A Beginner's Guide to Research

David Vernon

Carnegie Mellon University Africa

www.vernon.eu

www.vernon.eu/teaching.htm#research

A Beginner’s Guide to Research

David Vernon

Carnegie Mellon University Africa

www.vernon.eu

Revision 7.1

January 9, 2024

Table of Contents

1. Getting Started

Research is Difficult

Writing a thesis is hard work and getting started is perhaps the most difficult part. A research degree focuses on you becoming an expert in a particular topic and adding to the body of scientific and engineering knowledge on that topic. However, research is also concerned with learning, and especially with learning to work independently and being able to develop your own understanding of any given topic. Starting with very little, where do you begin? The following are some pointers on how to get started.

Understand the Problem

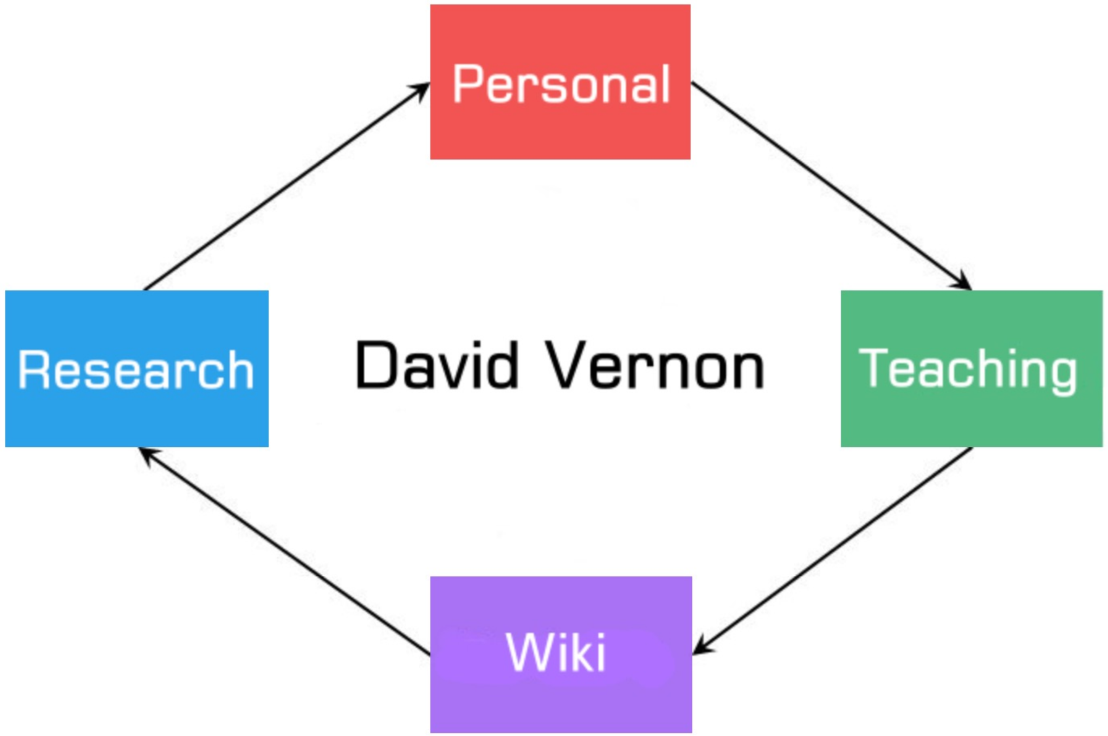
Your supervisor will have provided you with an outline description of your research topic. To begin with, and to test your understanding of this outline, you should try to expand on the problem statement. In your own words, try describing informally exactly what the final system should do, what data it will take as input, what data it will produce as output, and how these input and output processes are accomplished. Next, try to say exactly how the input data is transformed in order to produce the required output. Describe why this problem is relevant. Say why it is important to solve it. What are the consequences of finding a solution? What does solving this problem mean?

Start Reading

Your supervisor will suggest some initial reading: journal papers, conference papers, book chapters. Read them all.

Start Writing

Arm yourself with a pen and write a short synopsis of every paper and book chapter you read. Write down the key message (one or two sentences) and the main contribution (one or two paragraphs). It is also worth writing down one or two quotes from the paper if they provide some important insight into the topic. This is hard work. Don’t underestimate it but do be aware of how important it is to do it. The very act of writing helps crystallize ideas. Remember: summarize every paper you read. Don’t be tempted to copy the paper abstract: the point of the exercise is for you to express your understanding of the paper in your own words. This is a necessary part of the learning experience. You won’t learn if you don’t write it down in your own words.



david@vernon.eu
www.vernon.eu

Teaching Activities

If you are a teacher and would like a Powerpoint version of any of these courses, please contact me.

