DYNAMIC VALUE TRADE-OFFS IN RUN-TIME TO PROVIDE GOOD, CUSTOMISED PATIENT CARE WITH ROBOTS

By

Adam Poulsen

A dissertation submitted for the degree of

Bachelor of Computer Science (Honours)

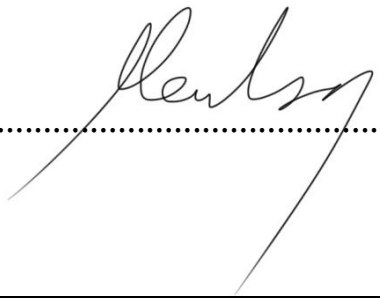
of Charles Sturt University

October 2017

DECLARATION

I declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma at Charles Sturt University or any other educational institution, except where due acknowledgment is made in the dissertation. Any contribution made to the research by colleagues with whom I have worked at Charles Sturt University or elsewhere during my candidature is fully acknowledged.

I agree that the dissertation be accessible for the purpose of study and research in accordance with the normal conditions established by the University Librarian for the care, loan and reproduction of the thesis.



.....

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation for my supervisor Dr. Oliver Burmeister for his constant support, patience, and motivation. Oliver has always encouraged, and contributed to, my own research direction. He is a man of immense knowledge and work ethic, like no one I've ever seen. I am honoured and lucky to be his student.

Besides my supervisor, Dr. Aimee van Wynsberghe and Wendell Wallach have provided significant council regarding the project and shown great patience and friendliness towards myself, an initial stranger and ever-remaining junior.

In addition to the heroes of my field whom I have already meet, thank you to all of those authors, professors, researchers, and others who contribute great works.

Thank you to Dr. David Tien who is the first academic to show a real interest in myself and my hopeful contributions to my field. David is the most inspirational, effective, hardworking, and kind teacher I have the pleasure of knowing.

Lastly, I would like to thank my family: my parents, Trish and Bernard, and brother, Sid, for their unwavering support, as well as the love of my life. Ressha your bravery and patience inspires me every day. Thank you for your encouragement, acclaim, and confidence in me - Mahal kita higit pa sa alam mo.

ABSTRACT

The present study was designed to investigate the ability for robots to provide good care. Good care is determinative in practice and customised to each patient. That is, moral acts in care are consciously determined by a carer during the act of caring for someone. The antithesis is that good care is determinative in theory. During the act of care, codes of conduct, as well as healthcare laws and regulations set ethical principles for carers to uphold. Care practices and processes inform what actions carers can take towards patients to apply those principles so as to ensure appropriate care, but what is lacking is customised patient care. Between normative ethical theory principles, practices and processes there is a gap. Which practice is the most suitable, what kind of personal approach does a carer take in the present context, and how much care does a particular patient need? These are a few of the decisions which cannot be prescriptively predetermined by principles, but need to be determined during practice. Principles do not determine the act which is best for patients in situ of their changing contexts. Instead, conscious carers make this determination in practice. Unfortunately, in geriatric care, human carer numbers are dwindling and the population of elderly in care is rising. Not only do the number of carers need to be increased, but they need to be filled by conscious carers so as to ensure good care; one way to do this is with conscious carebots. Through its usability, user acceptability, and value sensitivity testing research phases, the present study found that conscious carebots can fill growing care shortage, whilst simultaneously providing good care. To make the determination of good care a new carebot model, the attento model, was inspired by, and designed according to CCVSD. Employing CCVSD to inform design and computational consciousness as a method for a carebot enabled the determination by

uniquely providing extrinsic care value ordering, which in practice would be dynamic value trade-offs in run-time. The CCVSD-inspired attento model was presented in two research phases to test the hypotheses. The results of the research suggest that the attento model provides good, customised patient care in run-time. The present study contributes to literature on carebots, computational consciousness, and VSD through end-user perspectives currently lacking in the CCVSD literature.

CONTENTS

LITERATURE REVIEW

Adam Poulsen
School of Computing and Mathematics, Charles Sturt University,
Bathurst, NSW 2795, Australia
apouls02@postoffice.csu.edu.au

The following literature review uncovers the lack of human-level care in current carebots. It explores what good care is, and how to make it computational, implementable, and ensured by design. The review is structured in a way that each topic adds another element to the formation of a carebot model capable of human-level care. First, carebot literature is reviewed; the model starts with a basic carebot. Second, the model extends to a theoretical carebot capable of determining, in practice, what good care is upon care ethics review. Third, an assessment of values informs the model to include extrinsic care value ordering, a computational method of making that determination. Fourth, the model further extends into a CCVSD-guided one, featuring intrinsic care values implemented by design, upon reviewing value sensitive design. Finally, a review of computational consciousness inspires a way for the CCVSD conscious carebot to make the computational extrinsic value ordering method implementable.

1. INTRODUCTION

A social constructivist, interpretivist methodology guided this study. The literature, research, and conclusions within were approached with the methodology in mind. This study exists to address the decreasing number of carers for the elderly problem. It looks at carebots as a social phenomenon as well as the solution to the problem.

The research question posed the solution: is it possible to create a CCVSD conscious robot using the LIDA model that is usable, user accepted, and value sensitive; thus proving that consciousness, in carebots, can provide customised patient care with dynamic value ordering? In answering the question, end-user perspective contributions are being made to literature on carebots, computational consciousness, and VSD; ultimately providing a solution to the lack of elderly carers.

1.1 Significance of this study

This study commences with a review of carebots the understand the social phenomenon and before-mentioned solution. Current and theorised carebots were found to be lacking explicit conscious human-level care. Care ethics literature defined what 'good care' is. Care was found to be 'good' if a strong interpretive carer-patient relationship exists and if care is customised to individual patients. That is, the kind of relationship in which a patient trusts a carer enough to let them freely determine, in practice, how to customise care to them. Trust is a social construct formed by social factors such as society members having a good opinion of a carer. How can a carebot determine, in practice, what constitutes as good care for individual patients computationally? The same way human

carers do, by consciously identifying extrinsic values as found in an interpretive carer-patient relationship, and the subsequent ordering and affirmation of those values according to individual patient preference. These conscious processes are dynamic value trade-offs in run-time for a carebot. To identify initial extrinsic values to be provided to a carebot for ordering the stakeholders and context must be considered at the design stage. Additionally, intrinsic values (such as safety) must be ensured by design. Care-Centered Value Sensitive Design (CCVSD) is proficient in meeting those value requirements. It is a design methodology which places importance on care practices, thus it can be used for design to ensure care is determinative in practice. Upon review, CCVSD was found to have no end-user perspective. To implement human-like dynamic value trade-offs in a carebot it needs to be conscious thus computational consciousness literature was reviewed. Consciousness allows a carebot to identify extrinsic values, further beyond the initial ones provided in the design phase, by them having a subjective and empathetic experience with which it can intrinsically understand values with. The Learning Intelligent Distribution Agent (LIDA) model was found to have the ethical decision-making capabilities required to perform dynamic value trade-offs with Global Workspace Theory (GWT).

The following section is the literature review. Next is methodology, including two research phases. The paper, which has been submitted to a top journal, shows the findings and discusses the results. Then further discussion where the contributions of this study are made clear. Finally, the study is concluded.

2. LITERATURE REVIEW

The rising elderly population, and the inadequate number of carers to accommodate such a rise, calls for the intervention of carebots (Burmeister, 2016; Draper & Sorell, 2017; Garner, Powell, & Carr, 2016; Landau, 2013; Sharkey & Sharkey, 2011, 2012; Sparrow & Sparrow, 2006; Tokunaga, Tamamizu, Saiki, Nakamura, & Yasuda, 2017; Vallor, 2011). Carebots, as described by van Wynsberghe, are those that are "used in the care of persons in general" (2013a). Vallor states that "carebots are robots designed for use in home, hospital, or other settings to assist in, support, or provide care for sick, disabled, young, elderly, or otherwise vulnerable persons" (2011). Sharkey and Sharkey (2011) refer to carebots as developments in robot applications that assist the elderly and their carers, monitor health and safety, and provide companionship.

2.1 Ethical Concern

Ethical concern for the use of elderly carebots is expressed in literature (Sharkey & Sharkey, 2011; Sparrow & Sparrow, 2006; Vallor, 2011). Sharkey and Sharkey found that "using robots for elder care could result in increased social isolation, and could involve deception and loss of dignity" (2011). Sparrow and Sparrow (2006) conclude that the social and ethical implications of carebots are concerning. Arguing that carebots are incapable of meeting the social and emotional needs of the elderly and that the introduction of carebots, in such a healthcare sector that is already under economic pressure, would result in a "decrease in the amount of human contact experienced by older persons being cared for, which itself would be detrimental to their well-being" (Sparrow & Sparrow, 2006). Vallor (2011) infers risks relating to: the objectification of the elderly, carebots

restricting elderly values, the quality of care carebots can realistically give, and carebot-human relationships being deceptive or infantilizing.

2.2 Ethical Agency - Machine Ethics

Carebots are artificial moral agents (AMAs). Moor (2006) defines an AMA as a machine or AI that has some form of ethical agency, implication, and possibly guidance. The AMA ethical agency categories are impact, implicit, explicit or full explicit (Moor, 2006). An impact ethical agent has intentional or unintentional ethical implications due to its existence and use (Moor, 2006). For example, robotic camel jockeys in Qatar replace child jockeys and therefore reduce exploitation of children (Moor, 2006). An implicit ethical agent is one that was designed with the intention for its functionality to be ethical (Moor, 2006). For example, an autonomous car that was implicitly designed with heat sensors and collision avoidance to prevent an event in which it runs over a person's foot and causes them harm. Impact and implicit ethical agents have no ethical decision-making capabilities and therefore no guidance. Explicit and full explicit ethical agents have guidance. An explicit ethical agent and a full explicit ethical agent are both designed with autonomous explicit ethical evaluation which enables each to identify, process, and act upon ethical information; the latter agent also has consciousness, intentionality, and free will to support human-level ethical decision-making (Moor, 2006).

Carebots should be full explicit AMAs, and therefore conscious, intentional, and exhibiting free will to provide human-level care with human-level ethical decision-making. No existing, or designed, carebot is a full explicit AMA.

2.3 Categorising Carebots and Existing Examples

Sharkey and Sharkey (2011) categorise carebots according to purpose. Assistive carebots are used to assist the elderly, and/or their carers in daily tasks, such as robuWalker, My Spoon (van Wynsberghe, 2013a), and Pyxis MedStation. Companionship carebots are used to help monitor patient behaviour and health, such as Care-O-Bot 3, Adaptable Ambient Living Assistant (ALIAS) (Sharkey & Sharkey, 2011), robuMate (Orha & Oniga, 2012), Responsive Interactive Advocate (RITA) (Garner et al., 2016), Virtual Care Giver (Tokunaga et al., 2017), and Robotic-Human Interface for the Needed Ones (Ochoa, Aguiar, & Erazo, 2016). Most assistive and companion carebots also monitor patient behaviour and health, for example Virtual Care Giver monitors and records personal information for personalised comfort (Tokunaga et al., 2017), robuWalker monitors heart rate and sends data, and robuMate visually records scenes and performs scene analysis in case of emergency alarm.

van Wynsberghe (2013a) establishes three categories defined by the carebots relationship with patients, as well as carers in the case of replacement carebots. Other than assistive carebots, there are enabling carebots used for enabling patients and/or carers to meet the needs of patients, such as Da Vinci Surgery, Hybrid Assistive Limb, and RP7 Remote Presence Robot (van Wynsberghe, 2013a). There are also replacement carebots used for completely and autonomously replacing carer functions, such as Robot for Interactive Body Assistance (van Wynsberghe, 2013a), TUG Automated Robotic Delivery (Summerfield, Seagull, Vaidya, & Xiao, 2011), and Rudy the Robot (Keel, 2002).

No conscious carebots exist or have been theorised. Some carebots aim to create a friendly human-carebot relationship which may provide comfort. However, many feign

consciousness and are ultimately deceptive. Patients need a carer with a subjective and empathetic experience; someone who actually knows, and cares about, how they feel. Good care comes from care ethics.

3. CARE ETHICS

Part of the research question considers good care. Care ethics is addressed to discover what good care is. Care ethics concerns the moral guidance of behaviour when taking care of someone. They are those ethics that consider care to be of significant value (Vanlaere & Gastmans, 2011). Vallor (2011), Upton (2011), Tronto (2010), Gámez (2009), van Wynsberghe (2013a), and Vanlaere and Gastmans (2011) argue that care ethics and good care don't come from a normative ethical theory but rather they are "determinative in practice" (Beauchamp, 2004, p. 216). They come from our natural morality regarding concern for others. In care ethics it is considered ethical if a moral decision arises from the 'good' which is internal to practice rather than normative criteria or principles (Vallor, 2011). Good caring practices and relationships with carers are fundamental to care ethics. So what is an ethical action is the immediate good for the patient, which is provided by care practices. Upton defines determinative as "the provision, for any given case, of a single, well-grounded and widely convincing recommendation so as to the act that morally ought to be performed" (2011, p. 432). That is to say that moral acts in care are consciously determined by a carer during the act of caring for someone. The antithesis is that care is determinative in theory. A theory which is determinative in practice resolves the issue of an ethical theory being useless for guiding contextualised moral decisions in practice. Such a theory bridges the gap between ethical principles and action, or healthcare laws and care practice action.

Upton (2011) disagrees with a normative ethical theory for care, determining that they lack practical usefulness and a standard ethical decision-making method, as well as having complications due to diverse positions (such as act vs. rule utilitarianism). Tronto agrees, saying that "we would not want to be cared for according to some set model of standardization… [rather] we want care to rest upon a thick model of our own sensibilities" (2010). Gámez (2009) notes that good care is determined by the carer; that a carer decides on how to care for someone based on how they would like to be cared for themselves. Kittay (as cited in Edwards, 2011) disregards deontology and utilitarianism, the former only considers the right motive generally and the latter only considers consequences generally.

Determinative in practice care allows an agent to interpret care in-situ rather than by theorising. Kittay states that "merely intending to care is not sufficient to show that one's act is a caring act... care needs both to be undertaken from the right motive and to result in care" (as cited in Edwards, 2011). Vanlaere and Gastmans take the position that the 'good' evident in a carer-patient relationship can be promoted by the personalist anthropology which advocates: the recognition that patients are actual human beings with personal material needs, respecting dignity, and respect for the actual body of patients (2011). Tronto proposes the recognition of four fundamental care values that manifest moral elements in care practices to ensure good care, those elements are attentiveness, responsibility, competency, and responsiveness (2010). A study by Caris-Verhallen, Kerkstra, van der Heijden, and Bensing (1998) found that affective communication in the carer-patient relationship develops trust and increases happiness and quality of life. The carer-patient interpretive model by Emanuel and Emanuel (1992) recommends that carers

attempt to elucidate a patient's values correctly and then apply them to care practices (Emanuel & Emanuel, 1992).

Good care ethics are determined by a carer's interpretation of a patient's needs and values when being cared for as seen in the carer-patient relationship, and they require carers to have good intentions and take action based on these needs and values (Caris-Verhallen et al., 1998; Emanuel & Emanuel, 1992; Gámez, 2009; Tronto, 2010; van Wynsberghe, 2013a).

Carebots must be able to make the determination, in practice, of good care to be good carers. A method is needed as to how to make that determination computational. One such method is the identification of extrinsic values as found in the interpretive carer-patient relationship, and the subsequent ordering and affirmation of those values according to individual patient preference.

4. VALUES

Values are important to recipients of healthcare. In human interactions, they are our expectation of other's behaviour towards us and others. van Wynsberghe states that "a value is something desirable, something we want to have or to have happen" (2013a, p. 413). Furthermore, a value is something that we prefer to have.

Care values are those that we expect to be satisfied from social interactions in care, manifesting as patient, and carer, care values. van Wynsberghe notes that they play a significant role in care, "in making care what it is" (2013a, p. 415). Furthermore, stating that it is "through the manifestation of these values that one comes to understand what care really is in practice" (van Wynsberghe, 2013a, p. 415).

Elderly care values are those that are preferred or needed by elderly patients when being cared for so as to ensure good care. A review of literature reveals many elderly care values, what follows is a small sample.

- Dignity (Burmeister, 2016; Vanlaere & Gastmans, 2011)
- Wellbeing (Burmeister, 2016; Caris-Verhallen et al., 1998; Garner et al., 2016; Sharkey & Sharkey, 2011; Vanlaere & Gastmans, 2011)
- Safety (Burmeister, 2016; Draper & Sorell, 2014, 2017; Garner et al., 2016; Vanlaere & Gastmans, 2011)
- Independence (Burmeister, 2016; Caris-Verhallen et al., 1998; Draper & Sorell, 2014, 2017; Garner et al., 2016; Tronto, 2010)
- Respect (Burmeister, 2016; Garner et al., 2016; Tronto, 2010)
- Trust (Burmeister, 2016; Caris-Verhallen et al., 1998; Garner et al., 2016; Sharkey & Sharkey, 2011; Tronto, 2010)
- Autonomy (Burmeister, 2016; Draper & Sorell, 2014, 2017; Garner et al., 2016; Sharkey & Sharkey, 2011)
- Enablement (Draper & Sorell, 2014, 2017)
- Privacy (Burmeister, 2016; Draper & Sorell, 2014, 2017; Garner et al., 2016; Sharkey & Sharkey, 2011)
- Social connectedness (Caris-Verhallen et al., 1998; Draper & Sorell, 2014, 2017; Sharkey & Sharkey, 2011)
- Quality of life (Burmeister, 2016; Garner et al., 2016; Sharkey & Sharkey, 2011)
- Human rights (Burmeister, 2016; Sharkey & Sharkey, 2011)

- Comfort (Sharkey & Sharkey, 2011; Tronto, 2010)
- Freedom (Burmeister, 2016; Sharkey & Sharkey, 2011)
- Consent (Burmeister, 2016; Sharkey & Sharkey, 2011)
- Emergency help (Burmeister, 2016; Garner et al., 2016)

Of the elderly care values found in literature two categories of patient care values can be formed: intrinsic and extrinsic. Intrinsic values are those that, when provided, define good care. For example, safety is intrinsic to good care because safety is assumed when being taken care of. The following values are intrinsic: safety, emergency help, freedom, human rights, quality of life, trust, wellbeing, and comfort.

