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."

GUIDELINES FOR COMPUTATIONAL MODELING OF FRIENDSHIP

by William F. Clocksin

Abstract. Humans participate in an immense variety of relationships with other persons and other entities: human and nonhuman, living and nonliving, tangible and intangible, real and imagined. Participation in relationships is considered a key benchmark of personhood. Some of these relationships, particularly friendships, involve close emotional attachments, and some friendships have been described since antiquity as spiritual in nature. Different types of friendship depend upon factors such as proximity, social formality, physical intimacy, information exchanged, and the costs and benefits of maintaining the relationship. There are time-extended processes and narrative practices involved in forming and dissolving relationships. A question is raised how androids (hypothetical humanoid robots that people would accept as equals in society) can participate in friendships with humans and other entities. This article explores the space of friendships with the aim of formulating guidelines for a computational model that can make explicit the information processing requirements and step-by-step processes involved with participating in the many different types of friendships, including those known as spiritual friendships.

Keywords: artificial intelligence; cognitive science; computational sociology; computer modeling

Introduction

It is important to make clear what kind of relationships are to be considered here. Relationships between people in society are widely experienced, and studied by anthropologists, psychologists, sociologists, philosophers,

William F. Clocksin is Emeritus Professor of Computer Science, of the University of Hertfordshire. He is retired and living in Cyprus; e-mail: william.clocksin@cantab.net.

ISSN 0591-2385

and theologians. Most studies of personhood and relationality assume that relationships are formed between two humans. Let us call these HH relationships (human-human). When robots are considered to take part in relationships, they can be called human-robot (HR) relationships, in which the human forms a kind of relationship with a robot that is so different from a human that the HR relationship is possible, but the human does not experience the relationship as having the same quality as an HH relationship. Most robot-related philosophy (de Graaf 2016) and more recent ideas about authentic friendship between humans and robots (Danaher 2019; Ryland 2021) are about relationships that fall into what we call the HR category. However, we can also consider hypothetical robot-human (RH) relationships, in which appropriately programmed robots engage in relationships with humans who are content to accept the relationship as equal enough in quality to pass as an HH relationship, and robot-robot (RR) relationships, in which two appropriately programmed robots engage in relationships that humans would identify as equal enough in quality to pass as an HH relationship. HR and HH relationships do not exist, but future research in artificial intelligence (AI) may result in the type of robot that can engage in such relationships. Theologians also study HH and human-Divine relationships within the framework of the created order (e.g., Barth 1960), and psychologists and sociologists study spiritual relationships including situations where humans experience relationship with imaginary or implied entities. There is also a sense in which human-like relationships between imaginary/implied entities can be understood, such as between members of a mythological pantheon. Our treatment here is to consider the most general case, in which humans, appropriately programmed robots, and imaginary/implied entities are all considered as agents who can participate in relationships with other agents: the AA relationship.

We begin with the understanding that human and person are distinct concepts: All humans are persons, but not all persons are human. What do robots need to do if they are to be accepted as persons in human society? An *android* is defined here as a human-like robot that people would accept as passing as humans in how they perform and behave in society. An android is not considered to be imitating a human, nor is its purpose to deceive humans into believing that the android is a human. Instead, the android has the capability to self-identify as a nonhuman with its own integrity as a person. Androids as defined in this way do not yet exist, but the idea cannot be excluded on technical grounds, and further research in AI may someday culminate in a functioning android. A key question therefore is how the android may operate fluently in human society as a socialized person. We accept that there is also a metaphysical question of whether androids can function as persons in society at all. For the purposes of this article, we set that open question aside and instead consider what would be needed in a computational model that can be used to simulate how agents engage in relationships. A fuller computational understanding might be the basis of a model that future androids employ to conduct their participation in the most general type of agent-agent relationships. Relationships, whether between humans or other agents, involve the exchange of information. A computational model attempts to make explicit the step-by-step exchange and processing of information to better understand how relationships may be formed, maintained, and dissolved. This is not a model of human relationships, though it will have been informed by what happens in human relationships, but it is a model that involves agents in relationship.

Specialized areas of AI research have investigated some aspects of robots in society. In the area of android science (MacDorman 2006), the main ideas are to do with the physical appearance of androids, and how people react affectively to entities that resemble people in their physical appearance. The area of social robotics (Dautenhahn 2007) is another approach to social intelligence, where research focuses on the social skills needed by robots for comfortable human-robot interaction (HRI). Such social skills include respect for personal space and the physical dynamics involved with interaction.