A Beginner's Guide to Research

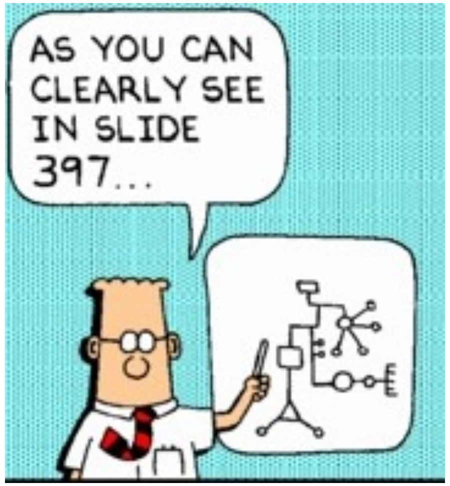
Getting Started
Moving Along
Reading
Writing: Good Writing Discipline, Good Writing Style
Writing a Research Proposal
Writing a Literature Survey
Writing Scientific Papers
Ph.D. and M.Sc. Dissertations
Looking Forward
Acknowledgements
Further Reading
Appendix I Some Rules of English Usage
Appendix II Writing a Literature Survey
Appendix III Typical Structure of a Thesis

Booklet, 660 kb: ; Slides, 66 Mb:

“Extreme Tutorial”

http://www.diffily.com/articles/digital-governance-management-tools.htm

HEALTH WARNING



HEALTH WARNING

Getting Started

Research is Difficult

- Become an expert in a particular topic
- Contribute to the body of scientific and engineering knowledge
- Learn to work independently

Understand the Problem

- In your own words, try describing informally exactly what the final system should do
- Inputs, outputs, transformations
- Why this problem is relevant?

Start Reading

- Initial reading
- Journal papers
- Conference papers
- Book chapters

Start WRITING

- Write a short synopsis of every paper and book chapter you read
  - Key message (one or two sentences)
  - Main contribution (one or two paragraphs)
  - One or two quotes from the paper
  - Assumptions, principles, techniques (these are organizing attributes)
- Writing helps crystallize ideas
- Don't copy the paper abstract
- You won't learn if you don't write it down in your own words

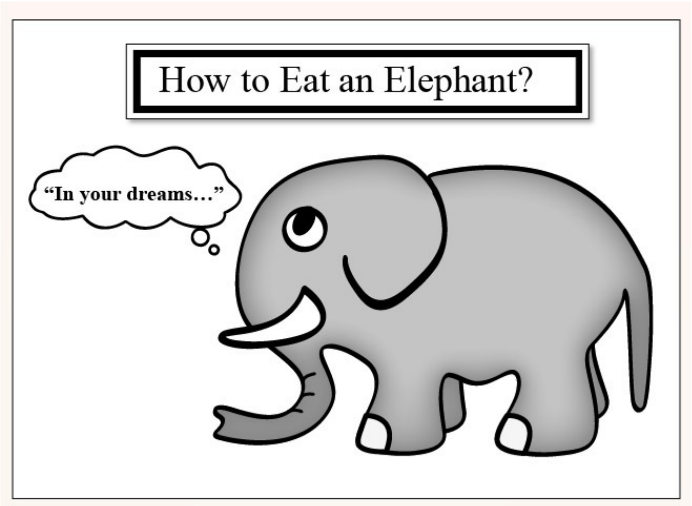
Start Writing

- Talk to your supervisor & colleagues
  - The goal of your work
  - The central problem
  - How you are going to solve it
  - How other people have solved it
  - What difficulties they encountered
  - Why their approach isn't as good as the one you are trying to use
- Write down these ideas
  - Use a pen and paper first
  - Type the notes later
  - Give a short presentation to your group

Start Listening

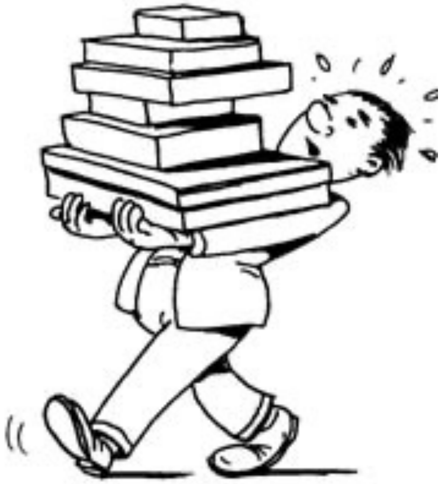
- Go to seminars
- Go to talks being given by other students
- Go to all group meetings
- Research is a social activity

Start Simple



- Carve up the problem
- Solve a simple version first
- Solve the complicated version next
- Structure your research goals
  - an essential but fairly easily achieved goal
  - a desirable but somewhat harder or risky goal
  - the ideal goal, something that would represent a real breakthrough

Work Hard



- Genius is 1% inspiration and 99% perspiration

- You can't get a Ph.D. or M.Sc. without a lot of hard work

  - early mornings
  - late nights
  - frustration
  - fatigue
  - & sometimes depression

Formalize

$$V(s) = K(s)\left(\frac{R(s)}{1+\alpha R(s)}\right)^{1/3}$$

vs.