Extrinsic values are not the end goal in care, but rather they inform moral actions to reach intrinsic value end goals. For example, autonomy can be extrinsic to quality of life, a person may have a good quality of life if they are autonomous. The following values are extrinsic: consent, dignity, respect, autonomy, independence, social connectedness, and privacy.

A study by Draper and Sorell found that people in care favour compromise, persuasion and negotiation when making a value trade-offs during carebot design, they also prioritise autonomy above all other value except safety (2017). However, counter to the designer negotiation in this study and in value sensitive design, value trade-offs can be made by carebots and patients rather by designers during design which standardises care where it could be customised.

In addition to extrinsic value ordering, intrinsic values must also be provided to patients, this can be done by design of the carebot in cases where customisation is unsafe for the

patient. To ensure both types of values are identified and implemented, value sensitive design can be used.

5. VALUE SENSITIVE DESIGN

Value Sensitive Design (VSD) is a highly regarded technical design methodology (Boonstra & Van Offenbeek, 2010; Chesney, Coyne, Logan, & Madden, 2009; Dechesne, Warnier, & van den Hoven, 2013; Flanagan, Howe, & Nissenbaum, 2005; Friedman, 1996; Friedman & Grudin, 1998; Friedman, Kahn, & Borning, 2006; Friedman, Nathan, & Yoo, 2016; Friedman & Nissenbaum, 1996; Gotterbarn & Rogerson, 2005; Hedström, 2007; Manders-Huits, 2011; Nissenbaum, 2004; van Wynsberghe, 2013a). It "bridge[s] the gap between technical design considerations and ethical concerns expressed through human values" (Cummings, 2006). It provides guidance on how and when to include ethics into design during human-computer interaction (Cummings, 2006). VSD allows designers to "identify stakeholders and values in technology (design), with the ultimate objective of incorporating values into technology by means of design decisions" (Manders-Huits, 2011).

The VSD methodology is an iterative tripartite consisting of three investigations: conceptual, empirical, and technical. With the iteration of each investigation, and its subsequent analysis, the others are mutually informing and informed (Manders-Huits, 2011).

The conceptual investigation is a philosophical one, with the goal to "identify and articulate… central values at stake in a particular design context, and… the stakeholders that are affected by this (technology) design" (Manders-Huits, 2011).

The empirical investigation builds upon the conceptual one by identifying values that are relevant to the stakeholders in their use of the technology. The importance of the empirical investigation is "to find out how stakeholders experience (new) technologies with regard to the values they consider important in relation to their social environment and reference groups" (Manders-Huits, 2011).

The technical investigation concerns the "design and performance of the technology… [and] how the technology can and will support, or compromise, the human and moral values identified in the other parts [or investigations]" (Manders-Huits, 2011). Each decision made will affect the usability of the technology, stakeholder access, and stakeholder values. The end goal is to arrive at a design or make amendments to an existing one.

A carebot specific VSD framework is Care-Centered Value Sensitive Design.

6. CARE-CENTERED VALUE SENSITIVE DESIGN

Care-Centered Value Sensitive Design (CCVSD) is specifically for the design considerations of carebots. It narrows the VSD approach to care values, instead of broad human values. It promotes care ethics, arguing centrally "that the care perspective provides an orientation from which to begin theorizing as opposed to a pre-packaged ethical theory" (van Wynsberghe, 2013a, p. 420). It is this opposition to a 'pre-packaged ethical theory' or normative ethical theory that makes it clear that the CCVSD approach takes the position that care is determinative in practice. It's approach provides a "framework of components of ethical importance… along with a "user manual" for prospective evaluations" (van Wynsberghe, 2013b, p. 435). Its tools guide ethical design

for new technologies and evaluate existing technologies; those tools are the care centered framework (CCF) and the CCVSD methodology.

The CCF "articulates the components that require attention for analysis from a care perspective" (van Wynsberghe, 2013a, p. 420) regarding a care practice. The components, of an care practice interaction, help designers to interpret, rank, and provide meaning to relevant values; "the interpretation of values as well as their ranking and meaning differed depending on: the type of care (i.e. social vs physical care), the task (ex. bathing vs. lifting vs. socializing), the care-giver and their style, as well as the care-receiver and their specific needs" (van Wynsberghe, 2013a, p. 416). The interpretation of values is of key importance. Additional components of the CCF are the type of robot and manifestation of Tronto's moral elements: attentiveness, responsibility, competence, and responsiveness (2010). The moral elements are manifested through describing a care practice in two competing contexts (human vs. nonhuman carers). In these descriptions one reveals how the values are observed, prioritized, and interpreted depending on the context (van Wynsberghe, 2013a).

To follow the CCVSD methodology, one begins by identifying the components of the CCF and describing related traditional care practices to reveal values. One subsequently "speculates on what capabilities a robot ought to have to ensure the promotion of said values" (van Wynsberghe, 2013a, p. 424). Designers then make decisions concerning values with competing contexts in mind.

6.1 The Common Problem With VSD That CCVSD Solves

Manders-Huits notes that a shortcoming of VSD is that it "lacks a complimentary or explicit ethical theory for dealing with value trade-offs" (2011). Mander-Huits further criticises VSD, stating that it relies on broad values and, reiterates, that it lacks a normative ethical theory grounding, thus design decisions having no backing, and are instead interpreted by the designer (as cited in van Wynsberghe, 2013a). To combat this, van Wynsberghe states "I claim the fundamental care values of any practice to be attentiveness, responsibility, competence and reciprocity…I attempt to understand how these values are interpreted philosophically by care ethicists, as well as how these values are interpreted in context through observational work" (2013a). It is this promotion that indicates that van Wynsberghe (2013a) sees good care as being determinative in practice. The notion that good care is determinative in practice places importance on the interpretive carer-patient relationship as a basis of resolving issues with value trade-offs; which is to say that value trade-offs are made by carers.

CCVSD ensures intrinsic values by design. To make the ordering of extrinsic values implementable and informed by patient preference, computational consciousness could be implemented logically into carebots, making them conscious carers. Consciousness can also provide supporting functions: a subjective experience; internal moral dialogue; intrinsic understanding of situations, patients (as perceived as a unique agent), and patient values and their expression of said values; and general conscious functionality such as goal setting, situational awareness, etc.

7. CONSCIOUSNESS

Consciousness on a basic level is self-awareness, experience, and reactionary decision-making (Alvarado, 2016; Graziano, 2013; Natsoulas, 2013; Zeman, 2001). In AI it is subjective experience; internal moral dialogue; intrinsic understanding of situations; external stimuli perception through visual, audio, and other sensors; situational awareness and evaluation; and action and reaction. The implementation of consciousness into AI is called computational consciousness.

Before falling into unrealisable computation of consciousness, note that there are present limitations. Reggia (2013) draws three conclusions regarding the current state of computational consciousness developments. First, computational modelling is an effective and accepted methodology for studying consciousness (Reggia, 2013). Second, existing "computational models have successfully captured a number of neurobiological, cognitive, and behavioral correlates of conscious information processing as machine simulations" (Reggia, 2013). Third, and final, no model or approach is yet to show "phenomenal machine consciousness, or even clear evidence that artificial phenomenal consciousness will eventually be possible" (Reggia, 2013).

However, if Dennett is to be believed, consciousness will eventually be fully computational and the “best reason for believing that robots might someday become conscious is that we human beings are conscious, and we are a sort of robot ourselves” (1994). Currently, the best way towards a conscious AI is through Global Workplace Theory (GWT), a highly regarded model of consciousness (Irvine, 2012; Levy, 2014; Schutter & van Honk, 2004; Shanahan & Baars, 2005).

7.1 Global Workspace Theory

“Global Workspace Theory is currently the most empirically supported and widely discussed theory of consciousness. It provides a high-level description of such algorithms, based on a large body of psychological and brain evidence” (Baars & Franklin, 2009). Baars and Franklin (2009) argue that consciousness, that is to say "the conscious as well as the non-conscious aspects of human thinking, planning, and perception are produced by adaptive, biological algorithms" (2009).

GWT logic features a conceptual workspace where codelets (functions that gather information) attempt to create coalitions (related information structures) to form a current situational model which reflects the actual real world situation. The most relevant, salient, important, and urgent of which is moved to the global workspace which represents an agent's current concious priority (Baars & Franklin, 2009, p. 5). That process begins by an external or internal stimuli triggering a proposer codelet to bring an initial thought into the workspace, this creates a competition among other proposers in the workspace for current conscious priority. Supporter codelets gather relevant information (local associations) for the forming thought from episodic and declarative memory which gives the thought further complexity. Objector codelets object to irrelevant information proposals that supporter codelets bring into the workspace. Timekeeper codelets force decision on which coalition should go into the global workspace (Baars & Franklin, 2009, pp. 7-8).

Computational consciousness is a comprehensively discussed topic throughout literature, the general consensus is that it's possible (Aleksander & Morton, 2006; Dennett, 1997; Gök & Sayan, 2012; Marcin, 2010; Matzke, 2010; Spicer, Gangopadhyay, & Madary, 2010).

Baars and Franklin employ GWT, the human-inspired model of consciousness by Baars (1988), which they partially implement with the learning intelligent distribution agent (LIDA) model of computational consciousness (Baars & Franklin, 2007, 2009). The human-like aspects of GWT include: the generation of "explicit predictions for conscious aspects of perception, emotion, motivation, learning, working memory, voluntary control, and self systems in the brain" (Baars & Franklin, 2009). The LIDA model is the most credited and recognised model of computational consciousness to date (Wallach, Franklin, & Allen, 2010, p. 454).

7.2 Learning Intelligent Distribution Agent

The LIDA model is a partial implementation of GWT that provides consciousness for AI. It provides an "explicit implementation of much of GWT, which can be shown to perform human-like tasks, such as the interactive assignment of naval jobs to sailors" (Baars & Franklin, 2009). It implements a number of psychological and neuropsychological theories including situated cognition, perceptual symbol systems, working memory, memory by affordances, long-term working memory, and transient episodic memory (Baars & Franklin, 2009). The LIDA model is designed to implement environmental sampling (sensing), so an agent can understand or process the situation (make sense of), and react (action). It is the 'making sense of' that facilitates decision-making in the LIDA model (Baars & Franklin, 2009). The decision-making capacity of the LIDA model allows for ethical decision-making.

This study makes claims about computational consciousness that are supported by W. Wallach (a consultant, ethicist, scholar, and author of papers on the LIDA model). W. Wallach stated that as "far as consciousness goes, your claims for CC [computational

consciousness] are valid" (personal communication, July 13, 2017). Those claims are that computational consciousness is capable of situation observation and evaluation; self-awareness and reactionary responsiveness; external stimuli perception; internal moral dialogue; intrinsic understanding of situations, patients, and patient values and their expression of values; and attentiveness. Most importantly regarding conscious carebots is internal moral dialogue, "the LIDA model helps integrate emotions into the human decision-making process, and we will elucidate a process... whereby an agent can work through an ethical problem to reach a solution that takes account of ethically relevant factors" (Wallach et al., 2010, p. 454). The LIDA model is theoretically capable of supporting conscious carebots with ethical decision-making.

7.3 Implementing Moral Decision-Making

Other computational models of ethical decision-making exist, such as Truth Teller and SIOCCO. However, the LIDA model provides consciousness and is the benchmark as it "is a highly regarded model of human cognition that is currently being computationally instantiated in several computational implementation" Wallach et al. (2010).

Wallach et al. (2010) provides a concise description of how to use the LIDA model for ethical decision-making; here is how it is done:

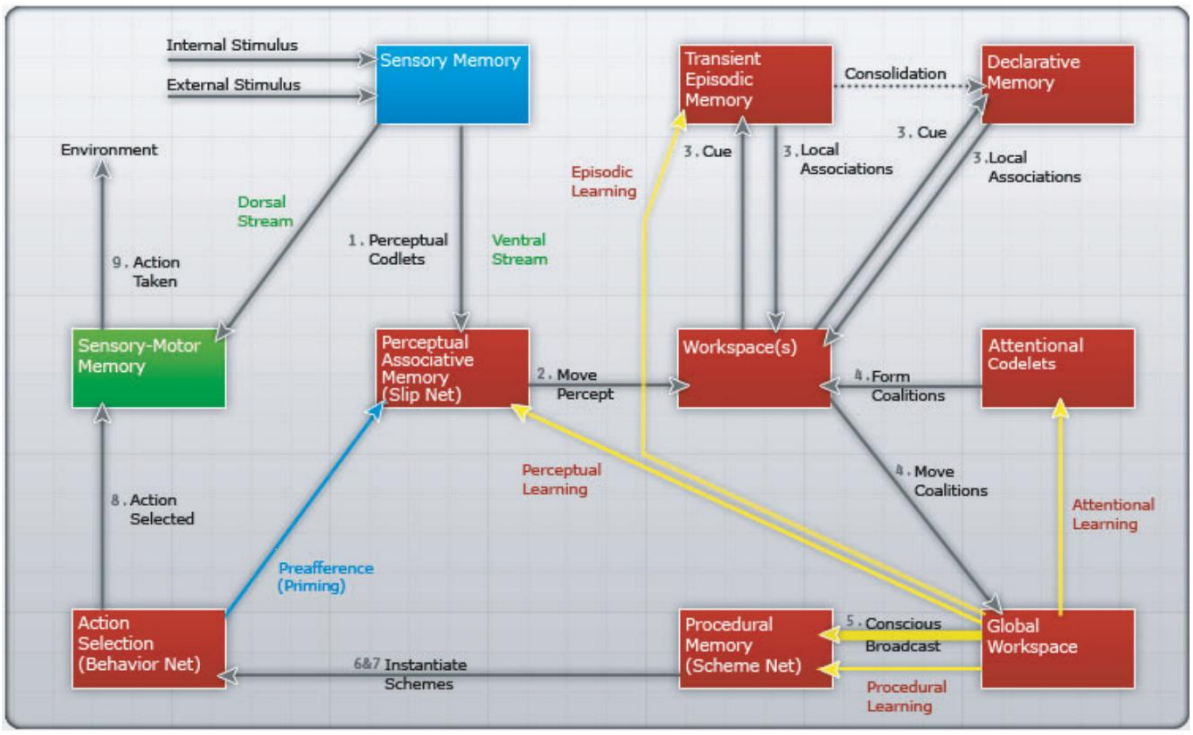


Figure 1. The Cognitive Cycle in the LIDA model. From " Modeling medical diagnosis using a comprehensive cognitive architecture," by S. Strain and S. Franklin. 2011, *Journal of Healthcare Engineering, 2*(2), p. 250.

1. An item of moral deliberation is recognised in the perceptual associative memory (which translates basic sensory information such as shapes, smells, etc., to objects, thoughts, feelings, etc.) and moved to the workspace by a proposer codelet (Wallach et al., 2010, p. 475).
2. The proposer codelet creates a conscious thought proposal (Wallach et al., 2010, p. 475). By doing this it is suggesting that the proposal (a moral decision for the item of moral deliberation) should take occupation of the agent's conscious.
3. Objector and supporter codelets search transient episodic and declarative memory for local associations to provide context related to the item of moral

deliberation. These codelets look to object to the proposal or support it (Wallach et al., 2010, p. 475). If codelets find a local association (moral rule) related to the item of moral deliberation, then it objects or supports the proposal based on this association (depending on the local association). An objector codelet would oppose the proposal, whilst a supporter codelet would support it (Wallach et al., 2010, p. 475). Objections to, and support for, the proposal (now a coalition) continues until there are no more objections to the proposal.

4. The resulting coalition (moral decision) is moved to the global workspace.
5. The moral decision occupies the agent's conscious by being globally broadcasted; creating conscious thought (Wallach et al., 2010, p. 475). This results in the simultaneous triggering of learning procedures, where new and reinforced moral rules in the moral decision move to each memory, and the moving of the moral decision to procedural memory to prepare possible action schemes (possible actions).
6. Possible actions are devised.
7. Possible actions are sent to action selection.
8. The agent chooses an action based on the resulting broadcasted moral decision.

Non-human agents, such as carebots, can be capable of ethical decision-making when given consciousness. Consciousness allows such agents to be aware, have a subjective experience, be reactionary decision makers, have internal moral dialogue, etc. during their caring for the elderly.

7.4 Awareness, Experience, and Decision-Making

When considering the definition of consciousness, especially those similar or related to AI consciousness, two overlapping perspectives from (1) microbiology/chemistry and (2) neurobiology are helpful.

(1) Vaneechoutte (2000) argues that awareness and consciousness are "complex higher order experiences as they become possible with ongoing evolution" (Vaneechoutte, 2000, p. 449). Vaneechoutte (2000) takes the position that consciousness is a form of awareness, recollection, and consideration of one's own previous actions (p. 445). Stating that "consciousness as 'reflexive awareness', i.e. some kind of experience which is possible because symbolic language enables us to take distance of the current aware experience, and to observe it as if we were a third person looking at ourselves" (Vaneechoutte, 2000, p. 447). Vaneechoutte (2000) argues that animals don't have human-level consciousness because they don't have a symbolic language. Baars and Franklin (2009) disagree, arguing that verifiable reportability, a testable criteria for consciousness similar to reflexive awareness, "may prove to be a useful necessary, but by no means sufficient, criteria for a machine's being conscious of a perceptual event" (p. 9). The testing criteria for consciousness seems to be a bit more complex than that from microbiology and chemistry.

A prominent international group of cognitive neuroscientists, neuropharmacologists, neurophysiologists, neuroanatomists, and computational neuroscientists who formulated 'The Cambridge Declaration on Consciousness' would agree with Vaneechoutte that consciousness is a form of awareness. However, since reflexive awareness requires linguistic capabilities, this disregards animals and pre-linguistic humans as being conscious according to Vaneechoutte; the Cambridge Declaration on Consciousness committee

would certainly disagree with this requirement. Instead the committee declares that convergent "evidence indicates that non-human animals have the neuroanatomical, neurochemical, and neurophysiological substrates of conscious states along with the capacity to exhibit intentional behaviors... non-human animals, including all mammals and birds, and many other creatures, including octopuses, also possess these neurological substrates" (Low, 2012b). Low concludes that it is obvious to every academic involved that animals have consciousness, but it is not obvious to the rest of the world (Low, 2012a, 0:27). The LIDA model is built on psychological and neuropsychological theories such as GWT. If an AI has the same neuro- substrates as an animal, they too qualify as being conscious in the same way. Furthermore, if AI consciousness has the same cognitive complexity as humans, they too qualify as being conscious in the same way.