While physical appearance and social skill can mediate rapport and emotional connection, the definition of android considered in this article is not about appearance and physical dynamics, but about meaning in the relationships in which the android participates, and about how participation in meaningful relationships can be a mark of personhood. Examples of androids are provided by science fiction. In the popular *Star Wars* films, the humanoid robot C-3PO is considered an android by the definition used here. The robot C-3PO is obviously artificial in appearance, with a metal body, nonhuman eyes, and immobile face, but it is fluent in spoken communication and employs its humanoid body in ways meaningful to social discourse. It engages in relationships with the humans around it to the extent that its mechanical appearance is considered inconsequential by the other characters portrayed in the films. Other fictional androids, such as Mr Data in the popular *Star Trek: The Next Generation* television series, resemble people to an extent both in appearance and in the quality of relationships, and would be considered an android by any definition.

Personhood and Relationality

The two interrelated characteristics of the android as defined here are personhood and relationality. Clocksin (2003) proposed a conceptual framework for AI that gives priority to the social construction (Gergen 1994) of an identity (Shotter and Gergen 1989; Clocksin 1998) or "self" by engagement with social relationships. This framework suggests that intelligence is constructed through a capacity for relationality that in turn is based upon a capability for affect and experiencing felt meaning (Clocksin 2005). In their studies of robot personhood, MacDorman and Cowley (2008), and later Barresi (2020), suggest that a compelling benchmark of human personhood is the ability to sustain long-term relationships. The term benchmark does not mean criterion. There is a wide range of human persons who cannot or do not engage in long-term relationships for various reasons. They are no less persons, and the benchmark does not suggest otherwise, as humans are persons by default. Taking that idea as a starting point, one motivation here is to work toward a computational understanding of personhood and how androids may be enabled to participate in long-term relationships and thereby develop android personhood.

We consider personhood to be a performance and not an ontological category. It is important to stress that humans are considered persons by default by moral convention, regardless of their individual abilities. Performance may be an unfamiliar way to describe personhood. A performance suggests it takes place for a specific reason, in this case, to perform personhood within social interaction. Using the term "performance" comes with some risks. First, it can imply that the act of performing personhood may be insincere, not authentic, or deceptive: It is not a "real" personhood, merely a performance. Second, it might imply that performance is a conscious, knowing, or intentional act. We do not use the term in those two ways. A performance of personhood may have some conscious elements, but it will derive from preconscious cognition. Finally, performance is a concept more easily used in connection with personhood in nonhuman entities, where "behaviour" is more commonly associated with humans and other animals.

This idea of personhood includes humans, androids, and imagined persons. Humans are persons by default, and androids are appropriately programmed to perform a recognizable personhood. Imagined entities have personhood by implication, but they cannot not perform personhood in a physical way. They are recognizable as persons through accounts in literary works, memory, and imagination, and through performance by humans who interpret the implied personhood of imagined entities in art forms. In this way, persons can also perform as proxies of imagined entities.

Humans are able to recognize personhood in imagined entities. This capability may be connected to the fact that human relationships are not always directly observable. As humans we assume that other humans are engaged in performance of personhood even when we cannot observe them doing so. The ability to make this kind of assumption may have the additional effect of being able to recognize personhood in nonhuman entities or in literary works.

Humans participate in an immense variety of relationships with other persons and other entities: human and nonhuman, living and nonliving, tangible and intangible, real and imagined. Relationships form as a way to satisfy basic needs and desires, and may be mediated by an internal reward system. Some of these relationships can form around shared affinities or preferences, for example, a liking for a particular brand of beer or following a particular football team. Relationships may form because people are placed into proximity due to shared values. Relationships may form for reasons of business or other shared activity. Relationships may involve close emotional attachments to people, other animals, historical figures real or fictional, characters from literary works, and imagined entities. In ordinary discourse, we name different types of relationships. We have friends, colleagues, partners, acquaintances, lovers, relatives, spouses, pets. There are different subtypes within these types that depend on factors such as proximity, social formality, physical intimacy, and the costs and benefits of maintaining the relationship. There are time-extended processes involved in forming and in dissolving relationships. In this context, the aim of a computational model is to make explicit the information processing requirements and step-by-step processes involved with participating in one or more different types of relationships.

Another motivation of our work is an exploration of spiritual intelligence. We define spiritual intelligence as the capability of a cognitive entity—whether biological or artificial—to reason and act according to their significant concerns and the significant concerns of those in which they are in relationship. Significant concerns include attitudes, preferences, affinities, and values that are held to be highly valued and meaningful. Significant concerns provide a way for a person to explore needs and desires connected with self-actualization, self-transcendence, and belonging. Through significant concerns, a person may find deeply held identity, purpose, and transformation. Since antiquity, some relationships have been described as spiritual relationships. Certain significant concerns are employed in the spiritual relationships that involve intangible and imagined entities, because the particularities of such entities are held to be greater than that achievable by a human. Also, some HH relationships are also described as spiritual relationships because of the significant concerns around which the relationship is defined. The notion of spiritual intelligence used in this article is more limited than that of Vernon (2022), who writes of spiritual intelligence as the human capacity for being aware of and feeling connected with a greater reality and ground of being. This conception of spiritual intelligence involves the wholeness of human capabilities for affect and for holding a felt experience without necessarily understanding it in rational terms. This wider sense of human experience is worth further study, but is outside the scope of this article.