$$C = {}^1\!/_2 \int_{t_1}^{t_2} \left[\left(\frac{d^3x}{dt^3}\right)^2 + \left(\frac{d^3y}{dt^3}\right)^2\right] dt$$

- As early as possible, you need to try to formalize the problem you are working on
- You will need to get a grounding in the background theory first

(One third power law vs. minimum jerk law for modelling biological motion)

Learn about Tools

You need to become and expert in two different domains

- The problem domain

  Theoretical issues by which we can model and solve the problem

- The solution domain

- The tools that will enable you to implement the solution

Moving Along

Make Notes

Keep a logbook of all work in progress

- Thoughts
- Ideas
- Notes of meetings with your supervisor
- Results
- Theoretical developments
- Calculations
- References
- Anything that is relevant to your work

Believe in Your Own Ability

- You will have to spend many hours every day, often with no reward, chipping away at the problem, hoping for progress

- To get through this, you need to believe in your own ability

Believe in Your Problem

- Believe that the topic you are working on is worthwhile
- If you are not convinced that solving this problem matters, find another topic
- You should enjoy your PhD / MSc!

Practice, Practice, Practice

Research is not a spectator event:

it is a contact sport!

Framing the Research Question

- Re-scope, if necessary
- PhD: ask and answer a hard question
- If the answer is not convincing, get a better answer or a different question

Jump Start Your Day

Before your leave work,
set yourself a task for the morning
that is easy (and productive)

Dealing With Criticism

"Remember, though, if we can't criticize ourselves, someone else will save us the trouble."

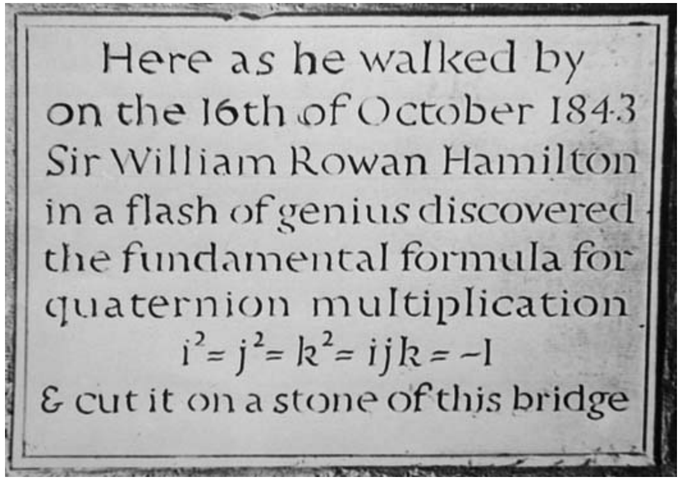
Drew McDermott, Artificial Intelligence Meets Natural Stupidity, ACM SIGART Newsletter, No 57, pp. 4-9, April 1976.

- You will have to take a lot of criticism
- Being wrong and being ignorant is normal
- Ignorant does not mean stupid.
  It means lacking knowledge
- Criticism is a positive act,
  not a negative one

Don’t Give Up!

- Take mini-holidays (10 minute ones!)
- Don’t think about the problem all the time!!!

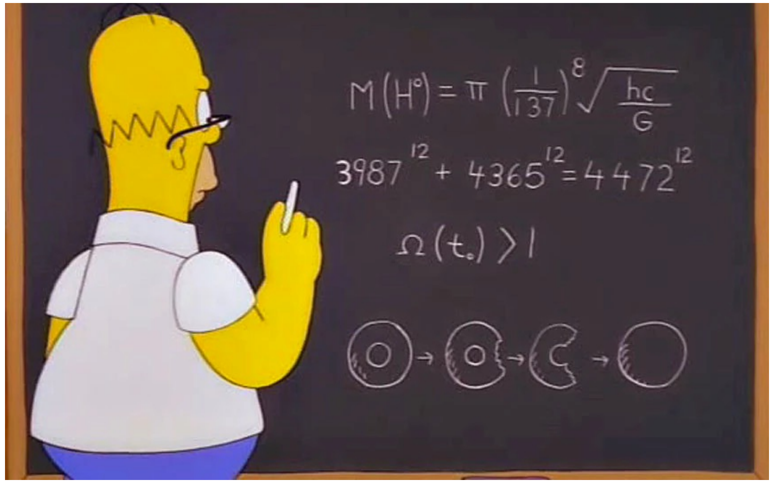
Be Prepared for Inspiration



- Inspiration strikes at strange times!
- Keep pen and paper handy,
  just in case!

  (or a pen-knife)

Get Yourself a Theory



- The great power of science and engineering is that it allows us to predict how systems will behave
- To be able to predict something, though, you must have a model: an abstract formulation.