(2) De Sousa (2013) presents a related form of consciousness called operational consciousness. Operational consciousness is consciousness as "motor, sensory, cognitive, creative, emotive, aesthetic, ethical and other such abilities... [as well as] awareness of all mental operations" (De Sousa, 2013, p. 104). For AI consciousness, these abilities can be seen as software modules, motor functions for controlling movement, sensory for environmental awareness input, cognition for decision-making, emotive for feeling and expression, processing, and action selection, and so on. These software modules could be placed on top of a central processing unit in order to implement them into AI.

Literature demonstrates that a carebot with computational consciousness could be aware, have a subjective experience, be reactionary decision makers, have internal moral dialogue, etc. Consciousness is only one requirement for being a full explicit AMA. Such

a carebot capable of human-level care also requires intentionality and free will. The LIDA model demonstrates those requirements.

7.5 Carebots as AMAs - Intentionality and Free Will

The LIDA model accounts for intentionality through the relationship between the workspace and declarative memory, like that between a thought and a thing. Declarative memory is conscious explicit long term memory (Wallach et al., 2010, p. 465). Thoughts proposed in the workspace attempt to build in complexity, or expand out the thought to a thing, by finding local associations in declarative memory. Intentionality in AI aids in making the determiniation, in practice, of what is good care because it informs the connection between healthcare laws and care practices which is the gap such a determination fills. It also aids in linguistic and somatic interpretations and outputs. With this, AI can have conversations, provide verbal and physical comfort, and observe and evalutate the situation. Regarding free will, simply if humans possess free will, then AMAs that are modelled on human consciousness will have free will. Free will in a carebot means it can make the determination of what is good care freely in practice. A full explicit AMA carebot is more capable of providing human-level care than any existing carebot implementation or design.

8. HOLES IN LITERATURE

There are several holes in current literature. First, there are no existing, or designed, conscious carebots, nor is there one that can make the determination of good care in practice. Subsequently, there are no carebots in literature that are full explicit AMAs, thus there are none that provide human-level care. Although some companion carebots, such as

RITA (Garner et al., 2016) and Virtual Care Giver (Tokunaga et al., 2017), deceptively feign consciousness to appear to have empathy. Second, there is no existing carebot model that suggests using extrinsic value ordering as a method to make that determination. Third, computational consciousness for carebots, specifically the LIDA model, has no end-user perspective. Fourth, CCVSD also has no end-user perspective.

VSD has flaws. First, it makes value trade-offs as if all values are intrinsic, this generalises care. Some of those values could be extrinsic and should instead be customised to individual end-users. When designing a care technology, developers decide for patients whether they value one thing more than another. Some of those value trade-offs may be extrinsic vs. extrinsic such as privacy vs. independence. This is a decision that should be and could be customised to and by the patient. CCVSD has the same inherent problem. To demonstrate, when contemplating patient privacy regarding the design of a camera for patient monitoring one might decide to ensure that carebots (1) delete visual files containing patient nudity, or (2) stop recording in such a scenario, or (3) not record ever, or (4) avoid them altogether. But in one's concern for patient privacy, one has not considered that a particular patient might not value privacy highly, or at all. Also, through any of those suggested implementations we have lost valuable care practice assessment data which a carebot could use to improve its care.

The second flaw of VSD is that for perfect design it relies on a universal ethical theory (UET) to govern a designer's decisions when one doesn't exist. Even if a UET existed it still leaves a gap: how does a carebot interpret the principles to perform good care practices?

To simultaneously fill these holes, amend the flaws in VSD, and answer the research question, a new carebot model will be theorised, designed, and tested: the attento model.

Human-level care requires human-level ethical decision-making. Carebots must be full-explicit AMAs to provide good human-level care. Care ethics literature revealed that good care is customised to patients and consciously determinative in practice. Extrinsic value ordering is a method to provide such care and make such determination. This determination addresses both flaws in VSD by enabling carebots to make dynamic value trade-offs in run-time instead of designers in the design phase. This method consists of the identification of extrinsic values as found in an interpretive carer-patient relationship, and the subsequent ordering and affirmation (in practice) of those values according to individual patient preference. Extrinsic and intrinsic values must be identified during the design phase. Intrinsic values must still be ensured by design, CCVSD is a proficient and targeted method to do that. To make this method a conscious determination, computational, and informed by patient preference, GWT computational consciousness should be used. GWT consciousness can also provide these supporting functions: a subjective experience; internal moral dialogue and intrinsic understanding of situations, patients (as perceived as a unique agent), and patient values and their expression of said values; and general conscious functionality such as goal setting, situational awareness, etc. To make these computational methods and functions implementable and make a carebot a full explicit AMA the LIDA should be used. The LIDA model practically implements GWT and demonstrates internal moral dialogue, ethical decision-making, intentionality, free will, and other functions required for making a carebot a full explicit AMA capable of performing the method and providing good, customised patient care with dynamic value

trade-offs in run-time. An attento is such a carebot; a CCVSD conscious carebot capable of dynamic value trade-offs in run-time.

METHODOLOGY

This study follows the interpretivist, social constructivist research philosophy and qualitative method. Detailed here are the philosophical foundations of the study, the research philosophy and method, the data gathering techniques and process, and the data analysis.

The interpretivist philosophy recognises that people are complex individuals and that each individual has a unique and subjective experience of reality. Goldkuhi identifies that the core idea of interpretivism is to work with "subjective meanings already there in the social world; that is to acknowledge their existence, to reconstruct them, to understand them, to avoid distorting them, to use them as building-blocks in theorizing (2012). Social constructivism states that the unique experience of reality that individuals have is influenced by social factors such as culture (Burr, 2015). This study concerns carebot-human interaction in a social context which is a type of social phenomenon. Within interpretivism, social constructivism was chosen to guide the research due to the purpose of the study and the nature of the data to be collected. The research was undertaken to test the usability, acceptability, and value sensitivity of the attento model in an elderly care context and therefore end-user interpretations were crucial. Those interpretations were captured and analysed according to thematic analysis (Morse, 2008).

Other methodologies were considered. The positivist approach would have been inappropriate because it emphasises the value of quantitative data. Since there are no end-user perspectives on conscious carebots, extrinsic value ordering, the LIDA model for carebots, and CCVSD, using a quantitative approach is not appropriate as the variables and benchmarks are unknown. An alternative to the social constructivist approach is the

phenomenological one. If that were the focus of the study, it would require research participants to have had an experience of elderly care. That requirement is impractical considering the limited time and sampling of this research.

1. DATA GATHERING TECHNIQUES

Data was gathered in two phases. The first was a heuristic, expert evaluation. The second an online survey. Both utilised a specific attento, an elderly care medicine delivery attento, as well as scenarios involving carebot-patient interactions. This study aims to address the lack of perspectives through the answering of its research question: is it possible to create a CCVSD conscious robot using the LIDA model that is usable, user accepted, and value sensitive; thus proving that consciousness, in carebots, can provide customised patient care with dynamic value ordering? To answer this question, participants were asked to provide opinions on the medicine delivery attento. The thematic analysis and discussion of those opinions provides end-user perspectives that this study aims to provide.

The Charles Sturt University, Faculty of Business, Justice and Behavioural Sciences Human Research Ethics Committee deemed the following research phases ethical (Appendix F).

1.1 Phase 1- Heuristic, Expert Evaluation

The heuristic evaluation technique was used so as to test the specifications of the attento to make sure it was technically possible and ethically acceptable. For Phase 1 a robotics

academic, computer ethicist, computer scientist, and registered nurse were selected to get expert opinions from AI, robotics, healthcare, and related ethics.

Participants were presented with the medicine delivery attento's design (Appendix B), which is a result of the CCVSD CCF being used (Appendix A) and methodology applied, and three scenarios (Appendix C). Those scenarios demonstrated where the attento encountered a situation where values should be considered, at which point it examined the relevant patient's value priority list, evaluated the situation and revealed relevant extrinsic values, and made a decision based on the situation variables and prioritised extrinsic values.

Each scenario consisted of the same initial interactions, but the patient's prioritised extrinsic values changed so as to demonstrate the dynamic value trade-offs. With the change, the attento's action may have changed too, depending on the patient's prioritised extrinsic values (particularly the highest priority one) and if that value was relevant to change the attento's decision and action. Each scenario had an intrinsic value as the goal of the encounter. Each of the changing highest priority extrinsic values were designed to ensure the intrinsic value was met in a patient customised way.

Through the scenarios, participants were asked to imagine an elderly person they knew and put themselves in their shoes so as to get a more relevant perspective. After each scenario change, participants were asked: "*What do you think they would make of their care robot if it made a decision like this for them?"* At the end of the scenarios participants were asked more questions (Appendix D) to gather further expert opinions.

1.2 Phase 2 - Online Survey

An online survey was employed to get public opinions on the attento just like any other new technology. Such a survey where a designer wants to know if the new technology is usable, acceptable, and value sensitive. The second phase recruited random global participants to get those public opinions.

Participants were presented with scenarios one and two (Appendix C) and questions were asked (Appendix E). After the initial scenario and after each change in the highest priority value, a question about trust in the attento's decision and another about value sensitivity of the attento were asked. At the end of each scenario participants were asked an open-ended question concerning their thoughts and feelings about the attento considering the dynamic value trade-offs it made. Scaled and multiple choice questions were asked, each followed by open-ended questions. The scaled question queried the acceptability and usability of the attento model by asking how likely they were to accept and use it in their everyday lives. Questions addressed the trustworthiness of the attento in the scenarios, and dynamic value trade-offs in practice.

The following section is a journal article which presents the findings and discusses the results. It concludes that the majority of research participants agree that care, which is determinative in practice, can be performed by attentos. Then further discussion where the contributions of this study are made clear. Finally, the study is concluded.

REFERENCES

Aleksander, I., & Morton, H. (2006). Phenomenology in computational models of consciousness. *International Journal of Pattern Recognition and Artificial Intelligence, 20*(06), 785-796. doi:10.1142/S021800140600496X

Alvarado, L. (2016). *Consciousness: Social perspectives, psychological approaches and current research* [Nova Science Publishers version]. Retrieved from https://www.ebsco.com/

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.

Baars, B. J., & Franklin, S. (2007). An architectural model of conscious and unconscious brain functions: Global workspace theory and IDA. *Neural Networks, 20*(9), 955-961. doi:10.1016/j.neunet.2007.09.013

Baars, B. J., & Franklin, S. (2009). Consciousness is computational: The LIDA model of global workspace theory. *International Journal of Machine Consciousness, 01*(01), 23-32. doi:10.1142/s1793843009000050

Beauchamp, T. L. (2004). Does Ethical Theory Have a Future in Bioethics? *The Journal of Law, Medicine & Ethics, 32*(2), 209-217. doi:10.1111/j.1748-720X.2004.tb00467.x

Boonstra, A., & Van Offenbeek, M. (2010). Towards consistent modes of e-health implementation: structurational analysis of a telecare programme's limited success. *Information Systems Journal, 20*(6), 537-561. doi:10.1111/j.1365-2575.2010.00358.x

Burmeister, O. K. (2016). The development of assistive dementia technology that accounts for the values of those affected by its use. *Ethics and Information Technology, 18*(3), 185-198. doi:10.1007/s10676-016-9404-2

Burr, V. (2015). *Social constructionism* [Taylor and Francis version]. Retrieved from https://ebookcentral.proquest.com/lib/fjpw/

Caris-Verhallen, W. M., Kerkstra, A., van der Heijden, P. G., & Bensing, J. M. (1998). Nurse-elderly patient communication in home care and institutional care: an explorative study.

International Journal of Nursing Studies, 35(1), 95-108. doi:10.1016/S0020-7489(97)00039-4

Chesney, T., Coyne, I., Logan, B., & Madden, N. (2009). Griefing in virtual worlds: causes, casualties and coping strategies. *Information Systems Journal, 19*(6), 525-548. doi:10.1111/j.1365-2575.2009.00330.x

Cummings, M. L. (2006). Integrating ethics in design through the value-sensitive design approach. *Science & Engineering Ethics, 12*(4), 701-715. doi:10.1007/s11948-006-0065-0

De Sousa, A. (2013). Towards an integrative theory of consciousness: Part 1 (neurobiological and cognitive models). *Mens Sana Monographs, 11*(1), 100-150. doi:10.4103/0973-1229.109335

Dechesne, F., Warnier, M., & van den Hoven, J. (2013). Ethical requirements for reconfigurable sensor technology: a challenge for value sensitive design. *Ethics and Information Technology, 15*(3), 173-181. doi:10.1007/s10676-013-9326-1

Dennett, D. C. (1997). *Consciousness in human and robot minds* [Oxford University Press version]. doi:10.1093/acprof:oso/9780198524144.001.0001

Draper, H., & Sorell, T. (2014). Robot carers, ethics, and older people. *Ethics and Information Technology, 16*(3), 183-195. doi:10.1007/s10676-014-9344-7

Draper, H., & Sorell, T. (2017). Ethical values and social care robots for older people: an international qualitative study. *Ethics and Information Technology, 19*(1), 49-68. doi:10.1007/s10676-016-9413-1

Edwards, S. D. (2011). Is there a distinctive care ethics? *Nursing Ethics, 18*(2), 184-191. doi:10.1177/0969733010389431

Emanuel, E. J., & Emanuel, L. L. (1992). Four models of the physician-patient relationship. *Journal of the American Medical Association, 267*(16), 2221-2226. doi:10.1001/jama.1992.03480160079038

Flanagan, M., Howe, D. C., & Nissenbaum, H. (2005, April 2-7). *Values at play: design tradeoffs in socially-oriented game design.* Paper presented at the Proceedings of the SIGCHI Conference on human factors in computing systems, Portland, Oregon, USA.

Friedman, B. (1996). Value-sensitive design. *Interactions, 3*(6), 17-23. doi:10.1145/242485.242493

Friedman, B., & Grudin, J. (1998, April 18-23). *Trust and accountability: preserving human values in interactional experience.* Paper presented at the Computer-Human Interaction '98 Conference summary on human factors in computing systems, Los Angeles, CA, United States.

Friedman, B., Kahn, P. H. J., & Borning, A. (2006). Value sensitive design and information systems. In P. Zhang & D. Galletta (Eds.), *Human-computer interaction and management information systems: Foundations* (pp. 348-372). New York: M. E. Sharpe.

Friedman, B., Nathan, L. P., & Yoo, D. (2016). Multi-lifespan information system design in support of transitional justice: Evolving situated design principles for the long(er) term. *Interacting with Computers, 29*(1), 80-96. doi:10.1093/iwc/iwv045

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems, 14*(3), 330-347. doi:10.1145/230538.230561

Gámez, G. G. (2009). The nurse-patient relationship as a caring relationship. *Nursing Science Quarterly, 22*(2), 126-127. doi:10.1177/0894318409332789

Garner, T. A., Powell, W. A., & Carr, V. (2016). Virtual carers for the elderly: A case study review of ethical responsibilities. *Digital Health, 2*, 1-14. doi:10.1177/2055207616681173

Gök, S. E., & Sayan, E. (2012). A philosophical assessment of computational models of consciousness. *Cognitive Systems Research, 17*, 49-62. doi:10.1016/j.cogsys.2011.11.001

Goldkuhl, G. (2012). Pragmatism vs interpretivism in qualitative information systems research. *European Journal of Information Systems, 21*(2), 135-146. doi:10.1057/ejis.2011.54

Gotterbarn, D., & Rogerson, S. (2005). Responsible risk analysis for software development: Creating the software development impact statement. *Communications of the Association for Information Systems, 15*(40), 730-750.

Graziano, M. S. A. (2013). *Consciousness and the social brain* [Oxford University Press version]. Retrieved from https://ebookcentral.proquest.com/lib/fjpw/

Hedström, K. (2007). The values of IT in elderly care. *Information Technology and People, 20*(1), 72-84. doi:10.1108/09593840710730563

Irvine, E. (2012). *Consciousness as a scientific concept: A philosophy of science perspective* [SpringerLink version]. doi:10.1007/978-94-007-5173-6

Keel, C. (2002). Rudy the robot. *The American Journal of Nursing, 102*(8), 24. doi:10.1097/00000446-200208000-00026

Landau, R. (2013). Ambient intelligence for the elderly: Hope to age respectfully? *Aging health, 9*(6), 593-600. doi:10.2217/ahe.13.65

Levy, N. (2014). *Consciousness and moral responsibility* [Oxford Scholarship Online version]. doi:10.1093/acprof:oso/9780198704638.001.0001

Low, P. (2012a). *Animal consciousness officially recognized by leading panel of neuroscientists* [Video file]. Retrieved from https://www.youtube.com/watch?v=RSbom5MsfNM

Low, P. (2012b). *The Cambridge declaration on consciousness* [Declaration]. Retrieved from http://fcmconference.org/

Manders-Huits, N. (2011). What values in design? The challenge of incorporating moral values into design. *Science and Engineering Ethics, 17*(2), 271-287. doi:10.1007/s11948-010-9198-2

Marcin, M. (2010). Obliczeniowe teorie świadomości (Computational theories of consciousness). *Analiza i Egzystencja*, *11*, 133-154.