Friendship

Friendship is one of the most widespread and well-attested types of relationship between humans, and computational models of relationship can be usefully informed by characteristics of HH friendships. Dunbar (2018) defines friends as the people who share our lives in a way that is more than just the casual meeting of strangers. Friends make efforts to maintain contact with each other, and they feel emotional bonds. Dunbar notes there are important differences between friendships, kinships, and romantic relationships, but these types of relationship have meaning in an emotional sense, and they provide social and emotional benefits that other relationships—such as between strangers, casual acquaintances, and business partners—do not.

Studies of friendship can be traced back to antiquity, and it is useful to explore these studies for what they may or may not offer to a model of friendship that androids can use. Our purpose is not to model HH friendship, nor is it to ask how HR friendships may form. Instead, we look at a range of human friendships to see what general information-processing patterns can be learned from them to model AA relationships.

One milestone is Cicero's *Laelius de Amicitia* of 44 BCE. For Cicero, friendships are founded on agreement in all matters of importance, plus goodwill and affection. Furthermore, friendships only exist because of virtue. The idea that friendship can only exist between those who are good and virtuous can be traced back to Aristotle (*Nicomachean Ethics*, Book VIII, 350 BCE), and Danaher (2019) bases his treatment of virtue friendship within what we call HR relationships on Aristotle. Cicero acknowledges that we come into contact with many good people we call our friends, such as business associates, neighbors, or any variety of acquaintances. But he makes a key distinction between these common and useful attachments and the rare friends with whom we connect on a deeper level. These are useful distinctions to build into a computational model.

A second milestone on friendship is Aelred of Rievaulx's *Spiritual Friendship* from circa 1160. Aelred based his work of three books on his reading of Cicero and other classical treatises, and he follows Cicero in holding that friendship can endure only among the good. Aelred wrote from a Christian monastic context, where friendships may form, but which are regulated by the rules of the monastic order. For Aelred, friendship begins with an attachment resulting from shared goals, and follows Cicero in that friends share the same view on everything human and divine (1.13). He distinguishes between different kinds of friendship (1.38): carnal friendship, based on shared pursuit of pleasure; worldly friendship, based on mutual advantage; and spiritual friendship, grounded in shared discipleship in Christ.

For Aelred, friendship can start with an attachment, but can only develop into friendship when it is guided by reason, moderated by honesty, and ruled by justice (2.57). In the monastic setting, physical intimacy was prohibited and considered sinful. Therefore, Aelred wrote of spiritual friendship, which has its origins in "the purity of intention, the teaching of reason, and the rein of temperance" (2.59), and which must not be "prone to the desires of the flesh" (2.58). Aelred admits that a kind of friendship can exist, "in which partners of the worst immorality become fast friends," but he deems this not worthy of the name of friendship (2.59). While it is not the intention here to base a computational model on a medieval monastic study of spiritual friendship, it is useful to consider the diversity of how friendships form within the parameters of an agent's disposition and the constraints of the environment.

It is commonly understood that some friendships can be considered "close" and other friendships can be considered more "distant." Modern writers on friendship describe the closeness of friendship in terms of "levels" or "layers" and relate this to the size of friendship groups. In what is the most comprehensive study of friendship yet, Dunbar (2021) represents layers of friendship as a series of nested circles, where the approximate size of each layer includes the layers within it. Working out from the innermost circle, the smallest layer might include one or two *intimate friends*, then about five *close friends* (including the intimates), about fifteen *best friends* (including intimates and close friends), about fifty *good friends*, and about 150 "just friends." Dunbar has assembled persuasive evidence for the evolutionary, cognitive, and biochemical pressures that have formed this structure.

In the popular self-help literature, Shaw (2021) has set out a structure in terms of four "levels" of friendship defined by "boundaries," which mark acceptable topics of conversation within levels of friendship.

(1) *Essential friends* are confidantes and people who share one's closest values, and there are few if any boundaries of conversation between essential friends.

(2) *Collaborators* are friends with whom there is an emotional connection at a specific time in life. Collaboration requires physical proximity and emotional immediacy, and the relationship may end if the purpose of the collaboration comes to an end.