Get Yourself a Benchmark

- The hallmark of good engineering is to
  - assess the system's performance
  - compare it to that of other similar systems
- Ideally, you should identify some quantitative metric by which to compare the systems

Reading

Recursive Reading

- Read the initial set of papers
- They will refer to other papers
- Get the main (relevant) papers cited and read these
- These refer to other papers ...
- Recurse until you achieve closure

Three Levels of Reading

- Shallow Reading
- Focussed Reading
- Deep Reading

Shallow Reading

- Mainly background material providing the context to the research
- The topics are relevant to your work but not directly related
- Often it is sufficient to read just the abstract, introduction, and conclusion
- Write a one- or two-line summary of the main issue addressed.

Focussed Reading

- Directly relevant and provide, for example,
  - alternative ideas on the topic of your research
  - Component techniques
- These papers should be read thoroughly, perhaps twice
- Write a one-paragraph summary
- These papers will typically go in your literature survey

Deep Reading

- Some papers (10-30) will be absolutely central to your work (e.g., competing models)
- Read them several times to really understand them
- Work through some examples
- Probably have to refer to other papers or text-books to understand some of the concepts described

Deep Reading

- Write a 1-2 page summary
  - If they contain mathematical results, you should include these
  - Explaining each term and the importance of the results
- After many careful readings, you should know as much about the topic as the author
- A good test of your understanding of a paper is to see if you can give a short presentation on it and explain it to other people in your group

Build a Bibliography

- For everything relevant you read, insert the full citation of the paper or book chapter in your bibliography so that you can refer to it in your subsequent writing

- Make sure you keep a complete citation index

  - Title of the article
  - Authors
  - Name of the conference/journal/book
  - Volume and number
  - Page numbers
  - If it's a chapter in a book and the author of the chapter is different from the editor of the book, you need to record both sets of names.

Variety

- The 'reading in' phase of the project can last quite a long time

  (there's a lot of reading and writing to be done)

- It can help to overlap with some of the other early tasks
  - Learning about the solution domain

Reading Means Writing

- To fully understand anything that you read, you must write it up in your own words

- If you can't express or speak about a given idea, then you haven't truly understood it in any useful way.

- Writing is an essential part of understanding

Writing

Clarity, clarity, clarity

Writing

- Writing is not easy

- The discipline of writing

  - the physical act of writing:
  - the process of assembling ideas and getting them down on paper

- The style of writing

  - elegance and simplicity:
  - the power of your writing to communicate an idea

Good Writing Discipline

- Keep records
  - Notes on papers you read
  - Notes on software you develop
  - Notes on ideas you have
  - Notes on tests you run
- Writing these notes serves several purposes
  - It helps to crystallize ideas and clarify them
  - It helps the learning process
  - It makes sure you don't lose or forget any important points.
  - It also acts as a basis for subsequent writing: for reports and papers

Good Writing Discipline

- Make writing a way of life!

- You should allocate a large proportion of your day to writing

- Writing should be an integral part of your working day

Good Writing Discipline

- Use pen and paper
- Write things down long-hand
- Later on, write these notes up more neatly and in a more organized fashion
- Once you get good at this, you can go straight from long-hand notes to typed document
- but it's very helpful at the beginning to first create an intermediate long-hand version.

Good Writing Style

- Effective writing is difficult

  It takes practice and a willingness to revise your work, many times

- Read good writing
  - Several previous dissertations
  - Conference papers
  - Journal papers
  - Magazine articles

Good Writing Style

- The popular scientific press, e.g., Scientific American or New Scientist, employs a particularly simple and effective form of written expression
  - Try to emulate their style
  - For a model of clarity in scientific writing, read https://www.nobelprize.org/nobel_prizes/medicine/laureates/2014/advanced-medicineprize2014.pdf

Good Writing Style

- Why is writing well so difficult?
- The goal of writing is to convey a message to the reader
- Writing and reading are sequential processes
- You have to construct the meaning of your message, in a linear time-line

Good Writing Style

- However, the meaning you intend to convey may emerge from many sources, not all related in a nice orderly fashion

- This creates a problem for the writer:

  how to order the messages contained in each sentence effectively

Good Writing Style

- Use short sentences and make sure the sentences are complete
  - A complete sentence has a subject followed (usually) by a verb, and then an object
- Simple sentence

  "Cognitive systems can adapt to changes in the environment."

Good Writing Style

- Use short sentences and make sure the sentences are complete
  - A complete sentence has a subject followed (usually) by a verb, and then an object
- Richer sentence

  "Cognitive systems can adapt autonomously to unexpected changes in the environment"

.

Good Writing Style

- Use short sentences and make sure the sentences are complete
  - A complete sentence has a subject followed (usually) by a verb, and then an object
- Include subordinate clauses
  (which will normally have a subject-verb-object structure of their own)
  - "Cognitive systems can adapt autonomously to unexpected changes in the environment, especially those that the designer did not anticipate."

Good Writing Style

- Use short sentences and make sure the sentences are complete
  - A complete sentence has a subject followed (usually) by a verb, and then an object
- Include subordinate clauses
  (which will normally have a subject-verb-object structure of their own)
  - "Cognitive systems can adapt autonomously to unexpected changes in the environment, especially those that the designer did not anticipate."