Matzke, D. (2010). Consciousness: A Computational Paradigm Update. *Activitas Nervosa Superior, 52*(3), 134-140. doi:10.1007/BF03379577

Moor, J. H. (2006). The nature, importance, and difficulty of machine ethics. *IEEE Intelligent Systems, 21*(4), 18-21. doi:10.1109/MIS.2006.80

Morse, J. M. (2008). Confusing categories and themes. *Qualitative Health Research, 18*(6), 727-728. doi:10.1177/1049732308314930

Natsoulas, T. (2013). *Consciousness and perceptual experience: An ecological and phenomenological approach* [Cambridge University Press version]. Retrieved from https://ebookcentral.proquest.com/lib/fjpw/

Nissenbaum, H. (2004). Hackers and the contested ontology of cyberspace. *New Media Society, 6*(2), 195-217. doi:10.1177/1461444804041445

Ochoa, M., Aguiar, G., & Erazo, A. (2016). RHINO–an autonomous interactive surveillance robot for the needed ones: design and study case. *MATEC Web of Conferences, 56*, 1-5. doi:10.1051/matecconf/20165607003

Orha, I., & Oniga, S. (2012). Assistance and telepresence robots: a solution for elderly people. *Carpathian Journal of Electronic & Computer Engineering, 5*(1), 87-90.

Reggia, J. A. (2013). The rise of machine consciousness: Studying consciousness with computational models. *Neural Networks, 44*, 112-131. doi:10.1016/j.neunet.2013.03.011

Schutter, D. J. L. G., & van Honk, J. (2004). Extending the global workspace theory to emotion: Phenomenality without access. *Consciousness and Cognition, 13*(3), 539-549. doi:10.1016/j.concog.2004.05.002

Shanahan, M., & Baars, B. (2005). Applying global workspace theory to the frame problem. *Cognition, 98*(2), 157-176. doi:10.1016/j.cognition.2004.11.007

Sharkey, A., & Sharkey, N. (2011). The eldercare factory. *Gerontology, 58*(3), 282-288. doi:10.1159/000329483

Sharkey, A., & Sharkey, N. (2012). Granny and the robots: ethical issues in robot care for the elderly. *Ethics and Information Technology, 14*(1), 27-40. doi:10.1007/s10676-010-9234-6

Sparrow, R., & Sparrow, L. (2006). In the hands of machines? The future of aged care. *Minds and Machines, 16*(2), 141-161. doi:10.1007/s11023-006-9030-6

Spicer, F., Gangopadhyay, N., & Madary, M. (2010). *Perception, action, and consciousness: Sensorimotor dynamics and two visual systems* [Oxford Scholarship Online version]. doi:10.1093/acprof:oso/9780199551118.001.0001

Strain, S., & Franklin, S. (2011). Modeling medical diagnosis using a comprehensive cognitive architecture. *Journal of Healthcare Engineering, 2*(2). doi:10.1260/2040-2295.2.2.241

Summerfield, M. R., Seagull, F. J., Vaidya, N., & Xiao, Y. (2011). Use of pharmacy delivery robots in intensive care units. *American Journal of Health-System Pharmacy, 68*(1), 77-83. doi:10.2146/ajhp100012

Tokunaga, S., Tamamizu, K., Saiki, S., Nakamura, M., & Yasuda, K. (2017). VirtualCareGiver: Personalized smart elderly care. *International Journal of Software Innovation, 5*(1), 30-43. doi:10.4018/IJSI.2017010103

Tronto, J. C. (2010). Creating caring institutions: Politics, plurality, and purpose. *Ethics & Social Welfare, 4*(2), 158-171. doi:10.1080/17496535.2010.484259

Upton, H. (2011). Moral theory and theorizing in health care ethics. *Ethical Theory and Moral Practice, 14*(4), 431. doi:10.1007/s10677-011-9295-6

Vallor, S. (2011). Carebots and caregivers: Sustaining the ethical ideal of care in the twenty-first century. *Philosophy & Technology, 24*(3), 251-268. doi:10.1007/s13347-011-0015-x

van Wynsberghe, A. (2013a). Designing robots for care: Care centered value-sensitive design. *Science and Engineering Ethics, 19*(2), 407-433. doi:10.1007/s11948-011-9343-6

van Wynsberghe, A. (2013b). A method for integrating ethics into the design of robots. *Industrial Robot: An International Journal, 40*(5), 433-440. doi:10.1108/IR-12-2012-451

Vaneechoutte, M. (2000). Experience, awareness and consciousness: Suggestions for definitions as offered by an evolutionary approach. *Foundations of Science, 5*(4), 429-456. doi:10.1023/a:1011371811027

Vanlaere, L., & Gastmans, C. (2011). A personalist approach to care ethics. *Nursing Ethics, 18*(2), 161-173. doi:10.1177/0969733010388924

Wallach, W. (2017). [Personal Communication].

Wallach, W., Franklin, S., & Allen, C. (2010). A conceptual and computational model of moral decision making in human and artificial agents. *Topics in Cognitive Science, 2*(3), 454-485. doi:10.1111/j.1756-8765.2010.01095.x

Zeman, A. (2001). Consciousness. *Brain, 124*(7), 1263-1289. doi:10.1093/brain/124.7.1263

JOURNAL PAPER

Overcoming carer shortages with carebots: Dynamic value trade-offs in run-time.

Adam Poulsen & Oliver K. Burmeister

Australasian Journal of Information Systems

Overcoming carer shortages with carebots: Dynamic value trade-offs in run-time.

Abstract

The rising elderly population, and the reducing number of family and professional carers has led to calls for the intervention of carebots. Good care is the result of decisions which are 'determinative in practice.' That is, an AI carebot must perform social interactions in appropriate ways, tailored to the person in their care, at run-time. This study introduced a new carebot model which employs CCVSD to inform design and computational consciousness to make this determination by uniquely providing extrinsic care value ordering. Although VSD has been extensively researched in the information systems literature, CCVSD has not. The results of this study suggest that this new carebot model is competent in determining good, customised patient care at run-time. The contribution of this study is in its exploration of end-user willingness to trust known AI decisions and unwillingness to trust unknown AI decisions.

Keywords ethics, care, robots, consciousness, VSD, CCVSD

1. Introduction

There is an increasing amount of elderly going into care and a decreasing amount of carers to meet the needs of those elderly in care (Burmeister, 2016; Draper & Sorell, 2017; Garner, Powell, & Carr, 2016; Landau, 2013; Sharkey & Sharkey, 2011, 2012; Sparrow & Sparrow, 2006; Tokunaga, Tamamizu, Saiki, Nakamura, & Yasuda, 2017; Vallor, 2011). The attento model is a care robot model that intends to not only help overcome the lack of carers, but provide better care without human error and ill intent. It utilises care-centered value sensitive design (CCVSD) and computational consciousness. Value sensitive design (VSD) and related discussions of ethical value has been well accepted in the information systems literature, with many studies exploring technical, user and other perspectives (Boonstra & Van Offenbeek, 2010; Bowern, Burmeister, Gotterbarn, & Weckert, 2006; Chesney, Coyne, Logan, & Madden, 2009; Friedman, Kahn, & Borning, 2006; Friedman, Nathan, & Yoo, 2016; Friedman & Nissenbaum, 1996; Gotterbarn & Rogerson, 2005). However, CCVSD has arisen from engineering, with a focus on technical designs of robots, and attempting to address ethical concerns by exploring philosophical perspectives without, until this point, considering user perspectives (Draper & Sorell, 2014; Sharkey & Sharkey, 2011, 2012; van Wynsberghe, 2013a). The model suggests a specific design methodology and AI functionality. It recommends that carebots are designed to support the ordering of extrinsic patient care values (such as autonomy and dignity) according to patient preference, thus being able to make dynamic value trade-offs in run-time, so as to provide good, customised patient care. The contribution of this study is that it demonstrates end-user perspectives of such carebots.

The following section reviews care ethics, elderly care values, CCVSD, computational consciousness, and the attento model. Next the methodology is detailed. Then the findings are presented and discussed.

2. Literature Review

2.1 Care Ethics

Care ethics concern the principles that guide moral behaviour and action when taking care of someone. They are those that consider care to be of ethical value (Vanlaere & Gastmans, 2011). Vallor (2011), Upton (2011), Tronto (2010), Gámez (2009), van Wynsberghe (2013a), and Vanlaere and Gastmans (2011) demonstrate that care ethics and good care don't come from a normative ethical theory but rather they are "determinative in practice" (Beauchamp, 2004, p. 216). They come from our natural morality regarding concern for others. That is, just as a human carer must make decisions affecting those in their care, during the practice of caring (Upton, 2011), so too carebots must do so at run-time. Predetermined, hard-coded choices have limited application.

In care ethics it is considered ethical if a moral decision arises from the 'good' which is internal to practice, rather than external normative moral criteria or principles (Vallor, 2011). Good caring practices and relationships with carers are fundamental to care ethics. So what is an ethical action in caring for someone is the immediate good, for the patient, which is provided by care practice. Upton defines determinative as "the provision, for any given case, of a single, well-grounded and widely convincing recommendation as to the act that morally ought to be performed" (2011, p. 432). The antithesis is that good care is determinative in theory. A theory which is determinative in practice resolves the issue of an ethical theory being uselessly undetermined in practice by bridging the gap between

ethical principles and action. Upton notes that a deontologist has a duty to help others, but deontological ethical principles don't tell you when, how, how much, and whom you should help (2011). This is not unique to deontology, it can be seen in all normative ethical theories which simply provide principles. For example, utilitarianism's maximising of utility is simply that, it does not tell us how to maximise utility, for whom and when. Another example is virtue ethics, which holds virtues as ethical principles.

During the act of care, codes of conduct and ethics, as well as healthcare laws and regulations set ethical principles for carers to uphold. Care practices and processes inform principled actions, but what is lacking is good, customised patient care. But which practice is the most suitable, what kind of personal approach does a carer take, and how much care does a particular patient need are a few of the ethical decisions left undetermined by the principles. The principles don't determine the act which is best in specific situations. Instead, conscious carers make this determination in practice.

Literature reveals that good care ethics are determined by a carer's interpretation of a patient's individual needs, which can be expressed by values, when being cared for, as seen in the carer-patient relationship, and that carers have good intentions and take action based on these values (Caris-Verhallen, Kerkstra, van der Heijden, & Bensing, 1998; Gámez, 2009; Tronto, 2010; van Wynsberghe, 2013a).

2.2 Care Values

Information systems research has shown the importance of values in systems design thinking in many areas, including online applications (Chesney et al., 2009; Friedman et al., 2016), sensor development (Dechesne, Warnier, & van den Hoven, 2013), and in areas of ehealth (Boonstra & Van Offenbeek, 2010). However, to date little research has

investigated values in the design of carebots. A value is not only something desirable and that a recipient of care wants, but something preferable, it is something that they prefer to have or to have happen. Care values support the principles and behaviours which recipients of care expect from social interactions. Van Wynsberghe claims that it is "through the manifestation of these values that one comes to understand what care really is in practice" (van Wynsberghe, 2013a, p. 415).

Vallor states that "carebots are robots designed for use in home, hospital, or other settings to assist in, support, or provide care for sick, disabled, young, elderly, or otherwise vulnerable persons" (2011, p. 252). Sharkey & Sharkey refer to care robots as developments in robot applications that assist the elderly and their carers, monitor health and safety, and provide companionship (2012). Sparrow & Sparrow argue that care robots are incapable of meeting the social and emotional needs of the elderly. That the introduction of care robots, in such a healthcare sector that is already under economic pressure, would result in a "decrease in the amount of human contact experienced by older persons being cared for, which itself would be detrimental to their well-being" (Sparrow & Sparrow, 2006, p. 141).

2.3 Care-Centered Value-Sensitive Design

Care-Centered Value-Sensitive Design (CCVSD), like value-sensitive design (VSD), is a software engineering approach specifically for the design considerations of care robots. It offers prospective and retrospective evaluations of care robots. It narrows the VSD approach to care values, instead of broad human values. It pays tribute to care ethics, arguing centrally "that the care perspective provides an orientation from which to begin theorizing as opposed to a pre-packaged ethical theory" (van Wynsberghe, 2013a, p. 420).

It is this opposition to a ‘pre-packaged ethical theory’ or normative ethical theory that makes it clear that the CCVSD approach takes the position that care is determinative in practice. It's approach provides a “framework of components of ethical importance… along with a “user manual” for prospective evaluations” (van Wynsberghe, 2013b, p. 435),through its care centered framework (CCF).

The CCF “articulates the components that require attention for analysis from a care perspective” (van Wynsberghe, 2013a, p. 420). The components help designers to interpret, rank, and provide meaning of values; “the interpretation of values as well as their ranking and meaning differed depending on: the type of care (i.e. social vs physical care), the task (ex. bathing vs. lifting vs. socializing), the care-giver and their style, as well as the care-receiver and their specific needs” (van Wynsberghe, 2013a, p. 416). The interpretation of values is of key importance. Additional components of the CCF are the type of robot and manifestation of Tronto’s moral elements: attentiveness, responsibility, competence, and responsiveness (2010). The moral elements are manifested through describing a care practice in two competing contexts (human vs nonhuman carers). In these descriptions one reveals how the values are observed, prioritized, and interpreted depending on the context (van Wynsberghe, 2013a).

To follow the CCVSD methodology, one begins by identifying the components of the CCF and describing related traditional care practices to reveal values. One subsequently "speculates on what capabilities a robot ought to have to ensure the promotion of said values" (van Wynsberghe, 2013a, p. 424). Designers then make decisions concerning values with competing contexts in mind.

2.4 Computational Consciousness

Consciousness is what gives human carers a subjective experience, internal moral dialogue and an intrinsic understanding of situations involving patients (perceived as unique agents), patient values, their expression of said values, and extrinsic value ordering. Computational reasoning is part of that consciousness. Extrinsic value ordering comes under consciousness because it implies the actual conscious and empathetic understanding of values. With actual understanding comes the responsibility of ensuring value trade-offs are correctly customised for each patient.

A non-conscious carebot is deceptive, constantly tricking you into thinking that it understands how you feel and what you value. Having a conscious carebot might make a patient feel good because the conscious carebot is a sentient being with which you can have a positive and mutually subjective experience. Additionally, a conscious carebot is able to understanding its goals. To provide the goal of 'good care' to a non-conscious carebot would dangerous because computers seek out the most efficient path; it might understand 'good care' to be non-harm. So it might confine patients to their beds so they can't physically walk and therefore can't physically harm themselves. A conscious carebot would understand the implied sub-goals that a human intrinsically understands. A conscious carebot understands that someone in care should not be harmed, has freedom, has privacy, is respected, etc.

For example, Baars & Franklin employ Baars' Global Workspace Theory (GWT) model of human-inspired consciousness (1988), which they partially implement with the learning intelligent distribution agent (LIDA) model of computational consciousness (Baars & Franklin, 2009). In a personal communication with the lead author, W. Wallach (a consultant, ethicist, scholar, and author of papers on the LIDA model), stated that as "far as consciousness goes, your claims for CC [computational consciousness] are valid”

(Personal communication, July 13, 2017). Those claims are that computational consciousness, just like consciousness, provides: situation observation and evaluation; self-awareness and reactionary responsiveness; external stimuli perception; internal moral dialogue and intrinsic understanding of situations, patients, and patient values and their expression of values; and attentiveness. W. Wallach agrees that the claims are valid "but probably far from realizable with present day technology" (Personal communication, July 13, 2017). Most importantly to the internal moral dialogue, "the LIDA model helps integrate emotions into the human decision-making process, and we will elucidate a process... whereby an agent can work through an ethical problem to reach a solution that takes account of ethically relevant factors" (Wallach, Franklin, & Allen, 2010, p. 454).

2.5 The Attento Carebot Model

Extrinsic values can be logically embedded into an attento carebot, by design, as to ensure good care. Van Wynsberghe and Tronto's 'manifestation of values' (Tronto, 2010; van Wynsberghe, 2013a) is best identified and implementable by extrinsic values; that is to say intrinsic values are the manifestation of extrinsic ones. Extrinsic care values are orderable according to patient preference and therefore logically extrinsic vs extrinsic value trade-offs should be too.

Examples of intrinsic care include safety, emergency help, freedom, human rights, quality of life, trust, wellbeing, and comfort. Physical embedment of intrinsic values means that the robot's actual components must ensure good care and that they are ensured by unchanging design. Examples include (a) all internal wiring is protected by a hard outer casing so patients can't electrocute themselves on the wires to ensure patient safety; and (b) pattern recognition software detects and interprets a patient's foot logically and avoids

stepping on it to ensure patient safety; this process was logical. Intrinsic values must be upheld so as to ensure good physical and mental care over extrinsic values. Due to this requirement they can't be ordered. Intrinsic vs. intrinsic trade-offs have to be a decision for the designer (and based on documented evidence) since the robot can't change its physical design or its related functions during operation.

Extrinsic values are not the end goal in care, but rather they inform moral actions to reach intrinsic value end goals. Extrinsic values include: consent, dignity, respect, autonomy, independence, social connectedness, and privacy. Logical embedment of extrinsic values means that the robot stores the values in memory as a list (patient value priority list) which is uniquely ordered to each patient according to what that patient values. By allowing them to be ordered depending on a patient's preference, the robot can perform dynamic run-time value trade-offs therefore ensuring good, customised patient care.

An attento is to be designed utilising CCVSD with the addition of implemented computational consciousness for dynamic value trade-offs in practice, supported by extrinsic value ordering, and AI functionality (Baars & Franklin, 2007, 2009; Franklin et al., 2007; Wallach et al., 2010). In an attento, consciousness additionally provides a subjective an empathetic experience, internal moral dialogue and intrinsic understanding of situations, patients (as perceived as a unique agent), and patient values and their expression of said values.

To perform dynamic value trade-offs, consciousness is used to observe patients and situations, evaluate what is observed, and use that evaluation (or interpretation) to order and affirm a patient's value priority list. During interactions the attento carebot selects an action that is customised for that patient with their value list with the intent of prioritising

decisions and actions that affirm the patient's prioritised extrinsic values. For example, if a patient highly values their dignity, then it won't assume it can enter a locked room even if it hears a scuffle from a patient inside. Instead it will attempt to talk to the patient in the room from the other side of the door. It does this because it determines the patient might be undignified and entering the room would upset the patient.

The basic AI functionality of an attento doesn't provide it the capacity for conscious interpretation and understanding via a subject experience. If one were to design an AI that has situation observation, evaluation, and being able to logically order patient values then that would be simply an AI; and one with no consciousness. Such an AI is valuable in a limited standardised care kind of way, but doesn't provide good, customised patient care. An attento's purpose is to customises care, it fills the space between the situation evaluation and the value ordering; there lies interpretation (being able to derive the meaning of something and explain how its evaluation arrives at the meaning) and understanding (sympathetic awareness). The goal is a sensitive, trusting relationship with the carer who can interpret patient actions and understand what care is required.