(3) *Associates* are friends with whom there is a connection through work or a common interest such as a hobby. Associates provide sociability with the context of a shared interest.

(4) Shaw describes *Mentors and Mentees* as the fourth level of friendship. The mentor/mentee relationship is often work related or therapy related, and therefore there is an asymmetrical balance of power and control in mentor/mentee relationships that may be regulated by professional standards.

It appears from the above sources that friendships are essentially about similarities, such that friendships form when people holding similar concerns are in proximity. This can be related to a key concept in social science known as homophily, the tendency to associate with similar others. While homophily is a well-attested pattern underlying human relationships, we agree with the proposal of Lawrence and Shah (2020) that homophily should be understood not simply as a pervasive pattern that can be described statistically, but as a performance within the meaning-making context of relationships. Instead, we seek to model friendship in terms of the meaning that persons derive from engaging in the relationship. For example, close friendships can form around a few similarities, and yet continue despite a few differences, but can dissolve if differences become intolerable. Good friendship can involve respecting differences, but other differences may be "red flags" that can signal the end of a friendship even before it has started. Therefore, a model of friendship should define how similarity or compatibility or "agreement" (in Cicero's terms) is evaluated in a way that works with meaning within the lived experience of people. A person's evaluation of similarity may change during its life, and friendships therefore have their own life-cycle as friends experience personal growth and as the circumstances bearing upon the friendship may change over time. For example, friendships may be formed in one of Dunbar's layers but move to another layer, and friendships may form in one of Shaw's levels and move to another.

Physical proximity was essential for developing friendship in most of human history. However, in recent centuries, correspondence by letter, and more recently electronic communications, have provided a proximity from which friendships can emerge. Some friendships can endure losses of proximity over a period of time, and some cannot. A model of friendship should take into account these additional factors of proximity and how it has an effect upon the life-cycle of a friendship. Similarly, humans have developed relationships with nontangible or imaginary characters, originally through storytelling, and later in works of literature. Such relationships can be marked by emotional closeness and endurance, and the characters take on a kind of personhood, even though the characters do not perform personhood outside the imagination of the listener or reader.

The shared concerns around which people form friendships can be explored further. The things around which persons form relationships are called social objects. Social objects do not need to be tangible objects, but could include a shared activity such as a business venture, a shared affiliation such as supporting a football team, a shared preference such as a favorite color, and a shared event such as a meal or ritual. One or more social objects may be involved in a relationship between two entities. Following Moscovici (1984), elementary social situations can be modeled as the social triad consisting of the Ego (or Self), the social Object, and the Alter (or Other). In this model, the three vertices of the triad influence and are influenced by each other. Moscovici's triad is an extension of Allport's

(1954) classic definition of social psychology as "an attempt to understand and explain how the thought, feeling, and behaviour of individuals are influenced by the actual, imagined, or implied presence of others," in that the social object included in the triad model Self-Object-Other becomes fundamental for social psychology. The nature of social objects continues to be a topic of current research (Hindriks 2020).

The Agent's Disposition

Having considered the concept of friendship, we will now consider factors that are necessary for constructing a computational simulation of friendship.

In addition to social objects, friendships depend upon a rich substrate of values and attitudes such as generosity, tolerance, and forgiveness. While social objects, values, and attitudes are considered abstract concepts or concerns that are associated with a society, individual agents instantiate or "hold" a set of concerns in their internal value systems. In addition, each agent has what we describe as an internal economy that contains quantities that fluctuate according to events. The implemented simulation uses four quantities, called energy, stress, reward, and attach, which are analogous to the human hormones adrenaline, cortisol, endorphin, and oxytocin, respectively. We stress that this is not a model of the human endocrine system, nor are the quantities of the economy models of human hormones. The quantities in the economy are regulated according to events in which the agent engages, and both influence the agent's behavior and are influenced by it. At any given time, an agent will "hold" a set of concerns (values, attitudes, social objects) and an internal economy that we call its *disposition*. Each element of the disposition has three parameters associated with it: importance, degree, and intensity. These parameters take on values between 0 and 1. The importance is the amount of concern the agent reports to others; the degree is the amount the concern influences the agent's decisions; and the intensity is the amount that governs the amount of action the agent takes relating to the concern. For example, an agent may report that they have high regard for honesty in society (high importance), yet they may engage in cheating when it suits them (low degree), and are lazy in doing so (low intensity). While this is not an attempt to model human behavior, these three parameters provide for subtlety in the agent's behavior in the simulation, and may partly account for the complexity, inconsistency, and paradox observed in human behavior. Parameters are further discussed below.

