford Handbook of 4E Cognition*. Oxford: Oxford University Press.
- Parsons, William B. 2021. *Freud and Religion: Advancing the Dialogue*. Cambridge: Cambridge University Press.
- Picard, Rosalind W. 1995. *Affective Computing*. Cambridge MA: MIT Press.
- Pickering, Andrew. 2010. *The Cybernetic Brain: Sketches of Another Future*. Chicago: University of Chicago Press.
- Sheridan, Thomas B. 2020. "A Review of Recent Research in Social Robotics." *Current Opinion in Psychology* 36 (December): 7–12.

- Shults, F. LeRon F., and Wesley J. Wildman. In press. "Simulating Religion." In *Cambridge Companion to Religion and AI*, edited by Beth Singler and Fraser Watts. Cambridge: Cambridge University Press.
- Turner, Léon. 2008. *Theology, Psychology and the Relational Self*. Farnham, UK: Ashgate.
- . 2012. "Science, Theology and the Relational Self." In *Science and the World's Religions. Vol 2: Persons and Selves*, edited by Wesley Wildman and Patrick McNamara, 61–90. Santa Barbara, CA: ABC/CLIO Press.
- . 2020. "Isolating the Individual: Theology, the Evolution of Religion, and the Problem of Abstract Individualism." *Zygon: Journal of Religion and Science* 55 (1): 207–28.
- Visala, Aku. 2014. "Imago Dei, Dualism, and Evolution: A Philosophical Defense of the Structural Image of God." *Zygon: Journal of Religion and Science* 49 (1): 101–20.
- Verney, Stephen. 1976. *Into the New Age*. London: William Collins.
- Walter, William Grey. 1953. *The Living Brain*. London: Gerald Duckworth.
- Watts, Fraser. 2016a. "Optimistic Simplification: Common Elements in Behaviour Therapy and Psychiatric Rehabilitation in the 1960s." *History and Philosophy of Psychology* 17 (1): 24–29.
- . 2016b. "Self-Conscious Emotions, Religion and Theology." In *Issues in Science and Theology: Do Emotions Shape the World?*, edited by Dirk Evers, Michael Fuller, Anne Runehov, and Knut-Willy Sæther, 201–10. Cham, Switzerland: Springer.
- . 2020. "The Evolution of Religious Cognition." *Archive for the Psychology of Religion* 42 (1): 89–100.
- Watts, Fraser, and Marius Dorobantu. 2023. "Shamanic and Doctrinal Religion: Applying Dunbar's Approach to Contemporary Religion." *Religion, Brain & Behavior*. <https://doi.org/10.1080/2153599X.2023.2168733>
- Wiener, Norbert. 1948. *Cybernetics: Or Control and Communication in the Animal and the Machine*. Cambridge, MA: MIT Press.
- . 1964. *God & Golem, Inc.: A Comment on Certain Points Where Cybernetics Impinges on Religion*. Cambridge, MA: MIT Press.
- Wilks, Yorick. 2010. "Introducing Artificial Companions." In *Artificial Companions in Society: Scientific, Economic, Psychological and Philosophical Perspectives*, edited by Yorick Wilks, 11–20. Amsterdam: John Benjamins.
- . In press. "Artificial Companions and Spiritual Enhancement." In *Cambridge Companion to Religion and AI*, edited by Beth Singler and Fraser Watts. Cambridge: Cambridge University Press.
- Williams, H. C. N. 1967. *Nothing to Fear*. London: Hodder & Stoughton.
- Williams, Rowan. 2018. *Being Human: Bodies, Minds, Persons*. London: SPCK.
- Zizioulas, John. 1985. *Being as Communion: Studies in Personhood and the Church*. Crestwood, NY: St Vladimir's Seminary Press.