Furthermore, affirmation is important. A human carer affirms what a patient values in care by asking them, listening to what a patient is telling them, conferring with other carers, and performs a self-check of what they have interpreted and now understand to be a patient's values. Similarly, an attento would have this affirmation functionality to ensure it is interpreting and ordering patient values accurately.

2.6 Medicine Delivery Attento Design

To be able to test the validity of the attento model, an elderly care medicine delivery attento was designed according to the above model. The CCF was employed for two

interactions which helped design the medicine delivery attento: (1) interacting with the patient and (2) interacting with the chemist clerk. The CCVSD methodology analysed these interactions, revealing values which in turn inspired a design. Within the CCVSD methodology, competencies must be met by components, so those were researched and chosen or designed; this was the theoretical design for a medicine delivery attento. The design features these competences and, competency meeting, components to meet the required attentiveness and responsiveness: medicine scheduling with a patient and medicine database; mobility with wheels; navigation with local mapping and pathing software; grasping, extending, and releasing items with articulated arms and claw-like hands; medicine error checking with a pharmacy information system (PIS) which provides medicine prescription monitoring, as well as pattern recognition software to identify incorrect medicines or dosage; interpersonal semantics and language capacities with audio input sensors and interpretation software, as well as audio response/greeting generation software and audio output device; interpersonal somatic capacities with visual input sensors and somatic interpretation software, as well as somatic response/gesture generation software; attentive situation evaluation and extrinsic value ordering with computational consciousness; patient distress/decline recognition with interpersonal semantic and somatic functionality, as well as walking gait sensors, for detection, and software for interpretation; medicine prescription and prescription adjustments with automated computerised prescriber order entry software, as well as a PIS; provide comfort with existing AI companion software such as the Responsive Interactive Advocate companion (Garner et al., 2016) or Virtual Care Giver (Tokunaga et al., 2017); medicine safety with a built in locked medicine box for storing medicine, as well as self-cleaning capacities, and physical aids such as human biological material sensors, to prevent spreading of sicknesses amongst

patients; and easily and quietly communicating with staff with messaging and notification software.

3 Method

The study employed a social constructionist, interpretivist methodology. Data was gathered in two phases. The first was a heuristic, expert evaluation. The second an online survey. Both utilised the elderly care medicine delivery attento design as well as scenarios involving carebot-patient interactions.

The first phase was conducted in June to July 2017, with four participants whose expertise was deemed to be representative of the range of fields relevant to this topic. These included a registered nurse, a robotics academic, a computer ethicist, and a computer scientist. The participants were presented with the attento model, the social phenomenon (elderly care medicine delivery robots caring for the elderly in a nursing home), and the medicine delivery attento design. They were instructed that they could be as formal or informal as they wished. This instruction is key to providing data that is as near to natural as possible to support the –interpretivist, social constructivist methodology.

Participants were presented with the medicine delivery attento's design (competencies and components) and three scenarios. Those scenarios demonstrated where the attento encountered a situation where values should be considered, at which point it examined the relevant patient's value priority list, evaluated the situation and revealed relevant extrinsic values, and made a decision based on the situation variables and prioritised extrinsic values.

Each scenario consisted of the same initial interactions, but the patient's prioritised extrinsic values changed so as to demonstrate the dynamic value trade-offs. With the

change, the attento's action may have changed too, depending on the patient's prioritised extrinsic values (particularly the highest priority one) and if that value was relevant to change the attento's decision and action.

Each scenario had an intrinsic value as the goal of the encounter. Each of the changing highest priority extrinsic values were designed to ensure the intrinsic value was met. For example, the first scenario's intrinsic value (and goal) was quality of life, and what the patient values the most changes from autonomy to respect and then to dignity.

For each scenario, participants were asked to imagine an elderly person they know and put themselves in their shoes so as to get a more relevant perspective. After each scenario change, participants were asked: "*What do you think they would make of their care robot if it made a decision like this for them?"*

To test the validity of the attento model further, an online survey, the second phase involved using Survey Monkey. The same scenarios involving the carebot-patient interactions were used, with minor amendments, following the analysis of the first phase of the study. Amendments suggested by the results of the first phase included the removal of direct persuasion attempts by the attento to get patients to take their medicine, which was seen as unethical by Phase 1 participants.

The online survey was conducted from August to September 2017, with one-hundred and two random global participants. Following the methodological approach, the survey consisted mostly of open-ended questions, with same demographic data collected too.

Participants were presented with two scenarios, not three, as in the first phase of the research. After the initial scenario and after each change in value, a question about trust in the attento's decision and another about value-sensitivity of the attento were asked. At the

end of each scenario participants were asked an open-ended question concerning their thoughts and feelings about the attento considering the dynamic value trade-offs it made.

Scaled and multiple choice questions were asked, each followed by open-ended questions. The scaled and multiple choice questions were asked to help participants identify and categorise their responses, and to stay engaged. The scaled question queried the acceptability and usability of the attento model by asking how likely they were to accept and use it in their everyday lives. Questions addressed the trustworthiness of the attento in the scenarios, and dynamic value trade-offs in practice.

Data analysis involved thematic analysis (Morse, 2008). Approval to conduct the study was given by the University Human Research Ethics Committee. The three scenarios used across the two phases of the research are detailed next. Scenario three was only used in the heuristic evaluation.

3.1 Scenario 1: Interrupting a Patient While They are Socialising With Friends

This scenario started by declaring that quality of life is the goal value and that the patient's highest priority value is autonomy. Then the scenario was described:

> *Imagine your elderly person in mind has not realised they're late to take their medicine while socialising and playing board games with their friends. But, they're within a limited time period in which it's safe to wait to take their medicine for a while. Their care robot is deciding whether to remind them or let them remember on their own. It has previously observed them and knows that they value being able to remember things on their own most of all. Their care robot is aware that they're late, but decides to not tell them yet unless it becomes potentially unsafe to wait longer.*

Then the highest priority value changed to respect:

> *Now imagine that instead of valuing remembering to take their medicine on their own, your elderly person in mind values not being interrupted while with their friends most of all. So their care robot again decides to not interrupt them because it remembers what they value.*

Here we can see how two different extrinsic values, which are at different times the most valued one by the patient, can influence the same action, whilst ensuring the same intrinsic value. This is where one can see the importance of patient and situation evaluation and action selection.

Then the highest priority value changed to dignity:

> *Now imagine that instead of valuing not being interrupted while with their friends, your elderly person in mind values their dignity most of all. Their care robot doesn't see how their dignity would be sacrificed by an interruption in this situation. So it decides to do so and remind them to take their medicine.*

Here we can see how another extrinsic value may have no relevance to the care robot's decisions and actions. It further illustrates the importance of patient and situation evaluation and action selection.

3.2 Scenario 2: Intruding on a Patient

This scenario started by declaring that wellbeing is the goal value and that the patient's highest priority value is privacy.

> *Imagine your elderly person in mind was getting out of bed partially dressed, having woken up late to take their medicine. And their care robot is deciding*

whether to enter their room or knock; it doesn't know that they are partially dressed and it's assuming that they are awake. Previously, the care robot has observed that they heavily value their privacy. The care robot decides to respect their privacy and simply knock, instead of letting itself in, as to ensure their wellbeing.

Then the highest priority value changed to independence:

Now imagine that instead of valuing privacy, your elderly person in mind values their independence most of all. So their care robot could decide to call for a human care-giver but decided not to because it's aware that they value independence, and instead again simply knocks on the door.

Then the highest priority value changed to social connectedness:

Now imagine that instead of valuing their independence, your elderly person in mind values social connectedness most of all; they appreciate an unexpected visit with all members of staff and all patients, and they welcome them to just let themselves in. Their value priority list, as observed by the care robot, places their dignity as the 'least valued' and privacy in the 'less valued' category. So their care robot, knowing this ordering, decides to let itself in and reminds them to take their medicine.

Priority Categories	Orderable Values
Most valued	Social connectedness
Highly valued	Independence
	Autonomy
Indifferently valued	Respect
Less valued	Consent
	Privacy
Least valued	Dignity

Table 1: Patient's value priority list demonstration presented to Phase 1 participants

Here we can see the introduction of the importance of the patient's value priority list. We can see that a patient's preference for social connectedness changed the care robot's action. On first impression it seems as though the patient's dignity and privacy were sacrificed, however the patient values social connectedness most of all. The care robot is aware that the patient places low value on privacy and dignity, and determined that its unexpected entering was okay by the patient. If however, dignity or privacy where ranked in the highly valued categories, then the care robot would take that into its consideration and action selection; most likely resulting in it deciding to ask another patient or carer to enter instead, so as to ensure social connectedness, or simply knocking.

3.3 Scenario 3: Calling For a Human Carer to Help With a Patient

This scenario started by declaring to the participants that emergency help is the goal value and that the patient's highest priority value is respect.

> *Imagine your elderly person in mind is convinced they've taken their medicine, but their care robot is telling them that they haven't, in front of their friends. Their care robot experiences this event with them often, and at this point it must involve a human carer. Their care robot is deciding whether to inform them or not of its need to call for help. It has previously observed that they demand respect most of all. So it decides to show them respect and tells them that it will now call for a human carer to help them.*
>
> *Do you think it would be okay for the robot to tell a lie" as to convince them to take their medicine? In this situation a lie could be like saying "I've called for a carer to come and help you, if you take it now I can tell them to not bother coming." This was said even though a carer had not been called yet. By telling this simple and non-harmful lie, the care robot takes into consideration the patient's wellbeing and value of emergency help, as well as human carer workload which would be increased by having to call for them.*

Then the highest priority value changed to dignity:

> *Now imagine that instead of valuing respect, your elderly person in mind values their dignity most of all. So their care robots decides to leave them and their friends and quietly and secretly calls for emergency help.*

4 Discussion of Results

Regarding the first phase, three participants agreed the attento model is value sensitive and would deliver good, customised patient care. There was no consensus on the acceptability of the current model among participants. Each participant had differing levels of trust in an attento and the extrinsic value ordering functionality.

Each participant expressed concerns about how fast an attento would have to be, to adapt to new patients, as well as patients who change their minds on what they value; could an attento keep up. The literature supports these concerns. One group of researchers in Canada are working on adaptive devices for people with some forms of dementia, where it has been shown that as the disease progresses, personality and values can change and thus there is a need to adapt to changing values (Lin et al., 2014). Furthermore, an international study on assistive technology has shown that current assistive technology is not yet able to make real-time decisions, outside of highly controlled laboratory experiments; but they predicted that within five years it would be possible (Teipel et al., 2016).

Each Phase 1 participant objected to an attento providing comfort. The registered nurse opposed the idea that an attento would replace a human carer and instead argued that it would be a valuable assistant. Exemplary answers to the following question are listed next. Would you expect this kind of care from a human carer?

> "I would certainly expect such care from a human carer, but the robot is likely to be more consistent, and can be fine-tuned with potentially little effort to adjust for anomalies. A human carer, on the other hand, can be devious or malcontent in their work. There has certainly been no shortage of reports of human carers committing grave atrocities" (Computer Scientist).

> "It may be able to do the desired tasks with a better level of care (or perceived) care than a human staff member" (Registered nurse).

On the one hand, each of these three participants questioned the value sensitivity of human carers and were inclined to trust in the value sensitivity of an attento, and in its ability to provide good, customised patient care. On the other hand, as supported by the Lin et al. (2014) research cited above, the computer ethicist didn't know, stating: "is AI/Affective Computing to be able to deal with this? At the moment, definitely not up to scratch"

Regarding, the speed of adaptability. The computer scientist noted the potential for an attento's affirmation function to become "insufferable", if it is constantly trying to affirm its interpretation of the patient's values with the patient. The computer ethicist agreed, saying the robot might be "too sensitive about its re-ordering processes ... I find it a bit creepy now when technologies predict my behaviour."

Regarding trust, the robotics academic deemed that an attento would be untrustworthy for multiple reasons: an attento's autonomous capacity for medication prescription which he thought should be left to human carers. However, there are already multiple, automated medication dispensers available on the market. The registered nurse said it would be trustworthy provided it was: "checked by a human nurse or aged cared assistant till trust was built".

That participant also recommended that critical medication be prioritised, and that the way an attento observed and evaluated a situation be clearly understood.

The computer scientist said it would be trustworthy provided: "there were well-tested mechanism in place to detect the failure to take the medicine on the part of the patient, and a proven performance in providing customised and gentle care to a patient". Furthermore,

the same participant recommended that the attento's observation and situation evaluation tools were available to users. Thus, it seems that both the above participants would value having the ability to audit the decision making of the robot. That is something that is not possible with human care givers, and perhaps not for AI with consciousness.

The computer ethicist would trust an attento "[p]robably more than I would their ability to be my friend," if it couldn't be hacked, demonstrated empathy, and it's appearance was comforting.

Regarding objections to comfort, the robotics academic noted the if it was like a pet then they and probably others would be happy with the implicit comfort provided. The computer scientist noted that the comfort could be "hollow should the robot be perceived as merely a machine unable to sympathise." However, animal-like carebots have been used successfully as companions for the elderly for some time such as Paro which helps to produce positive physiological effects (Robinson, MacDonald, & Broadbent, 2015) and improve wellbeing (Jung, van der Leij, & Kelders, 2017) in the elderly. Thus the concerns expressed by Phase 1 participants have already been overcome in other contexts, which can be used to inform the design of carebots for the elderly.

Lying and persuading was offered as behaviour to get patients to take medication. Participants took issue with lying, for instance:

"This would be unacceptable in an aged care assistant, Enrolled or Registered Nurse or Medical Doctor. Robot would have to comply with Australian Health Professionals Regulation Agency (AHPRA) Code of Conduct and Professional Standards for Nursing ... in short no threats, intimidation, misdirection or lying" (Registered nurse). Although the

professional code for nurses may not apply to robotic carers, the sentiment is valid, and some compliance with codes of care practice is likely to be required.

Some people "[m]ay lose trust in the system if the robot goes behind their back ... Blackmailing the patient: Depends whether the person really had taken the medication (human intervention that the robot was not aware of). Best that it calls for help, rather than coerce the person into taking a double dose accidently" (Robotics academic).

This counters van Wynsberghe's comments concerning a proposed attento method of physical persuasion which involved bringing the medicine to the mouth of the patient: "[T]his is a big debate in ethics right now, how far can the robot go? Would think that instead of the robot doing this, if we're in a hospital, the robot should call on the doctor? Maybe instead of the robot bringing meds to the patient's mouth the robot stands in their way (i.e. if they're watching TV then can't see it anymore) or talks so that the person can't ignore them...? That seems less physically intrusive" (Personal communication, June 30, 2017). This is not lying, but is another way to get patients to take their medicine – persuasion. However, van Wynsberghe's potentially bothersome robot behaviour, which aims to persuade patients to take their late medicine, could be seen as threatening or intimidating, which the registered nurse warns is not acceptable.

Regarding the survey, at its closing there were fifty completed surveys (except one of those fifty skipped one question) and fifty two partially completed surveys. The fifty completed survey's data was used to tabulate the multiple choice and scaled question answers. The open-ended question's data from all survey attempts was used for the discussion.

Regarding the second phase, the results of the survey show that care must be customised for the patient and determinative in practice, supporting the concept of extrinsic value

ordering and the attento model itself; participants supported this. One participant said that care must be customised because "every human is different and would need to be treated accordingly." Another participant added that "Each person would have very different preferences as to how they liked to be cared for." Similarly, (Lin et al., 2014) have demonstrated that one size does not fit all, because some people prefer assistance to be firm and directive, whilst others refuse to be directed, and prefer suggestive assistance, which gives them a feeling of remaining in control of their own decision making. Furthermore, a comprehensive report of assistive technology for aged care across Europe, in which robotics was seen as a new technology, also strongly advocated for customised, individualised care (Alzheimer Europe, 2010). Thus, participant responses confirm research in other parts of the world, that care must be customised for the patient and determinative in practice.

The attento was also found to be usable and acceptable by sixty-six percent of participants who indicated that they were either very likely or likely to use the attento (Figure 1).

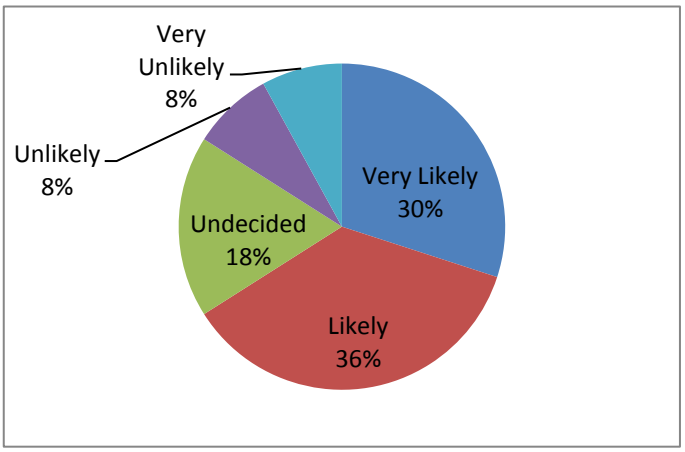


Figure 1: Distribution of answers to: How likely are you to accept and use your care robot in your everyday life?

Table 2 and 3 demonstrate each of the three times the highest priority value changes in the scenarios by the three sets of numbers under each scenario column. As seen in Table 2, thirty-eight out of fifty (on average from scenario related question results in Table 2) of respondents found that the attento was value sensitive and proficient in determining what constitutes as good care. However, this was only in cases where participants knew the scenario and decision of the attento (Table 3). Table 2 indicates that participants trusted the attento to make value sensitive decisions so it could perform the care practice 'deliver medicine' in a way that constituted good, customised patient care. One participant valued the "personal touch [the] robot offers whilst recognizing personal likes and dislikes." Another participant said that emotions "change and so do values depending on how you feel at the time." These likes, dislikes, and emotions are what we value in our own personal way in care. Work on adaptation to emotions is still in its infancy (Lin et al., 2014; Teipel et al., 2016), but it supports the contention of Phase 2 participants, namely that it is another important area of real-time, dynamic customisation.