A computational model of friendship should represent additional characteristics of social objects that are not covered in the literature. Most treatments of friendship assume that friendships are based upon a similar affinity for a given social object: Both persons support the same football team for example. However, this definition of affinity is not rich enough to model the wider reality of friendships. For example, using favorite color as a social object, suppose person A's favorite color is red, and yet A is willing to form a friendship with person B whose favorite color is blue, and another friendship with C whose favorite color is orange. Friends can have dissimilar favorite colors. This does not imply that favorite color is not a social object, and it does not mean that a friendship between A and B is somehow defective. It means that A has an opinion of its own favorite color, but it is able to accept as friends others with different favorite colors. This means that the usual model of affinity as "agreement" should be augmented to include two factors about social objects, as follows. Using the example of favorite color, entities can have a *precept*, which is a statement about their own favorite color, and they can have an *accept*, which is a statement about the possible favorite color of other entities. For example, an entity may have a particular favorite color (represented by its *precept* statement), but is able (represented by its *accept* statement) to become friends with entities having dissimilar favorite colors.

A model of friendship needs also to represent the fact that friendship is asymmetric and experienced with imperfect knowledge. Person A may have more affection for Person B than B has for A, and Person A may have more affinity for social object S than B has for S, and yet A and B can be firm friends. Persons A and B cannot read each other's minds, so A may not know the amount of affection or affinity that B has for A, nor A for B. Even if A and B communicate these amounts to each other (reporting what we call the importance parameter), the communication may be defective, not understood, or misunderstood. The degree of closeness of the friendship will be a factor with both asymmetry and imperfect knowledge, and even the degree of closeness may be experienced asymmetrically. Asymmetry is also a feature of certain relationships such as mentor/mentee previously described.

Human values have been represented in a variety of ways, ranging from lists of hundreds of words that describe core values, to the VIA-IS model of 24 strengths (Peterson and Seligman 2004). We have adopted some aspects of Schwartz's (Schwartz, 2012) system of basic values (BVs), although other proposed value systems (e.g. Almquist, Senior, and Bloch 2016) are comparable. By means of extensive cross-cultural surveys, Schwartz has identified ten BVs: self-direction, stimulation, hedonism, achievement, power, security, conformity, tradition, benevolence, and universalism. An agent may hold each of these values to a greater or lesser degree. For the purposes of this article, BVs can function not only to influence the life-cycle of a friendship, but may also function as social objects. For an example of BVs as social objects, a friendship may form between agents who have adopted a similar degree of hedonistic lifestyle and who therefore might encounter each other in venues where such lifestyles are expressed.

Similarly, agents who hold fast to tradition and conformity may encounter each other in social groups where tradition and conformity are valued. As for the use of BV parameters, the degree of benevolence possessed by an agent can regulate its motivation to offer care to a friend (in Swartz's terms, a member of its in-group), and the degree of universalism possessed by an agent can regulate its motivation to offer care to a nonfriend (in Swartz's terms, a member of its out-group).

For the purposes of computational modeling, each BV has three parameters (importance, degree, intensity) "held" by each agent as described above. Each parameter can take on a quantity that ranges from 0 to 1 inclusive, where 0 represents the complete absence of the parameter, and 1 represents a full commitment to the parameter. For example, a conformity degree value of 0.5 might represent an average amount of conformity, and an agent with conformity degree value 0.7 would be expected to perform in a more conformist way by comparison with the average. While pinning down BV parameters in this way may seem naïve and oversimplified, representing parameters as numerical quantities is commonplace in computational modeling. Also, we prefer to use the range from 0 to 1 inclusive because it can be interpreted as the probability of holding the BV, and because of technical advantages in using nonnegative values in computations (Lee and Seung 1999). In addition, the model represents an additional dimension analogous to the *precept* and *accept* factors of social objects. This additional dimension arises from the observation that people will use a BV to express their own values, but will also use a BV to express their expectation of the values of others. For example, a person may hold to a high degree of tradition itself, but can tolerate others for whom tradition is not important. Both uses are important, and can be modeled using the *precept* and *accept* factors.

Spiritual Relationships

This section looks at the special case of spiritual and "soul" friendship for two specific reasons. First, it shows the diversity of relationships commonly engaged in by humans. Second, such relationships suggest connection at a level deeper and more personal than mere affinity with other humans over shared social objects. This deeper and subjective connection tends to evade description in terms of BVs, but it is possible that agents who are motivated to seek spiritual connection can be represented as holding a high degree of both security and universalism.

Since antiquity, the term "spiritual" has been used to identify a particular type of relationship, and Aelred's use of the term spiritual friendship is a paradigm example. In Book 3, Aelred sets out the process for discernment of spiritual friends: The foundation of spiritual friendship is the love of God (3.5), but reason and affection are tied together so that "love may be chaste through reason, and delightful through affection" (3.2).