Good Writing Style

- Remember that, if you remove all the extra supporting words, you should be left with a valid sentence.

  It's a good idea to check all your sentences this way

Good Writing Style

- Good writing strikes a balance between short sentences and longer more descriptive ones

- Full stops mean pauses

  - too many pauses and the text sounds disconnected
  - too few and it can be hard to follow the story line

- Strike a balance but favour brevity over complexity

Good Writing Style

Pictures and diagrams

- Make sure each one has a self-contained explanatory caption
- Never refer to a picture or diagram in the main text without saying what it is
- Never say

  "Figure 2.3 shows the results of the noise test"
- Instead, say

  "Figure 2.3 shows the results of the noise test. These results demonstrate the robustness of the system to Gaussian noise with a standard deviation of 1.5 or less."
- If you have copied the figure from a book or article you must cite the source

Good Writing Style

- Make the paragraph your unit of construction
  - Each paragraph should bind one or more sentences about a specific subject or idea
  - If the subject or idea changes, start a new paragraph
- Omit unnecessary words. They distract the reader.
  - Don't write "This is a system the performance of which is very adaptive".
  - Instead, write "This is an adaptive system".

Good Writing Style

- Write in a way that comes naturally
  - Speak the sentence
  - If it sounds correct, trust your ear and use the sentence
  - If it sounds unnatural, rewrite it
- Avoid fancy words; they don't impress anyone

Good Writing Style

- Be clear in your expression
  - Write clearly
  - Write concisely
  - Write precisely
- You have a message to convey
  - make sure you know what it is
  - keep it in focus throughout
  - Don’t stray from the key point you are making

Good Writing Style

- Present your argument in a structured manner
  - Ensure that what comes first depends as little as possible on what comes later
  - Ensure that what comes later builds on and adds to what you have just stated
- If the idea you are trying to convey is getting lost in a sea of words and phrases, draw a line through the sentence and start again.

Good Writing Style

To learn how to write a good dissertation or paper

- Read other good dissertations or papers
- Practise your own writing

Good Writing Style

- Don’t take short-cuts
- Explain what you mean
- Don’t leave the reader to struggle trying to figure out the real meaning of your carefully constructed but complicated sentence

  She may conclude there is none
- Explain all acronyms the first time you use them

Good Writing Style

- Let information flow

  In each sentence, lead your reader from familiar information to new information

- Place material you want to emphasize at the end of the sentence.

Good Writing Style

Be brief

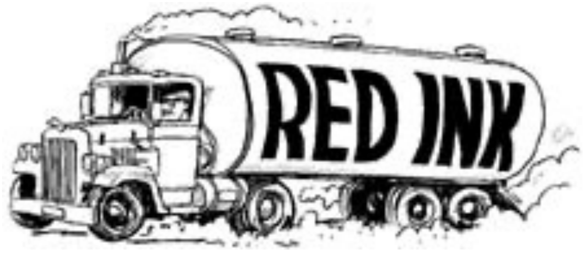
This isn't easy

"I have made this [letter] longer, because I have not had the time to make it shorter"
Blaise Pascal, "Lettres provinciales", Lettre XVI, 1657.

Good Writing Style

Revise and rewrite

Make an attempt, and then be prepared to revise it, repeatedly



Design and Implementation of the Situation Model Framework

David Vernon

Institute for Artificial Intelligence, University of Bremen, Germany.

Abstract

This paper presents the options for the design and implementation of the situation model framework and its integration in the CRAM cognitive architecture. Two options are presented, one symbolic, based on internal simulation using the Unreal Engine, and one sub-symbolic, based on multi-modal associative memories implemented with deep convolutional and recurrent neural networks.

Keywords:

1 The Situation Model Framework

The Situation Model Framework was introduced by Schneider, Albert, and Ritter (2020) as the basis for understanding how cognitive behaviour, in general, and flexible context-sensitive cognitive behaviour, in particular, is realized in humans, animals, and machines. The long-term vision is to explore situation models both at a functional level and at a mechanistic level. The functional level corresponds to the computational model level in Marr's three-level hierarchy of abstraction (Marr & Poggio, 1977; Marr, 1982), often referred to as the Levels of Understanding framework, while the mechanistic level corresponds to the representation and algorithm level and the hardware or wetware level (Poggio, 2012); see Figure 1. Thus, situation models, fully developed, can be viewed at three levels of abstraction: as a computational model, as a representation and process, and, ultimately, as an implementation of flexible context-sensitive cognitive behaviour. The situation model perspective does not yet commit to a specific computational theory, representation, or process. Rather, at present, it is to be viewed as a framework in which to set such a theory, representation, and process. As such, it sets out in explicit terms the assumptions — or foundations — on which to base and build such a theory. Underpinning the situation model framework are three cross-cutting themes. First, action plays a key role in cognition. Control of action is seen as the process that integrates the many different components of an intelligent systems and lies at the heart of flexible context-sensitive cognitive behaviour. Second, complex behaviours can emerge by scaffolding simpler behaviours. Third, perception, memory, and action are tightly linked in the context of a given task by prioritized control mechanisms, i.e. attention.