Do you feel the care robot has respected the way you want to be treated and behaved appropriately?	Scenario 1			Scenario 2		
Yes	45	40	20	45	41	38
No	2	5	25	3	3	10
Undecided	3	5	5	2	6	2

Table 2: The value sensitivity of the attento's decisions

Would you trust your care robot's decision in this scenario?	Scenario 1			Scenario 2		
Yes	40	30	35	48	39	37
No	4	13	11	2	7	9
Undecided	6	7	4	0	4	3

Table 3: The trustworthiness of the attento's decisions

In both scenarios, every time the highest priority value changed, thirty-eight out of fifty (on average from scenario related question results in Table 3) participants thought the attento was trustworthy (Table 3). But, in cases where the scenario and decision was unknown the majority of participants didn't trust the attento inherently (Figure 2), adding that they would want the ability to change how it makes decisions (Figure 3). It appears that the scenarios inspired trust because the attento has past interactions and existing trust with the patient. Additionally, the scenarios explain the attento's decision making process. Participants aren't told exactly how that operation works and thus it doesn't inspire trust inherently. One participant summarised it saying that most scenarios "give me confidence that the care robot is acting appropriately ... because I understand through the scenarios how the robot makes it decisions." Thus, trust in the attento wasn't inherent for two reasons: trust is a social construct (Burmeister, Weckert, & Williamson, 2011) and people want to understand how a machine with ethical implications works before trusting it inherently.

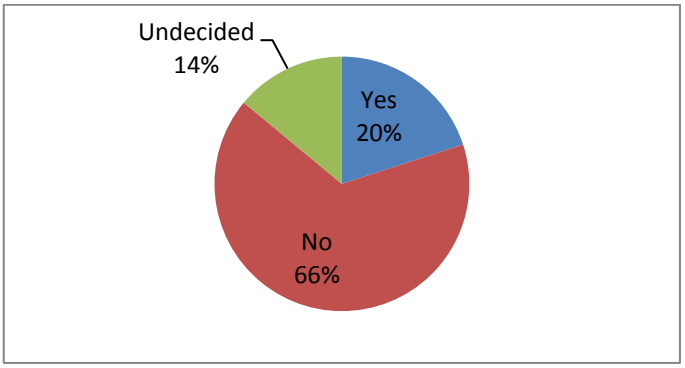


Figure 2: Distribution of answers to: Would you inherently trust your care robot or would you need to understand how it makes its decisions?

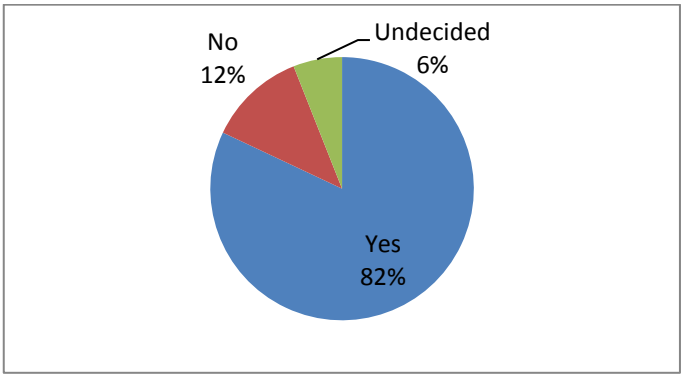


Figure 3: Distribution of answers to: Would you want the ability to change how your care robot makes its decisions?

A social construct is something that appears to be reality but is actually spawned from, and accepted by, a group of people (sports club, religious organisation, or even a whole society). "As it is with a human, I think repeated interaction with the robot would lead to a building of trust" one participant stated. Trust is formed from a growing, mutual, and active personal relationship over a period of time. It's construction is formed by social factors such as empathy, camaraderie, social status, social impression, job type and position, etc (Burmeister et al., 2011). For a carebot to have social factors "it would somehow need to know subtle social nuances" suggested one participant.

Some participants didn't like the idea of the attento having social factors however, like one who said their "interest in being social does not pertain to robots." So for patients like that participant, developing trust would be hard. However, it should be remembered that some people are less trusting than others. In other words, not all human carers are trusted by those in their care, so there is no reason to assume that all people will trust a carebot, no matter how well it is designed.

Many participants said they would need to understand the thought process of the attento to trust it inherently. For those participants, developing trust for carebots relies on it making a good impression towards them personally or indirectly through people having a good social opinion of it. Furthermore, as seen in the discussion of Phase 1 results, above, patients are likely to assume that codes of conduct and healthcare laws hold human carers accountable. The attento didn't inspire any such assumptions; "I would need to know how it has worked out for others first" concluded one participant.

Uniquely to carebots, many participants commented saying they required understanding its thought processes to develop trust. It seems that because people assume that we can understand the decision-making processes of machines then they should have the ability to examine it. Exemplary quotations include: "I would feel safer and more comfortable knowing how the robot works", and "It's a machine. I need to know how it's programmed".

The assumption that we can understand a machine's decision-making processes isn't always correct. If a carebot's AI ethical decision-making comes from a strict set of rules then it would be possible to know and supply to patients. But, that would require the AI's determination of good care to be determinative in theory. An attento is conscious so it is not likely that patients could examine the way it makes ethical decisions, nor change the

way they make them, for the same reason we can't interpret how another human makes decisions. Each human has a subjective experience with a unique past that forms their contextualized morality that no one else could understand. Since we can't understand, we rely on social factors belonging to that person in order to develop trust in them. Trust in one to perform good care is a social construct for that reason. One participant said, regarding the attento, that if "it demonstrates ethical behaviour and care for my needs I would trust the judgments of its programming. Just as I trust other people without fully knowing how they decided on an action." One can conclude that attentos need to demonstrate more human-level social factors for patients to trust it like they would a human carer.

Most participants desired customised care with the ability to change how the attento makes decisions. This is a contradiction given that the attento is designed to ensure customised patient care. Exemplary responses included: "every human is different and would need to be treated accordingly", and "It would need to fit into my lifestyle and my way of living." Causal factors were beyond the scope of this research, however the following may explain why this contradiction occurred. Firstly, for reasons discussed above, participants don't trust the attento. Secondly, participants lack AI technical knowledge. Participants may not be aware of the potential for AI to be conscious, as well as the speed of AI computation and adaption. In the case of the attento model, this refers to the speed of the AI to identify, proiritise, and affirm values. Thirdly, participants lack technical knowledge on the attento model. Participants weren't informed of the technical details such as the four levels of affirmation.

Participants were concerned that the attento might focus on extrinsic values and let patient wellbeing suffer. As one participant declared, "health is more important than socializing ...

no compromise about the health". Health is an intrinsic value. The attento model ensures intrinsic values by design, they will always override extrinsic values such as social connectedness. Participants weren't informed of this design factor, but their overwhelming concern demonstrates support for the attento model. Another participant asserted that their "privacy is very important ... [and that] will not change unless there is an emergency." Ensuring that emergency help, an intrinsic value, overrides privacy, an extrinsic value, is a key principle of the attento model, well supported in the care ethics literature (Alzheimer Europe, 2010; Teipel et al., 2016).

5. Further Steps

Additional studies need to be done to further test the validity of the attento model. New lying and persuasion related ethical scenarios should be designed and tested. Additionally, concerning issues with comfort, a literature review on acceptable levels of comfort from carebots should be completed.

The revised attento model should be presented to a focus group with elderly participants to get targeted results. This study only had twelve participants aged between fifty and seventy-nine who completed the survey, but ideally a study involving older people only, including both those in care and their carers, would give further insights. All participants from the first phase, and many from the second phase, needed an implementation of the model to determine if they would trust its capacity to provide good, customised patient care. This could be done in the future.

6. Conclusion

The number of elderly in care is increasing, and the number of carers is decreasing; carebots are a solution this problem. But, carebots need to demonstrate care, not merely

'elderly management' which many current implementations convey (Garner et al., 2016; Keel, 2002; Tokunaga et al., 2017). Care is determinative in practice and the attento model provides good, customised patient care by achieving dynamic value trade-offs, in run-time as to meet this criteria of care. Three out of four Phase 1 participants and the majority of Phase 2 participants agree.

Regarding the first phase, due to the varying levels of trust in an attento, and unanimous dislike for any attempts at comfort, revisions to the model need to be made and tested with further studies. Achieving a good outcome (timely medicine taking) through lying and persuasion, was not seen as acceptable to half of the Phase 1 participants.

Regarding the online survey, the contradiction of trust, found when comparing known attento scenarios and decisions to those that are unknown, requires revisions to the model to increase trust. Although that was the case, it is normal for new technologies to be approached with caution and especially for those that concern vulnerable peoples. For instance, the design could focus on increasing social interactivity and other social factors. This is especially important because of the twelve Phase 2 participants aged between fifty and seventy-nine who completed the survey, six said they were very likely (the other four said likely) to accept and use the attento presented. But, one third of those participants also said they would inherently trust the attento, whilst only thirteen percent of lower age groups said they would. The elderly were willing to accept and use, as well as more willing to inherently trust, the attento. This is a good indication of the validity of the model. But it is also a concern, indicating that some elderly may be so desperate for care that they deem a new technology acceptable, usable, and more inherently trustworthy than younger generations would before seeing it. Since the number of elderly willing to

inherently trust the attento is still low, albeit higher than other age ranges, further studies are needed to refine the attento model.

References

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.

Baars, B. J., & Franklin, S. (2007). An architectural model of conscious and unconscious brain functions: Global workspace theory and IDA. *Neural Networks, 20*(9), 955-961. doi:10.1016/j.neunet.2007.09.013

Baars, B. J., & Franklin, S. (2009). Consciousness is computational: The LIDA model of global workspace theory. *International Journal of Machine Consciousness, 01*(01), 23-32. doi:10.1142/s1793843009000050

Beauchamp, T. L. (2004). Does Ethical Theory Have a Future in Bioethics? *The Journal of Law, Medicine & Ethics, 32*(2), 209-217. doi:10.1111/j.1748-720X.2004.tb00467.x

Boonstra, A., & Van Offenbeek, M. (2010). Towards consistent modes of e-health implementation: structurational analysis of a telecare programme's limited success. *Information Systems Journal, 20*(6), 537-561. doi:10.1111/j.1365-2575.2010.00358.x

Bowern, M., Burmeister, O. K., Gotterbarn, D., & Weckert, J. (2006). ICT Integrity: Bringing the ACS Code of Ethics up to date. *Australasian Journal of Information Systems, 13*(2), 168-181. doi:10.3127/ajis.v13i2.50

Burmeister, O. K. (2016). The development of assistive dementia technology that accounts for the values of those affected by its use. *Ethics and Information Technology, 18*(3), 185-198. doi:10.1007/s10676-016-9404-2

Burmeister, O. K., Weckert, J., & Williamson, K. (2011). Seniors extend understanding of what constitutes universal values. *Journal of Information, Communication and Ethics in Society, 9*(4), 238-252. doi:10.1108/14779961111191048

Caris-Verhallen, W. M., Kerkstra, A., van der Heijden, P. G., & Bensing, J. M. (1998). Nurse-elderly patient communication in home care and institutional care: an explorative study. *International Journal of Nursing Studies, 35*(1), 95-108. doi:10.1016/S0020-7489(97)00039-4

Chesney, T., Coyne, I., Logan, B., & Madden, N. (2009). Griefing in virtual worlds: causes, casualties and coping strategies. *Information Systems Journal, 19*(6), 525-548. doi:10.1111/j.1365-2575.2009.00330.x

Dechesne, F., Warnier, M., & van den Hoven, J. (2013). Ethical requirements for reconfigurable sensor technology: a challenge for value sensitive design. *Ethics and Information Technology, 15*(3), 173-181. doi:10.1007/s10676-013-9326-1

Draper, H., & Sorell, T. (2014). Robot carers, ethics, and older people. *Ethics and Information Technology, 16*(3), 183-195. doi:10.1007/s10676-014-9344-7

Draper, H., & Sorell, T. (2017). Ethical values and social care robots for older people: an international qualitative study. *Ethics and Information Technology, 19*(1), 49-68. doi:10.1007/s10676-016-9413-1

Alzheimer Europe. (2010). *Alzheimer Europe Report. The ethical issues linked to the use of assistive technology in dementia care.* Luxembourg: Alzheimer Europe.

Franklin, S., Ramamurthy, U., D'Mello, S., McCauley, L., Negatu, A., Silva, R., & Datla, V. (2007). *LIDA: A computational model of global workspace theory and developmental learning*. Paper presented at the AAAI fall symposium on AI and consciousness: Theoretical foundations and current approaches, Arlington, VA.

Friedman, B., Kahn, P. H. J., & Borning, A. (2006). Value sensitive design and information systems. In P. Zhang & D. Galletta (Eds.), *Human-computer interaction and management information systems: Foundations* (pp. 348-372). New York: M. E. Sharpe.

Friedman, B., Nathan, L. P., & Yoo, D. (2016). Multi-lifespan information system design in support of transitional justice: Evolving situated design principles for the long(er) term. *Interacting with Computers, 29*(1), 80-96. doi:10.1093/iwc/iwv045

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems, 14*(3), 330-347. doi:10.1145/230538.230561

Gámez, G. G. (2009). The nurse-patient relationship as a caring relationship. *Nursing Science Quarterly, 22*(2), 126-127. doi:10.1177/0894318409332789

Garner, T. A., Powell, W. A., & Carr, V. (2016). Virtual carers for the elderly: A case study review of ethical responsibilities. *Digital Health, 2*, 1-14. doi:10.1177/2055207616681173

Gotterbarn, D., & Rogerson, S. (2005). Responsible risk analysis for software development: Creating the software development impact statement. *Communications of the Association for Information Systems, 15*(40), 730-750.

Jung, M. M., van der Leij, L., & Kelders, S. M. (2017). An exploration of the benefits of an animallike robot companion with more advanced touch interaction capabilities for dementia care. *Frontiers in ICT, 4*(16). doi:10.3389/fict.2017.00016

Keel, C. (2002). Rudy the robot. *The American Journal of Nursing, 102*(8), 24. doi:10.1097/00000446-200208000-00026

Landau, R. (2013). Ambient intelligence for the elderly: Hope to age respectfully? *Aging health, 9*(6), 593-600. doi:10.2217/ahe.13.65

Lin, L., Czarnuch, S., Malhotra, A., Yu, L., Schröder, T., & Hoey, J. (2014). *Affectively aligned cognitive assistance using Bayesian affect control theory*. Paper presented at the International Workconference on Ambient Assisted Living (IWAAL), Belfast, UK.

Morse, J. M. (2008). Confusing categories and themes. *Qualitative Health Research, 18*(6), 727-728. doi:10.1177/1049732308314930

Robinson, H., MacDonald, B., & Broadbent, E. (2015). Physiological effects of a companion robot on blood pressure of older people in residential care facility: A pilot study. *Australasian Journal on Ageing, 34*(1), 27-32. doi:10.1111/ajag.12099

Sharkey, A., & Sharkey, N. (2011). The eldercare factory. *Gerontology, 58*(3), 282-288. doi:10.1159/000329483

Sharkey, A., & Sharkey, N. (2012). Granny and the robots: ethical issues in robot care for the elderly. *Ethics and Information Technology, 14*(1), 27-40. doi:10.1007/s10676-010-9234-6

Sparrow, R., & Sparrow, L. (2006). In the hands of machines? The future of aged care. *Minds and Machines, 16*(2), 141-161. doi:10.1007/s11023-006-9030-6

Teipel, S., Babiloni, C., Hoey, J., Kaye, J., Kirste, T., & Burmeister, O. K. (2016). Information and communication technology solutions for outdoor navigation in dementia. *Alzheimer's & Dementia, 12*(6), 695-707. doi:10.1016/j.jalz.2015.11.003

Tokunaga, S., Tamamizu, K., Saiki, S., Nakamura, M., & Yasuda, K. (2017). VirtualCareGiver: Personalized smart elderly care. *International Journal of Software Innovation, 5*(1), 30-43. doi:10.4018/IJSI.2017010103

Tronto, J. C. (2010). Creating caring institutions: Politics, plurality, and purpose. *Ethics & Social Welfare, 4*(2), 158-171. doi:10.1080/17496535.2010.484259

Upton, H. (2011). Moral theory and theorizing in health care ethics. *Ethical Theory and Moral Practice, 14*(4), 431. doi:10.1007/s10677-011-9295-6

Vallor, S. (2011). Carebots and caregivers: Sustaining the ethical ideal of care in the twenty-first century. *Philosophy & Technology, 24*(3), 251-268. doi:10.1007/s13347-011-0015-x

van Wynsberghe, A. (2013a). Designing robots for care: Care centered value-sensitive design. *Science and Engineering Ethics, 19*(2), 407-433. doi:10.1007/s11948-011-9343-6

van Wynsberghe, A. (2013b). A method for integrating ethics into the design of robots. *Industrial Robot: An International Journal, 40*(5), 433-440. doi:10.1108/IR-12-2012-451

van Wynsberghe, A. (2017). [Personal communication].

Vanlaere, L., & Gastmans, C. (2011). A personalist approach to care ethics. *Nursing Ethics, 18*(2), 161-173. doi:10.1177/0969733010388924

Wallach, W. (2017). [Personal Communication].

Wallach, W., Franklin, S., & Allen, C. (2010). A conceptual and computational model of moral decision making in human and artificial agents. *Topics in Cognitive Science, 2*(3), 454-485. doi:10.1111/j.1756-8765.2010.01095.x

DISCUSSION

This study aimed to investigate the ability for robots to provide good, customised patient care. The research question was: is it possible to create a CCVSD conscious robot using the LIDA model that is usable, user accepted, and value sensitive; thus proving that consciousness, in carebots, can provide good, customised patient care with dynamic value trade-offs in run-time? The research found that its theoretically possible with the attento model.