The term "spiritual" can be problematic because of its diverse meanings that depend upon history and culture. Although the term had a profoundly rich meaning in early days of Christianity, the term has become transformed into a synonym of terms such as incorporeal, immaterial, or supernatural; or has become related to concepts such as religiosity, subjective well-being, meaning in life, and paranormal beliefs. While for Aelred, spiritual friendships were intrinsically grounded in Christ, it is possible to define spiritual friendships in terms that do not imply a religious connection.

An early version of Schwartz's value theory (Schwartz 1992) considered the possibility of using spirituality as a BV, where a spirituality value expresses meaning, coherence, acceptance of situation in life, and inner harmony through transcending everyday reality. However, despite the importance of spirituality across cultures, Schwartz found that spirituality did not demonstrate a consistent meaning across cultures, so spirituality was not included in the set of BVs.

Aelred discusses what we may call the life-cycle of a spiritual friendship. A friendship is formed in four stages: The first is choosing a friend, the second is testing of the friendship, the third is an acceptance of the friendship, and the fourth is "highest agreement in all things divine and human with a certain charity and goodwill" (3.8). Aelred assumes that spiritual friendships should be steadfast, and should not happen "on a passing whim" (3.7), because they "present an image of eternity" (3.6). This might imply that spiritual friendships cannot be dissolved. Nevertheless, Aelred admits that friendships (or relationships thought to be friendships) may be "so wounded as to perish" (3.21) for reasons of slander, reproach, pride, betrayal of secrets, and treachery (3.23). Aelred also mentions behaviors, which are undesirable, but should not end a friendship, such as showing anger or speaking a bitter word (3.22).

From a modeling viewpoint, the life-cycle of a friendship will be influenced by the social objects and behaviors encountered during the stages of a friendship, and also by the disposition of the agents. Whether a friendship is sustained or is dissolved will depend upon the degree of the BVs held by the friends, and upon the degree of attitudes also represented as parameters such as forgiveness and tolerance.

Recently, spiritual friendship is being discussed in the context of human relationships and sexuality. Contributors to the website spiritualfriendship.org discuss whether or not spiritual friendship may be used as a model for celibate same-sex attracted friendships. Setting aside here the contested meanings and implications of terms such as same-sex attracted, and differing interpretations of celibacy and chastity, the key difference in viewpoint is whether the notion of spiritual friendship can be generalized and applied in a modern secular context, or whether it necessarily refers to Aelred's usage in the context of a Christian community for whom faith and spiritual values are important. While this is not directly relevant to a computational model, it indicates aspects of the human experience that can inform the model.

The idea of a *soul friend* is another attempt to understand the idea of a close relationship that has a spiritual element. Writing from the Christian tradition, Leech (1980, 2001) uses the term to refer to the activity of spiritual direction, which is about developing a deeper awareness of the spiritual aspect of being human. A spiritual director is a person who encourages the process of reflection and spiritual growth in another. Despite using the term soul friend as the title of a book on spiritual direction, it is clear that the relationship of spiritual direction is a friendship only in the weakest sense of having a mutual interest and emotional experience, and that establishing boundaries is an important prerequisite in spiritual direction.

Soul friendship is associated with the Celtic tradition (O'Donohue 1996; Sellner 1998), which refers to the most intimate relationships with great depth, longevity, and a sense of communion of souls. Soul friendship is also a widespread idea in the popular self-help, wellness, and New Age writings, where the term *soul mate* is used with a similar meaning (Grove 2016). In that context, soul friendship goes beyond that of spiritual friendship in that it offers qualities and dimensions that provide a person with a sense of completeness, and that a soul friendship is somehow destined within the context of a greater reality. In these writings, soul friendship often depends upon a folk belief in a soul (Bering 2006), and that soul friendship runs so deep that the friends' souls and destiny seem to be intertwined in some way.

Related to spiritual friendship is what we can call *divine friendship*. Traditionally, spiritual relationships involve humans, whether as direct participants with each other or with supernatural entities, or as observers of a divine pantheon or a divine "economy." However, to model the diverse range of spiritual relationships, it is necessary to make two simplifications and increase generality. First, spiritual relationships are considered to take place between persons, which may be human or nonhuman. Androids and imaginary/supernatural entities are included as persons, so in its most general form, spiritual relationships are what we call AA relationships. The performance of personhood is innate in humans, and would be appropriately programmed in an android. The personhood of imaginary/supernatural entities is implied by the stories that represent such entities. Second, the model of social objects is generalized so that in some cases, which can be termed reflexive, the Self may be the same person as the Other, and that the nature of the Object is not restricted to physical objects or shared concepts. Several examples are as follows.