An important feature of the situation model framework is the ability to create rich situation models on-the-fly. However, a cognitive agent can also envisage a number of alternative situation models because models depend on focus and context. Together with short-term memory, this brings the agent into the high-level situation involving a network of related situation models. Now the situation model framework principles can be applied again at this higher level again, e.g. by simulating outcomes of alternative actions on the models. We argue that The capacity to operate in this generative manner on models, rather than just use them, is the distinctive

An Overview of Different Options for the Implementation of the Situation Model Framework in the CRAM Cognitive Architecture

David Vernon

Institute for Artificial Intelligence,
University of Bremen,
Germany.

Abstract

This report presents the options for the design and implementation of the situation model framework and its integration with the CRAM cognitive architecture. Two options are presented, one symbolic, based on internal simulation using the Unreal Engine, and one sub-symbolic, based on multi-modal associative memories implemented with deep convolutional and recurrent neural networks. The CRAM cognitive architecture provides a computational infrastructure for cognition-enabled robot manipulation in everyday activities. The design, implementation, and integration of the situation model framework will enable more flexible cognitive behaviour by allowing tasks to be planned and executed by exploiting new actions, created on the fly in response to contextual requirements. Two forms of new action are anticipated. One form is based on recombination of perception-action-outcome behavioral episodes, using known actions in new circumstances, as well as generating novel sequences of known actions. The other form is based on generating new motions.

Keywords:

1 The Situation Model Framework

The Situation Model Framework was introduced by Schneider, Albert, and Ritter (2020) as the basis for understanding how cognitive behaviour, in general, and flexible context-sensitive cognitive behaviour, in particular, is realized in humans, animals, and machines. The situation model perspective does not yet commit to a specific computational theory, representation, or process. Rather, at present, it is to be viewed as a framework in which to set such a theory, representation, and process. As such, it sets out in explicit terms the assumptions — or foundations — on which to base and build such a theory. This report documents the first steps to construct and implement such a theory.

Three cross-cutting themes underpin the situation model framework. First, action plays a key role in cognition. Whereas reactions are elicited by earlier events, actions are initiated by a motivated agent, they are defined by goals, and they are guided by prospective information (Vernon, von Hofsten, & Fadiga, 2011; Vernon, 2014). Essentially, actions are organized by goals and not by their trajectories or constituent movement, although of course these are needed to accomplish the action. Control of action is seen as the process that integrates the many different components of an intelligent systems and lies at the heart of flexible context-sensitive cognitive behaviour. Second, complex behaviours can emerge by scaffolding simpler behaviours. Third,

Project Description

1 Background and Objectives

The Success of the Inclusive Digital Transformation of Africa Depends on Acceptance, Trust, and Cultural Sensitivity

This research project is motivated by the recognition that socio-economic development in Africa must be sensitive to people's culture for it to be successful (Olasunkanmi, 2011). Dignum (2023) drives this home when, in the recently published book *Responsible AI in Africa* (Eke et al., 2023), she says "research and development of AI systems must be informed by diversity, in all the meanings of diversity, and obviously including gender, cultural background and ethnicity". While the overarching agenda of the inclusive digital transformation of Africa is widely recognized to have the potential to be a positive disruptive influence many aspects of the lives of African citizens, the transition from recognition of potential to realization of benefits is not a straightforward matter. The transition depends on turning technological invention into innovation, requiring widespread adoption. However, adoption, especially of AI, depends on trust (Alupo, Omeiza, & Vernon, 2022), which, in turn, depends on social and cultural factors (Lee & See, 2004). The successful deployment of social robots in Africa, therefore, depends on the robots being accepted by African citizens. Culturally sensitive robot behavior, the focus of this proposal, is a prerequisite for this.

Social Robotics

There is an increasing need for artificial intelligence technology that is capable of interacting effectively with humans. This includes social robots which serve people in a variety of ways. The global social robotics market was valued at USD 1.98 billion in 2020 and is expected to reach USD 11.24 billion by 2026, registering a compound annual growth rate (CAGR) of 34.34% during the period of 2021-2026 (Research and Markets, 2022).

Social robots are designed to operate in everyday environments, often in open spaces such as hospitals, exhibition centers, and airports, providing assistance to people, typically in the form of advice, guidance, or information. The people interacting with the robot have no special training and they expect the robot to be able to interact with them on *their* terms, not the robot's. There are two aspects to this expectation.

First, it means that social robots need to be able to interpret the intentions of the people with whom they are interacting. This is difficult to achieve because humans do not necessarily articulate their specific needs explicitly when they interact with social robots (or, indeed, with other humans). As Sciutti et al. (2018) note, "the ability of the robot to anticipate human behavior requires a very deep knowledge of the motor and cognitive bases of human-human interaction".