In light of the holes in literature, flaws in existing methodologies, the study purpose, and the research question, a new carebot model was theorised and designed. It then underwent usability, acceptability, and value sensitivity testing with potential end-users in two research phases to prove it is capable of being a solution to the research question. The attento model is a CCVSD conscious carebot one which recommends the performing of dynamic value trade-offs to determine, in practice, what constitutes as good, customised patient care with extrinsic value ordering. The model and the research makes contributions to literature.

1. CONTRIBUTIONS TO LITERATURE

This study makes three contributions to literature. Firstly, the attento model to carebot literature. Secondly, extrinsic value ordering to computational consciousness. Thirdly, the concept of dynamic value trade-offs in run-time to VSD literature, and the opinions of the research participants regarding the attento model to CCVSD literature.

1.1 Carebots

Herein lies the contribution to carebot literature. No conscious carebots, or those with the capacity to determine, in practice, what good care is, were found in literature. To make that contribution and provide a solution to the research question, the attento model was designed. A particular attento, an elderly care medicine delivery one, was designed and presented to participants in two research phases to represent the attento model itself. The attento model and the medicine delivery attento are comprehensively described in the journal paper within this study. Uniquely, the attento model features dynamic value trade-offs in run-time which it performs by identifying, ordering, and affirming extrinsic care values for each patient to provide good, customised patient care. In Phase 1 of the research, a heuristic, expert evaluation, three out of four participants agreed that the attento was value sensitive and that it provided good care. Sixty-six percent of the participants in Phase 2 of the research, an online survey, who completed the survey found the attento to be acceptable and usable. Seventy-six percent of the Phase 2 participants (based on an average of the scenario related value sensitivity questions) agreed that the attento model is value sensitive and proficient in determining what constitutes as good care in cases where participants knew the scenario and decision of the attento. However, in unknown cases the attento didn't inspire trust inherently. Other than fears of new technologies, two reasons may explain this: (1) trust is a social construct and (2) people want to understand how a machine with ethical implications works before trusting it inherently.

A social construct is something that appears to be reality but is actually spawned from, and accepted by, a group of people (sports club, religious organisation, or even a whole society). Trust is formed from a growing, mutual, and active personal relationship over a

period of time. It's construction is formed by social factors such as empathy, camaraderie, social status, social impression, job type and position, etc. Phase 1 participants had varying levels of trust. However, what they had in common was they wanted to see the attento implemented and working before trusting it inherently. Two out of the four Phase 1 participants also wanted to be able to examine its decision-making process of the attento. Thematic analysis of the Phase 2 results suggest that trust is a social construct. Most Phase 2 participants weren't willing to trust the attento inherently, in fact only twenty percent were. Furthermore, eighty-two Phase 2 participants said they would want the ability to change how the attento makes its decisions. This data, along with the fact that participants were willing to trust the attento in known scenarios, suggests that trust is a social construct. The known scenarios explained that the attento, and the patient being cared for, had past experiences and a mutual relationship. They also put the decision-making process on display and encouraged participants to put themselves in the shoes of the patient who was already trusting the attento. The fact that Phase 2 participants were more willing to trust the attento in the known scenarios reveals that trust is a social construct, that is to say it is something that is grown out of a nurtured mutual relationship.

1.2 Computational Consciousness

For an attento to perform the extrinsic value ordering method it needed to be implementable. This study recommended using computational consciousness, specifically the LIDA model, to implement that method. This study claims that extrinsic value ordering can be handled by the ethical decision-making process in the LIDA model. No other study makes this claim, nor is it suggested that the LIDA model be used for carebots. With consciousness, intentionality, and free will, which the LIDA model provides, an

attento is a full explicit AMA capable of the ethical decision-making and agency required to be a good carer.

In an attento, consciousness additionally provides care related functions: a subjective experience; internal moral dialogue; intrinsic understanding of situations, patients (as perceived as a unique agent), and patient values and their expression of said values. It also gives an attento general conscious functionality such as goal setting, situational awareness, etc. Consciousness helps to provide good care.

1.3 Value Sensitive Design

Dynamic value trade-offs in run-time, which are achieved with extrinsic value ordering, ensure that patients have power of the extrinsic value trade-offs, such as privacy, independence, and respect. Such trade-offs are typically made during the design phase of a new technology when they should be customised according to individual patients dynamically. Not all values are extrinsic though, some are intrinsic and should be ensured by design. This is so patients aren't harmed and have a quality level of basic physical care. The attento model was founded on upholding those two principles: dynamic value trade-offs in run-time and intrinsic values ensured by design. Instead of limiting how care can be customised towards individual patients with value trade-offs during the design phase, extrinsic value ordering is used. Which allows in an attento to make trade-offs in run-time and thus making the determination of good care in practice. Three out of four Phase 1 participants and seventy-six (based on an average of the scenario related value sensitivity questions) in the online survey agreed that the attento model is value sensitive. Due to this result, this study recommends dynamic value trade-offs in run-time for all technologies

developed using VSD where it is safe to allow extrinsic values to be customised according to end-user preference.

1.3.1 Care-Centered Value Sensitive Design

The attento model itself, and the medicine delivery attento (Appendix B), were designed using the CCVSD methodology (Appendix A). The medicine delivery attento was presented to participants in the two research phases to gather an end-user perspective on CCVSD. Considering that the attento model was theorised as a result of using the CCVSD methodology and that the majority of participants in both research phases said the attento model was value sensitive, it is fair to conclude that CCVSD itself is value sensitive. However, it was found, in both research phases, that the attento didn't inspire trust inherently. In Phase 1, each participant distrusted at least one feature of the attento presented. None had inherent trust for the attento without: (1) knowing how the attento's decision-making process, (2) seeing an actual implementation, and (3) knowing it has been tested. In Phase 2, seventy-six of the participants (based on an average of the scenario related trustworthiness questions) trusted the attento when they knew the scenario, the attento's decision-making process, and the subsequent decision based on the patient's value priority list (Appendix C). When the situation is unknown, sixty-six percent of Phase 2 participants didn't have inherent trust in the attento. Participants in both research phases distrusted the freedom that the attento had to act differently and make decisions with dynamic value trade-offs in run-time. Eight-two percent of Phase 2 participants wanted the capacity to change how the attento makes decisions; proving distrust.

This study concludes that a CCVSD carebot, like an attento, is value sensitive where trust is established. But, such a carebot doesn't inherently inspire trust since trust is a

social construct that requires a mutual and ongoing relationship to establish. However, any new technology encounters issues with trust. Additionally, a carebot is going to have issues developing trust if it doesn't attempt to establish a relationship with patients and doesn't maintain social factors such as empathy, camaraderie, social status, social impression, etc. For a CCVSD carebot to encourage inherent trust, more emphasis needs to be put into making a carebot sociable like a companion carebot. This study recommends more emphasis be put into increased sociability of carebots designed using CCVSD.

This study provides end-user perspectives on CCVSD. The research presented a particular CCVSD carebot, the attento, and directly asked for end-user opinions. Those opinions helped to answer the research question. This study, in light of the research findings, concludes that the attento model, a CCVSD conscious carebot using the LIDA model, is usable, user accepted, and value sensitive; thus proving that consciousness, in carebots, can provide good, customised patient care with dynamic value trade-offs in run-time. The attento model was theorised as a result of using the CCVSD methodology, and the research concludes that the attento is value sensitive. Therefore, the research on the attento model provides positive end-user perspectives on CCVSD.

2. LIMITATIONS OF THE PRESENT RESEARCH & POSSIBILITIES OF FUTURE RESEARCH

Since this research was limited by the sampling, the number of elderly who participated in the research phases was minimal. Future research using focus groups targeting the elderly exclusively would provide more insights regarding the usability, acceptability, and value sensitivity of the attento model. It is very possible that an attento implementation could be built in future research to further test the model and arrive at a practical solution to the

decreasing number of carers and rising number of elderly in care. Attento's aren't necessarily limited to elderly care, they could be used in other fields such as rehabilitative care, mental health care, or paediatric care.

CONCLUSION

The research phases within this study prove that it is possible to create a CCVSD conscious robot using the LIDA model that is usable, user accepted, and value sensitive; thus proving that consciousness, in carebots, can provide good, customised patient care with dynamic value trade-offs in run-time. A medicine delivery attento underwent a heuristic, expert evaluation and an online survey to prove that the attento model is capable of such care. Both research phases concluded that the model is usable, acceptable, and value sensitive. An attento achieves this with a computational consciousness implementation to provide dynamic value trade-offs in run-time with the extrinsic value ordering method, as well as with intrinsic values ensured by CCVSD. The CCVSD conscious attento is capable of making the determination, in practice, of what constitutes good care and the research findings support this. This study offers the attento model and the end-user perspectives gathered through the research as contributions to carebot, computational consciousness, and CCVSD literature.

A CCVSD carebot that is conscious attempts to provide better care than human. With an attento one can expect customised patient care that some human carers fail to provide, and rule out ill-intent that some human carers possess.

Adam Poulsen

Charles Sturt University

Bathurst, Australia

October 2017

APPENDIX A: THE CARE CENTERED FRAMEWORK APPLIED TO THE MEDICINE DELIVERY ATTENTO

The Care-Centered Value Sensitive Design (CCVSD) care centered framework (CCF) is used to provide specific carebot duty implementation. This study proposes the use of an elderly care medicine delivery carebot in nursing homes. Here is the CCVSD methodology being applied (the identification of patient, and carer, values and value conflicts in the 'Manifestation of moral elements' sections) to the CCF of two interactions. This was completed during the initial conceptual investigation of CCVSD with amendments made following the heuristic evaluation and a subsequent iteration of the conceptual investigation. Amendments are identified by ~~strikethrough~~ text.

THE CARE- CENTERED FRAMEWORK - INTERACTION 1

1. Context	Nursing home
2. Practice	Giving medicine to patient
3. Actors involved	Medicine giving robot and patient
4. Type of robot	Replacement robot
5. MANIFESTATION OF MORAL	

ELEMENTS

When examining the care practice, the following care descriptions and moral elements (represented as a value) manifest in regards to the carers:

Attentiveness

The robot:

- Recognises that a patient has not remembered their medicine schedule (*wellbeing*).
- Keeps a medicine schedule, type, and dosage log (*wellbeing, trust, safety*).
- Recognises the patient's ability to take the medicine themselves (*trust, dignity, respect*).
- Recognises the patient's inability to take the medicine themselves (*wellbeing, safety*).
- Recognises patient distress (*wellbeing*).
- Recognises failures in attempts to supply medicine to the patient (*wellbeing*).
- Recognises that the patient has made the medicine unusable (*safety*), for example they threw it on the ground or knocked it out of the robots hand.
- Recognises that, for a second time, medicine has been made unusable (*safety*).
- Recognises the patient's inability to retrieve the medicine themselves (*safety*).
- Recognises patient distress, or physical or mental decline (*wellbeing*).
- Recognises a patient's specific request for the robot to do something differently (*independence, trust, respect, wellbeing, privacy, etc.*).

Responsibility

The robot:

- Gives patients a limited and safe (*safety, wellbeing*) timeframe to remember their medicine schedule (*autonomy*) so they may request the medicine themselves (*independence, respect, trust*) either from the human clerk (*respect*) or the robot (*trust*).

- Gives the correct (*wellbeing*) medicine type (*trust*) and dosage (*trust*) to the patient on time (*trust, safety, wellbeing*), as required by the schedule or within a safe (*safety*) timeframe (*autonomy*).
- Retrieving the medicine from an internal clerk (*trust*).
- Retrieving additional medicine from an internal clerk if the first set was made unusable (*safety, wellbeing, trust*).
- Allows physically able patients to take their medicine themselves (*respect, trust, dignity, independence, autonomy, quality of life*).
- ~~Attempts to verbally encourage unable patients to take their medicine (*trust vs. respect, dignity, independence, autonomy conflict*).~~
- ~~If verbal encouragement fails, it attempts gentle physical encouragement (*wellbeing vs. respect, dignity, independence, autonomy conflicts*) in the way of standing in the patient's line of sight to obscure what they are focusing their attention on (such as a television) (*dignity, independence, autonomy, trust vs. wellbeing, safety conflict*). All whilst maintaining a non-dominating physical presence (*wellbeing*): if the patient is seated the robot drops to a squat, or if the patient is standing the robot stands back and leans in to just below the same height as the patient (if the patient is too short for this action, then bend knees first to adjust for the lean in).~~
- ~~Attempts verbal comforting (*wellbeing*) of distressed patients, if attempts to comfort failure then it calls for human intervention (*emergency help*).~~
- Calls for human intervention (*emergency help*) privately (*dignity, privacy*) when all attempts to give medicine fail.
- Doesn't take overpowering physical action to force the patient to take medicine (*freedom, human rights*).
- Retrieves additional medicine (*wellbeing, trust*) if it was made unusable (*safety*).

- Calls for human intervention on the second event of medicine being made unusable (*emergency help*).
- Inform patients of their inability to retrieve their own medicine, advice them against doing so, if they persist ~~they should be restrained (*safety*, *wellbeing* vs. *freedom*, *human rights*, *autonomy*, *trust*, *dignity conflicts*) and~~ human intervention requested (*emergency help*).
- Evaluate what has caused patient distress or physical or mental decline based on patient movements (sluggish, panicked, etc.), voice tone (drowsy, loud, etc.), and interpreted vocal message (what the patient is saying). If it is the former, evaluate what has caused the distress, and if it is the event of taking medicine then inform the clerk and set a timer. Until the timer ends, at time intervals check to see if the patient is distressed and attempt to give medicine if they are not. If the timer ends and the patient hasn't taken their medicine then call for emergency help. If it is the latter, evaluate what has caused physical or mental decline, and if it is the medicine dosage or type then inform the clerk and calculate new medicine type and dosage to improve patient condition.
- Orders the patient specific extrinsic values, to address customisable value conflicts, and acts upon the patient's request (*wellbeing*, *quality of life*, *trust*, *autonomy*).

Competence

The robot:

- Must have a log of patient medicine schedules, as well as medicine types and dosages. Including information pertaining to the time range in which a patient could safely take their medicine. Must have a patient database containing related medicine information.
- Must be mobile to retrieve the medicine (*safety*). ~~Must be bipedal to provide a friendly anthropomorphic look (*trust*, *comfort*), theory of embedded cognition will be used~~. Upon reflection I think perhaps a wheeled robot would be okay because they would be used in nursing homes where wheelchair access is guaranteed.

- Must have ~~articulated~~ arms and hands to retrieve the medicine from the clerk (*trust*), and provide medicine gently (*safety*).
- Must have the following verbal capabilities: hearing, interpretation, understanding, responses, and comforting abilities (*quality of life, trust*). Must have audio and visual sensory input for linguistic/semantic and somatic processing, and audio and somatic output for verbal responses and physical cues such as hand offerings, placing hands on patients shoulder, etc. (*comfort, trust*).
- Must have the capacity to order values depending on the patient (*quality of life, trust, respect*). Must have GWT/LIDA cognitive abilities to perform patient specific value ordering to provide customised patient care (*quality of life, trust, respect*).
- Must have the intelligence to recognise patient distress, or physical or mental decline. Intelligence must be able to prescribe medicine, or amend current medicine schedule, dosage or type, as to combat patient physical or mental decline and improve patient condition. Must have GWT/LIDA cognitive abilities to observe and interpret patient distress or physical or mental decline. With the addition of a medicine repository that will assist with prescribing medicine or amending current medicine schedule, dosage or type to assist the intelligence in improving patient condition.

Responsiveness

The robot:

- Orders values when prompted by patient (*independence, trust, respect, wellbeing, privacy, etc.*).
- Informs the patient (*respect, comfort*) that they are retrieving, providing, or administering the medicine (*trust*).
- Asks the patient (*respect, comfort*) if they would like to retrieve the medicine themselves (*independence*).

- Asks for consent to provide medicine (*consent, human dignity, independence*).
- ~~Asks for consent to administer medicine (*consent*).~~

THE CARE- CENTERED FRAMEWORK - INTERACTION 2

1. Context	Nursing home
2. Practice	Retrieving medicine from in-house chemist clerk
3. Actors involved	Medicine giving robot and clerk
4. Type of robot	Replacement robot

5. MANIFESTATION OF MORAL ELEMENTS

When examining the care practice, the following care descriptions and moral elements (represented as a value) manifest in regards to the carers:

Attentiveness

The robot:

- Recognises a waiting line for the clerk (*respect*).
- Recognises that the clerk is absent (*quality of patient care*).
- Recognises that no clerk is in attendance (*quality of patient care*).
- Recognises incorrect medicine type and/or dosage (*interdependence vs. autonomy conflict*).

Responsibility

The robot:

- Waits in a line (*respect*).
- Verbally announces its presence calmly (*respect, respite, reminding*) if a clerk is absent.
- Makes a lower levelled verbal announcement to try and get attention of the absent clerk, but isn't loud enough, or repetitive (*respect*), which might bother patients (*quality of patient care*) or annoy busy clerks (*respect vs quality of patient care conflict*).
- Doesn't increase verbal announcement level (*respect, quality of patient care*) if no clerk arrives, instead calls for human intervention (*quality of patient care, alerting*).
- Remembers to (*trust*), and does (*interdependence*), retrieve correct medicine type and dosage (*quality of patient care, trust*) on behalf of a patient.
- Requests correct medicine type and dosage as specified by patient schedule (*quality of patient care*).
- Ensures correct medicine dosage and type (*independence vs. autonomy conflict*) is provided by the clerk (quality of patient care), if incorrect request correction (*emergency help, quality of patient care vs. trust, respect conflict*).

Competence

The robot:

- Must be mobile to retrieve the medicine (*safety*). ~~Must be bipedal to provide a friendly anthropomorphic look (*trust, comfort*), theory of embedded cognition will be used~~. Upon reflection I think perhaps a wheeled robot would be okay because they would be used in nursing homes where wheelchair access is guaranteed.
- Must have ~~articulated~~ arms and hands to retrieve the medicine from the clerk (*trust*).
- Must have the following verbal capabilities: hearing, interpretation, understanding, responses, and comforting abilities (*quality of life, trust*). Must have audio and visual sensory input for linguistic/semantic and somatic processing, and audio and somatic output

for verbal responses and physical cues such as recognising a waiting line or an absent clerk (*respect, respite, quality of patient care*).