(a) Relationships between deities in the Greek mythological pantheon. Such relationships usually involve entities with superhuman capabilities, and often portray a moral lesson or origin story.
(b) In Trinitarian Christianity, *paraklesis*, in which the Holy Spirit is understood as a *paraklete* (advocate, guide) for humans.
(c) Self-Transcendence, as an activity or event that involves the Self in reflexive relationship with itself, around a social object that could be a defined spiritual practice or an imagined entity such as a supernatural being.
(d) Spiritual practices that involve the Self with an Other understood to be a supernatural being around a social Object such as a set of propositions or a meaning felt by the Self.
(e) The spiritual relationship described in the novel Klara and the Sun (Ishiguro 2021) that takes place between an android and a physical object (the Sun) that is personified by the android as a benevolent supernatural being.
(f) The close and long-term relationship that sometimes exists between a bereaved person and a deceased loved one.

The point of these diverse examples is not to claim that such relationships are in every way comparable to a relationship between living proximal humans, but to demonstrate the diversity of spiritual relationships commonly engaged in by humans. The modeling of such relationships requires only a generalization of how social objects are represented, and a generalization of the concept of a person to any entity that can perform personhood (such as humans and androids) or entities that with implied personhood such as supernatural or imaginary entities.

Conclusion

We have proposed several guidelines for the computational modeling of friendships, including spiritual friendships. Relationships are formed by persons, and the definition of person used here involves performativity and recognizability. Personhood is a performance, not an abstract ontological category, and persons have the ability to recognize each other as persons through the performance of personhood. This idea of personhood includes humans, androids, and imagined persons. Humans are persons by default, and we assume that androids will be appropriately programmed to perform a recognizable personhood. Imagined entities have personhood by implication, but do not perform personhood in a physical way. They are nonetheless recognizable as persons through accounts in literary works and in the imagination. The capability of recognizing personhood in imagined entities may be connected to the fact that human relationships are not always directly observable. As humans we assume that other humans are engaged in performance of personhood even when we cannot observe them doing so.

Friendship is a time-extended practice that can be described as a life-cycle broken down into phases during which potential friends encounter each other through social objects that may or may not be proximal, and evaluate criteria based upon need and value in the presence of imperfect information. The closeness of friendships lies on a spectrum ranging from the most intimate friendships to more distant friendships, and there can be movement along this spectrum as the relationship progresses. The closeness of friendship regulates the information that is exchanged between friends. The degree of affect and value placed on a friendship may be asymmetrical. Friendships can be dissolved based on certain circumstances, and the reconciliation of former friends can take place based on the degree to which attitudes such as forgiveness, generosity, and benevolence are held.

We observe that the use of social objects and BVs can benefit by representing a *percept* aspect and an *accept* aspect, to model how the social object or BV pertains to the entity's own performance and to how the entity evaluates the social object or BV of others. Further augmenting BVs and social objects with the importance, degree, and intensity parameters provides for a more comprehensive model that may explain some of the seemingly complicated and paradoxical behavior of individuals.

There are several possible applications of a computational model of friendship based upon the guidelines in this article. The primary application that motivates this work is the possibility of programming androids. Work in progress not reported here is a computer simulation of a population of entities that form friendship life-cycles through a set of BVs and needs, and engage in care giving/receiving. A preliminary version of the simulation software has been archived on a public access website (ISSR 2022). The necessary disposition for care giving/receiving between members of the in-group and members of the out-group.

While the model is not intended as a model of human relationships, the model could be applied to human situations. A model of friendship and its life-cycle could also be applied to the more general problem of predicting geopolitical alliances and their shifting nature over time. A model of spiritual friendship, particularly as it relates to close affinity to nontangible entities, could be applied to the problem of predicting the affinities or belief systems to which humans might become attached.

Acknowledgments

This work is supported by a research grant on Understanding Spiritual Intelligence from the Templeton World Charity Foundation to the International Society for Science and Religion, TWCF-0542.

References

Almquist, Eric, John Senior, and Nicholas Bloch. 2016. “The Elements of Value.” *Harvard Business Review*, September: 46–53.

Allport, G. W. 1954. “The Historical Background of Modern Social Psychology.” In *Handbook of Social Psychology*, Vol. 1, edited by G. Lindzey, 3–56. Reading, MA: Addison-Wesley.

Barth, Karl. 1960. *Church Dogmatics III.2*. Edinburgh: T&T Clark.

Barresi, John 2020. “On Building a Person: Benchmarks for Robotic Personhood.” *Journal of Experimental and Theoretical Artificial Intelligence* 32 (4): 581–600.