Second, and conversely, humans have expectations of the robot's behavior and, specifically, they expect the robot to act in a trustworthy, culturally-sensitive, socially-acceptable manner, and they have a distinct preference for robots that exhibit legible and predictable behavior (Sciutti et al., 2018). Since people make predictions based on what they are used to, this is somewhat easier to achieve, provided the robot behaviors are tuned to the socio-cultural context in which they are operating.

People use spatial, non-verbal, and verbal communication when interacting with other people. So too must social robots, if they are to be effective. However, successful interaction requires acceptance and trust, which depend on social and cultural norms. These norms impact on the nature of the robot's non-verbal and verbal expression as well as its appearance and spatial behaviour. Consequently, they determine the acceptance of social robots and the effectiveness of their interaction (Bartneck et al., 2020). While the case for culturally competent robots has been well made (Bruno et al., 2017b; Khaliq et al., 2018), and while there are studies on cultural differences in the acceptance of robots in the West and East, e.g., (Kaplan, 2004; Bartneck et al., 2005; Bruno et al., 2017a), similar studies of the cultural factors that impact of acceptance in Africa have not been reported (Bartneck et al., 2020).

Research Objectives

This research proposal is concerned with the second aspect of effective interaction by social robots identified in the previous section: the need for predictable and culturally-acceptable patterns of robot behavior. In other words, the robot must adapt to, or be adapted to the cultural environment. However, rather than attempt to learn these patterns progressively over time through interaction (a research goal that would involve far more effort than is feasible in a project of the size supported by Afretec), we aim to

[1] The recent survey by (Lim et al., 2021) briefly mentions Egypt, Tunisia, Libya, and Sudan but only to contrast perceptions with the Gulf region when interacting with an Arabic robot.

identify these patterns through ethnographic research and then embed them in reconfigurable and reusable interaction primitives that can be utilized when developing the interaction behaviors for the application and environment at hand. Thus, we aim to identify the interaction patterns that are socially and culturally acceptable in Africa, and the specific *traits* in appearance and behavior that will make social robots capable of predictable, effective, and engaging interaction by reflecting the social and cultural norms of African people.

The factors that underpin effective human-robot interaction include spatial interaction (proxemics, localization and navigation, socially appropriate positioning, initiation of interaction, communication of intent), nonverbal interaction (gaze and eye movement, deictic, iconic, symbolic, and beat gesture, mimicry and imitation, touch, posture and movement, and interaction rhythm and timing), and verbal interaction (speech, speech recognition, language understanding, speech generation) (Bartneck et al., 2020). These spatial, non-verbal, and verbal *interaction factors* must be adjusted to reflect the *traits* that would make social robots acceptable in Africa.

These traits will also be used to adjust the eight accepted design patterns for sociality in human-robot interaction (Kahn et al., 2008) so that they reflect social and cultural norms in Africa. These design patterns include the initial introduction, didactic communication, moving in motion together, personal interests and history, recovering from mistakes, reciprocal turn taking, physical intimacy, and claiming unfair treatment or wrongful harms.

Having identified the verbal and non-verbal social and cultural norms of human interaction that are prevalent in different countries in Africa, we will encapsulate them in the behavioral traits of social robots so that these robots engage with African people in a manner that is consistent with their expectations of acceptable — respectful — social interaction, rather than using inappropriate or insensitive social behaviors and modes of interaction from the West or the East.

In pursuing this research, we recognize that there are many different cultures in Africa, with many different norms for deictic, iconic, and symbolic manual gesturing, as well as gestures involving eye gaze, head tilt, eyebrows, and body posture, generally. Similarly, there are many different ways in which spoken language can express nuances of meaning by modulating amplitude and timbre. The outcomes of the research will take the form of a suite of software primitives, integrated in an application programming system architecture, and a set of design patterns that can be recruited during human-robot interaction, deploying the spatial, non-verbal and verbal communication channels that are best suited to the social and cultural needs of the interaction.

The software primitives, system architecture, and design patterns will be evaluated in two complementary use-cases. Each use case will be conducted in two phases, so that evaluation after the first phase can provide feedback and allow the results of the research to be adjusted and improved, if necessary.

Technical Merits & Support for Capacity Building

It is increasingly accepted that AI systems need to understand, and interact in, the social world of humans. This is particularly true in robotics, which is viewed by many as "cognition-enabled transferable embodied AI" (euROBIN, 2023) and, especially, in social robotics. As we noted above, effective interaction is essential for acceptance, trust, and adoption. This implies that social robots must be able to recognize cultural traits in humans, infer their intentions, and behave in a manner that is culturally legible and predictable by adhering to social and cultural norms.