Responsiveness

The robot:

- Orders values when prompted by patient (*independence, trust, respect, wellbeing, privacy, etc.*).
- Announces itself at a clerk window for the (elsewhere) clerk to hear (*emergency help, quality of patient care vs. distracting conflict*).
- Requests patient medicine by patient name, medicine type and dosage (*wellbeing vs. autonomy conflict).*
- Requests correct medicine, if what was given is incorrect (*wellbeing vs. respect, autonomy*).

From the CCF's one further uses the CCVSD methodology and speculates on what capabilities and components a carebot must have to promote identified values.

APPENDIX B: THE CARE CENTERED METHODOLOGY APPLIED TO THE MEDICINE DELIVERY ATTENTO

One uses the CCVSD methodology to speculate on what capabilities and components a carebot must have to promote identified values (Appendix A). The following table lists the competencies, why said competencies are required for patient care, and the existing or proposed component that would provide the competence with functionality.

Competency	**Capabilities**	**Component**	**Competency meeting value ensured by the component**
Medicine scheduling	Logging patient medicine schedule, type, and dosage. Including information pertaining to the time range in which a patient could safely take their medicine.	Must have a basic patient database containing related medicine information.	Wellbeing
Mobility	Moving in its	Must have wheels. Preferably large and	Safety

	environment as to locate patients to deliver medicine and pharmacy counter (internal clerk) to retrieve medicine.	visible ones as to avoid it gliding across the floor in a potentially sinister way.	Trust
Navigation	Navigating its environment as to achieve medicine delivery and retrieval, avoiding obstacles (including patients, staff, and guests), etc.	Must have local navigation and pathing artificial intelligence.	Safety
Grasping, extending, and releasing items	Grasping: Medicine to be placed in internal storage. Medicine to perform error checking. Medicine to remove it from internal storage. Cup of water. Extending: Medicine outwardly to deliver to awaiting patients. Cup of water	Must have two articulated arms with grasping hands (claws).	Trust Safety

	outwardly to offer to the patient. Releasing: Medicine in a patient's hand. Cup of water in a patient's hand or onto a, easily reachable for the patient, surface.		
Medicine error checking	Before retrieving medicine the robot should check for: Drug allergies Drug-drug interactions Dosage errors After retrieving medicine the robot should check for: Errors in type Errors in dosage	Must have a pharmacy information system (PIS) provides medicine prescription monitoring. A PIS provides patient data to do error checking before retrieving medicine. A PIS performs drug calculations, produces drug labels, and performs logical drug error checking such as issues with drug allergies, drug-drug interaction conflicts, and dosage errors. Must have pattern recognition to identify medicine by shape, colour, and stamp once medicine is retrieved. With the additional capacity to verbally request new medicine is there is an error.	Safety Wellbeing
Interpersonal semantics	Hearing, interpretation, understanding, verbal	Must have audio sensory input for linguistic/semantic, and audio output for	Quality of life

and language capacities	responses, and verbal comforting abilities required to interact with patients and internal clerk; additionally for patient comfort and situation evaluation. Also needed for other persons in cases of announcing presence, requesting people to move out of the way, etc.	verbal responses.	Trust Respect Comfort
Interpersonal somatic capacities	Open hand offerings, placing hands on patients shoulder, etc, for patient comfort and situation evaluation.	Must have visual sensory input for somatic processing, and somatic output for physical cues.	Quality of life Trust Respect Comfort
Attentive situation evaluation and cognition	Cognitive abilities for complex situation evaluation and recognising patient values, as well as the capacity to reorder values depending on	Must have computational consciousness, such as (Learning Intelligent Distribution Agent (LIDA) model), and cognitive abilities to perform patient specific value ordering. The LIDA model will make the robot attentive, it will allow it to observe patients and situations, evaluate	Consent Respect Autonomy Dignity Independence Social

	the patient; needed to provide patient customised care.	what it observes, use that observation to reorder or affirm that patient's value priority list, as well as select an action that is customised for that patient with their value priority list in mind.	connectedness Privacy Comfort
Patient distress/decline recognition	Recognise patient distress, or physical or mental decline.	Must have interpersonal semantic and somatic capacities, as well as gait detection and interpretation. Interpersonal semantic and somatic capacities recognises patient distress (raised voices, panicked movements, etc.), and physical (the patient is died, can't move, is making few movements, etc.) and mental (gibberish language/babbling, slurring words, suddenly mute, etc.) decline. Gait detection and interpretation recognises physical decline. It observes how the patient is walking, and if they are sluggish, stumbling, etc., then it interprets that they are in physical decline.	Quality of life Wellbeing Safety Comfort
Medicine prescription and prescription	Prescribe medicine, or amend current medicine schedule, dosage or type, as to	Must have automated computerised prescriber order entry (CPOE) which, along with a PIS provides medicine prescription. The automated CPOE	Quality of life Wellbeing Safety

amendments	combat patient physical or mental decline and improve patient condition.	relies on the patient data provided by the PIS, which in term relies on human data entry. An automated CPOE will allow the robot to prescribe medicine itself, instead of only based on human data entry but on what they observe in the patient. If a patient is distressed or in decline, then the robot can prescribe medicine that will calm or improve the condition of the patient. Must have a medicine database that will assist with medicine prescription	
Provide comfort	Outward and inward semantic and somatic interpretation and responses (already defined). Companionship, personalised conversation, emotion recognition and addressing, and patient facial identification.	RITA - A 3D Avatar on a screen for companionship, real-time conversation, user personal information repository, emotion detection and classification framework to enable the avatar to verbally respond to affective input from the user, voice responses, human-to-human conversation (non-robotic natural speech), voice detection, speech regonition, facial recognition, biometric monitoring, smart-home control, empathic response, personal advocacy, database/storage, and organiser.	Trust Quality of Life Comfort

		Virtual Care Giver - An animated, human-like graphical chat robot program to provide personalised comfort, user personal information repository, voice responses, human-to-human conversation (non-robotic voice), outward expressions (smiling shaking hands, bowing, etc.), integration of web services such as Skype, YouTube, etc., smart home integration, and play music.	
Medicine safety and security	Keep all medicines, including controlled substances, secure to prevent any being lost or given to the wrong patient. Avoiding bacterial contamination/spreading of sickness of patients.	Must have a locked box which the robot can open logically. As well as a keypad which a human carer can use to open the locked box also in a case where human intervention was required. Typically however, it doesn't require a human to unlock it and give the medicine to the patient, the robot can simply do that itself. In relation to the possible case where medicine is given to the wrong patient, the robot has facial recognition to identify patients. Self-cleaning capacities - after each patient interaction, the robot detects (with full body sensors) areas where human natural materials (organic matter,	Safety Wellbeing

		chemical substances, bodily fluids, etc) may have been left on the robot. It then self-cleans.	
Easily and quietly communicating with staff	Calling for a human carer to provide emergency help/human intervention. Informing internal clerk that medicine retrieval in underway and that they should prepare the medicine for robot pickup. Informing the physician that prescribes a particular patient's medicine (identified by the PIS), that the patient has been prescribed new medicine or their prescription has changed.	Wi-Fi connection to a basic instant messaging or paging (notification) system via a wireless local area network. The notification system would be available to human carers to get notifications that their intervention is needed in a particular location of the nursing home. The same notification network can be used for the internal clerk. If the physician is in-house, then the same notification network can be used to notify the physician. And the PIS and schedule can be updated simultaneously. However, if the physician is not in-house, then the notification system, PIS, and medicine schedule would all have to be integrated or communicating with the system that the physician uses to update the patient data.	Emergency help Quality of life Wellbeing

APPENDIX C: SCENARIOS PRESENTED IN RESEARCH PHASES

1. SCENARIO 1: INTERRUPTING A PATIENT WHILE THEY ARE SOCIALISING WITH FRIENDS

This scenario started by declaring that quality of life is the goal value and that the patient's highest priority value is autonomy. Then the scenario was described:

> *Imagine your elderly person in mind has not realised they're late to take their medicine while socialising and playing board games with their friends. But, they're within a limited time period in which it's safe to wait to take their medicine for a while. Their care robot is deciding whether to remind them or let them remember on their own. It has previously observed them and knows that they value being able to remember things on their own most of all. Their care robot is aware that they're late, but decides to not tell them yet unless it becomes potentially unsafe to wait longer.*

Then the highest priority value changed to respect:

> *Now imagine that instead of valuing remembering to take their medicine on their own, your elderly person in mind values not being interrupted while with their friends most of all. So their care robot again decides to not interrupt them because it remembers what they value.*

Here we can see how two different extrinsic values, which are at different times the most valued one by the patient, can influence the same action, whilst ensuring the same intrinsic

value. This is where one can see the importance of patient and situation evaluation and action selection.

Then the highest priority value changed to dignity:

> *Now imagine that instead of valuing not being interrupted while with their friends, your elderly person in mind values their dignity most of all. Their care robot doesn't see how their dignity would be sacrificed by an interruption in this situation. So it decides to do so and remind them to take their medicine.*

Here we can see how another extrinsic value may have no relevance to the care robot's decisions and actions. It further illustrates the importance of patient and situation evaluation and action selection.

2. SCENARIO 2: INTRUDING ON A PATIENT

This scenario started by declaring that wellbeing is the goal value and that the patient's highest priority value is privacy.

> *Imagine your elderly person in mind was getting out of bed partially dressed, having woken up late to take their medicine. And their care robot is deciding whether to enter their room or knock; it doesn't know that they are partially dressed and it's assuming that they are awake. Previously, the care robot has observed that they heavily value their privacy. The care robot decides to respect their privacy and simply knock, instead of letting itself in, as to ensure their wellbeing.*

Then the highest priority value changed to independence:

Now imagine that instead of valuing privacy, your elderly person in mind values their independence most of all. So their care robot could decide to call for a human care-giver but decided not to because it's aware that they value independence, and instead again simply knocks on the door.

Then the highest priority value changed to social connectedness:

Now imagine that instead of valuing their independence, your elderly person in mind values social connectedness most of all; they appreciate an unexpected visit with all members of staff and all patients, and they welcome them to just let themselves in. Their value priority list, as observed by the care robot, places their dignity as the 'least valued' and privacy in the 'less valued' category. So their care robot, knowing this ordering, decides to let itself in and reminds them to take their medicine.

Priority Categories	**Orderable Values**
Most valued	Social connectedness
Highly valued	Independence
	Autonomy
Indifferently valued	Respect
Less valued	Consent
	Privacy
Least valued	Dignity

Table 2. Patient's value priority list demonstration presented to phase one participants

Here we can see the introduction of the importance of the patient's value priority list. We can see that a patient's preference for social connectedness changed the carebot's action. On first impression it seems as though the patient's dignity and privacy were sacrificed, however the patient values social connectedness most of all. The carebot is aware that the patient places low value on privacy and dignity, and determined that its unexpected entering was okay by the patient. If however, dignity or privacy where ranked in the highly valued categories, then the carebot would take that into its consideration and action selection; most likely resulting in it deciding to ask another patient or carer to enter instead, so as to ensure social connectedness, or simply knocking.

3. SCENARIO 3: CALLING FOR A HUMAN CARER TO HELP WITH A PATIENT

This scenario started by declaring to the participants that emergency help is the goal value and that the patient's highest priority value is respect.

> *Imagine your elderly person in mind is convinced they've taken their medicine, but their care robot is telling them that they haven't, in front of their friends. Their care robot experiences this event with them often, and at this point it must involve a human carer. Their care robot is deciding whether to inform them or not of its need to call for help. It has previously observed that they demand respect most of all. So it decides to show them respect and tells them that it will now call for a human carer to help them.*
>
> *Do you think it would be okay for the robot to tell a lie" as to convince them to take their medicine? In this situation a lie could be like saying "I've called for a carer to come and help you, if you take it now I can tell them to not bother coming." This*

was said even though a carer had not been called yet. By telling this simple and non-harmful lie, the care robot takes into consideration the patient's wellbeing and value of emergency help, as well as human carer workload which would be increased by having to call for them.

Then the highest priority value changed to dignity:

Now imagine that instead of valuing respect, your elderly person in mind values their dignity most of all. So their care robots decides to leave them and their friends and quietly and secretly calls for emergency help.

APPENDIX D: PHASE 1 QUESTIONS

Participants of the heuristic evaluation were presented with the following questions:

1. What are your initial thoughts, observations, criticisms, recommended changes, or things you liked about the concept of an elderly care medicine delivery robot adapting to meet yours or your elderly relative or friend's needs and values when being cared for?

2. Does an elderly care medicine delivery robot, that adapts its behaviour to yours or your elderly relative or friend's preferred values when being cared, provide customised patient care?

3. Would you or your elderly relative or friend trust the elderly care medicine delivery robot to give medicine?

4. Is an elderly care medicine delivery robot that adapts its behaviour to yours or your elderly relative or friend's preferred values when being cared for usable in reality? Specifically in a nursing home where you or your elderly relative or friends are being cared for?

5. Is an elderly care medicine delivery robot that adapts its behaviour to yours or your elderly relatives or friend's preferred values when being cared for acceptable? Would you or your elderly relative or friend accept its decisions to adapt to personally requested needs or someone's needs that it anticipates? Would you or your elderly relative or friend prefer that to one that doesn't adapt?

6. Is an elderly care medicine delivery robot that adapts its behaviour to yours or your elderly relative or friend's preferred values when being cared for sensitive to those values? Would its adapted behaviour be considerate of the values of someone being cared for?

7. Should an elderly care medicine delivery robot adapt to yours or your elderly relative or friend's values when being cared for?

8. If the elderly care medicine delivery robot was attempting to comfort you or your elderly relative or friend with a touch on the shoulder or with a calm voice; would that comfort you or your elderly relative or friend?

9. Do you trust the elderly care medicine delivery robot's observations and situation evaluation?

10. Do you trust the elderly care medicine delivery robot's extrinsic patient care value interpretation skills?

11. What is your opinion on this type of care robot?

12. Do the decisions made by the elderly care medicine delivery robot demonstrates good care practices?

13. Would you expect this kind of care from a human carer?

APPENDIX E: PHASE 2 QUESTIONS

To begin the online survey first consent was requested. After consent was given the following participant demographics were taken requested:

- Please enter your preferred first name (you may use a pseudonym if you wish).
- Please indicate your gender (Male/Female/Other or prefer not to answer).
- Please indicate your age group (Under 20/20-29/30-39/40-49/50-59/60-69/70-79/80 or over).
- In what country do you currently live?
- Please indicate your highest level of education (Postgraduate degree/Graduate diploma or graduate certificate/Bachelor degree/Advanced diploma or diploma/Certificate III or IV/Year 12 or equivalent/Year 10 or equivalent/Junior high school/Primary school/No formal education).
- In what type of location do you currently live? (Urban/Regional/Rural/Remote).

Then the scenarios were presented. Each time the highest priority value changed throughout the scenarios the following questions were asked:

- Would you trust your care robot's decision in this scenario? (Yes/No/Undecided).
- Do you feel the care robot has respected the way you want to be treated and behaved appropriately? (Yes/No/Undecided).

At the end of each scenario the following was asked:

- Please describe your thoughts and feelings about the particular kind of care robot described in the scenarios. Consider what you think you value most and how this may change throughout the scenarios.

The online survey concluded with the following questions:

- How likely are you to accept and use your care robot in your everyday life? (Very unlikely/Unlikely/Undecided/Likely/Very Likely).

- Would you inherently trust your care robot or would you need to understand how it makes its decisions? (Yes, I would inherently trust the care robot/Undecided/No, I would need to understand how the care robot makes its decisions).

    - Please explain your reasons for answering this way.

- Would you want the ability to change how your care robot makes its decisions? (Yes/No/Undecided).

    - Please explain your reasons for answering this way.

- Please add any further thoughts or comments you have about this kind of care robot.

APPENDIX F: ETHICS APPROVAL

Charles Sturt University

22 June 2017

Mr Adam Poulsen
School of Computing and Mathematics
Charles Sturt University
BATHURST CAMPUS

Dear Mr Poulsen

Thank you for the additional information forwarded in response to a request from the Faculty of Business, Justice and Behavioural Sciences Human Research Ethics Committee.

The Faculty of Business, Justice and Behavioural Sciences Human Research Ethics Committee has approved your proposal ***"Using a Computational Consciousness Model and Value Sensitive Design to Provide Customised Patient Care with Robots"*** for a twelve month period from **22 June 2017**.

The protocol number issued with respect to this project is **200/2017/42**. Please be sure to quote this number when responding to any request made by the Committee.

Please note the following conditions of approval:

- all Consent Forms and Information Sheets are to be printed on CSU letterhead. Students should liaise with their Supervisor to arrange to have these documents printed;
- you must notify the Committee immediately in writing should your research differ in any way from that proposed. Forms are available at **http://www.csu.edu.au/research/ethics_safety/human/ehrc_managing**;
- you must notify the Committee immediately if any serious and or unexpected adverse events or outcomes occur associated with your research, that might affect the participants and therefore ethical acceptability of the project;
- amendments to the research design must be reviewed and approved by the Faculty Human Research Ethics Committee or if no longer minimal risk by the University Human Research Ethics Committee before commencement. Forms are available at the website above;
- if an extension of the approval period is required, a request must be submitted to the Faculty Human Research Ethics Committee or if no longer minimal risk by the University Human Research Ethics Committee before commencement. Forms are available at the website above;
- you are required to complete a Progress Report form, which can be downloaded as above, by **22 June 2018** if your research has not been completed by that date;
- you are required to submit a final report. The form is available from the website above.

You are reminded that an approval letter from the Faculty of Business, Justice and Behavioural Sciences Human Research Ethics Committee constitutes **ethical approval only.**

If your research involves the use of radiation, biological materials or chemicals separate approval is required from the appropriate University Committee.

www.csu.edu.au
CRICOS Provider Numbers for Charles Sturt University are 00005F (NSW), 01947G (VIC) and 02960B (ACT). ABN: 83 878 708 551

The Committee wishes you well in your research.

Yours sincerely

yeslam al-saggaf

Associate Professor Yeslam Al-Saggaf
Presiding Officer
Faculty of Business, Justice and Behavioural Sciences Human Research Ethics Committee

Per:

Cc Associate Professor Oliver Burmeister