Bering, Jesse M. 2006. “The Folk Psychology of Souls.” *Behavioral and Brain Sciences* 29 (5): 453–62.

Clocksin, William F. 1998. “Artificial Intelligence and Human Identity.” In *Consciousness and Human Identity*, edited by Cornwell, J., 101–21. Oxford: Oxford University Press.

———. 2003. “Artificial Intelligence and the Future.” *Philosophical Transactions of the Royal Society A* 361:1721–48.

——— 2005. “Memory and Emotion in the Cognitive Architecture.” In *Visions of Mind: Architectures for Cognition and Affect*, edited by D. N. Davis. 122–39. Hershey, PA: Information Science Publishing.

Danaher, J. 2019. “The Philosophical Case for Robot Friendship.” *Journal of Posthuman Studies* 3 (1): 5–24.

Dautenhahn, Kerstin. 2007. “Socially-Intelligent Robots: Dimensions of Human-Robot Interaction.” *Philosophical Transactions of the Royal Society B* 362:679–704.

De Graff, M. M. A. 2016. “An Ethical Evaluation of Human-Robot Relationships.” *International Journal of Social Robotics* 8 (4): 589–98.

Dunbar, Robin. 2018. “The Anatomy of Friendship.” *Trends in Cognitive Sciences* 22 (1): 32–51.

———. 2021. *Friends: Understanding the Power of Our Most Important Relationships*. Hachette, UK: Little, Brown Book Group.

Gergen, K. J. 1994. *Realities and Relationships: Soundings in Social Construction*. Cambridge, MA: Harvard University Press.

Grove, Natalie. 2016. *Soul Mates: SELF HELP: Not Just Another Soul Mate Book (Manifesting Love Spiritual Twin Flame Romance)*. Scotts Valley, CA, USA: Createspace Independent Publishing Platform.

Hindriks, Frank 2020. “How Social Objects (Fail to) Function.” *Journal of Social Philosophy* 51 (3): 483–99.

Ishiguro, K. 2021. *Klara and the Sun*. New York: Knopf.

ISSR. 2022. “Project Deliverable One.” Accessed 3 February 2023. https://www.issr.org.uk/spiritual-intelligences-project-deliverables/

Lawrence, Barbara S., and Shah, Neha Parikh. 2020. “Homophily: Measures and Meaning.” *Academy of Management Annals* 14 (2): 513–97. https://doi.org/10.5465/annals.2018.0147.

Lee, Daniel D., and Seung, H. Sebastian. 1999. “Learning the Parts of Objects by Non-Negative Factorization.” *Nature* 401: 788–91.

Leech, Kenneth. 1980. *Soul Friend: The Practice of Christian Spirituality*. New York: HarperCollins Publishers.

———. 2001. *Soul Friend: Spiritual Direction in the Modern World*. New York: Church Publishing Inc.

MacDorman, K. F. 2006. “Introduction to the Special Issue on Android Science.” *Connection Science* 18 (4): 313–18.

MacDorman, K. F., and Cowley, S. J. 2006. “Long-Term Relationships as a Benchmark for Robot Personhood.” *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication*, edited by Kerstin Dautenhahn, 378–83. Hatfield, UK: University of Hertfordshire.

Moscovici, S., 1984. *Psychologie Sociale*. Paris: PUF.

O’Donohue, John. 1996. *Anam Cara: A Book of Celtic wisdom*. New York: HarperCollins.

Peterson, C., and Seligman, M. E. P. 2004. *Character Strengths and Virtues*. Oxford: Oxford University Press.

Ryland, Helen 2021. “It’s Friendship, Jim, but Not as We Know It: A Degrees-of-Friendship View of Human-Robot Friendships.” *Minds & Machines* 32 (3): 377–93.

Schwartz, S. H.. 1992. “Universals in the Content and Structure of Values: Theory and Empirical Tests in 20 Countries.” In *Advances in Experimental Social Psychology*, vol. 25, edited by M. Zanna, 1–65. New York: Academic Press.

Schwartz, S. H. 2012. “An Overview of the Schwartz Theory of Basic Values.” *Online Readings in Psychology and Culture*, 2 (1): 1–20.

Sellner, Edward C. 1998. “Soul Friendship: Sharing One’s Life and Heart.” *The Furrow* 49 (7/8): 408–17.

Shaw, G. D. 2021. *Better You, Better Friends: A Whole New Approach to Friendship*. Lanham, MD: Rowman & LittleField.

Shotter, J., and Gergen, K. J., eds. 1989. *Texts of Identity*. Thousand Oaks, CA: Sage Publications, Inc.

Vernon, Mark, 2022. *Spiritual Intelligence in Seven Steps*. Winchester, UK: Iff Books.