A complete culturally competent robot requires at least five elements: (i) cultural knowledge representation, (ii) culturally sensitive planning and action execution, (iii) culturally aware multimodal human-robot interaction, (iv) culture-aware human emotion recognition, and (v) culture identity assessment, habits, and preferences (Bruno et al., 2017a), as well as intention recognition and some capacity for forming a theory of mind (Vernon, Thill, & Ziemke, 2016). The research project proposed here focusses on the first three of these, i.e., the generation of culturally sensitive robot behavior. This presents a crucial technical challenge which is an essential step in building a research capacity for developing complete culturally competent social robots. While some research programs focus on robots that can learn these behaviors from demonstration, a difficult problem given the nuanced nature of non-verbal communication and the tonal nature of verbal communication[2] which we intend to address in future research, our approach in this project is to catalogue the behaviors based on enthnographic research and embed them in reconfigurable software design patterns. In the short term, as we build the required research capacity, this is a more tractable approach and will produce reusable results, while still being compatible with the goal of developing culturally competent robots by combining top-down and bottom-up approaches based on the predetermined profiles of a cultural group

[2]Watch a video of the Kismet robot saying the same thing in different tones for a convincing demonstration of the nuanced nature of verbal and non-verbal communication: https://robots.ieee.org/robots/kismet/?gallery=video1.

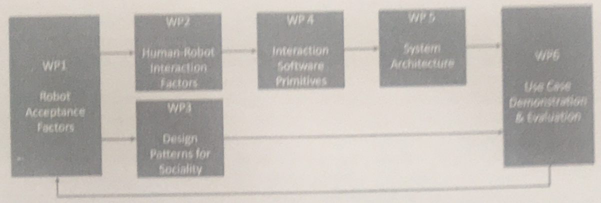


Figure 1: PERT chart showing the dependencies between the technical work packages, and highlighting the iterative development cycle.

Table 1: List of work packages

WP No.	Work Package Title	Lead Partner	Person Months	Start Month	End Month
1	Robot Acceptance Factors	CMU-Africa	5.64	1	27
2	Human-Robot Interaction Factors	CMU-Africa	3.57	1	27
3	Design Patterns for Sociality	Wits	3.57	1	27
4	Interaction Software Primitives	CMU-Africa	13.50	3	30
5	System Architecture	CMU-Africa	7.73	7	33
6	Use Case Demonstration and Evaluation	Wits	6.15	1	36
7	Dissemination and Impact	Wits	2.75	1	36
8	Project Management	CMU-Africa	1.31	1	36

and the cultural profiles derived from the behaviors of individuals, respectively (Khaliq et al., 2018).

Our plan is to embed these culturally-sensitive interaction primitives in a ROS-based application system architecture, the de facto standard for robot middleware, and make them freely available on GitHub. By doing so, we will be creating a reusable resource for all roboticists in Africa. This will make a significant contribution to the capacity of these roboticists to develop their own bespoke culturally-sensitive social robot applications.

Furthermore, these research results will provide a foundation for future research in human-robot interaction, significantly improving the likelihood of success when bidding for new research projects in social robotics and human-robot interaction.[3]

In carrying out the research described in this proposal, we also lay the foundation for future research by exposing students to the development of new techniques, rigorous research methodology, and effective project management. To facilitate student involvement through short-term one-semester assignments, the tasks outlined in the work package descriptions in Section 5 will, where possible, be configured as well-scoped micro projects.

Taken together, this technically-important goal and the impact of its three capacity-building results (freely-available reusable software, enhanced research resources, and trained researchers) contribute to the United Nations Strategic Development Goals 9.b "Support domestic technology development, research and innovation in developing countries", 10.2 "... empower and promote social ... inclusion of all" (United Nations, 2022), and 17.6 "Enhance North-South, South-South, and triangular regional and international cooperation, as well as contributing to the objective of the African Union Digital Transformation Strategy for Africa 2020 – 2030 to "build inclusive digital skills and human capacity ... to lead and power digital transformation including ... robotics" (African Union, 2022).

2 Approach and Activities

We adopt both a user-centric perspective and an agile and iterative approach in this project. This is reflected in the work plan; see the Pert chart in Figure 1 and the list of work packages in Table 1. WP1 is fundamentally user-driven and focusses on identifying the cultural and social norms, and behavioral traits

[3]The CMU-Africa PI is a member of consortia in two Horizon Europe proposals on social robotics and social AI, with specific responsibility for integrating socio-cultural factors in a cognitive architecture, and is a cooperation partner in a proposal for a German excellence cluster on joint action in robotics. The current Afretec project will increase the likelihood of successful involvement on future proposals and projects, helping to increase research capacity even further.

that define respectful, engaging interaction in African countries. It will achieve this through ethnographic user studies that create the data and theory necessary for the development of the HRI factors in WP2, the design patterns for sociality in WP3, the interaction software primitives in WP4, and the system architecture in WP5 that forms the basis for the two demonstration and evaluation use cases in WP6. Monitoring research progress, meanwhile, is also done in WP6 through user studies that test and validate the targeted use case functionality at the end of years 2 and 3 of the project, taking appropriate action after year 2 to adjust and augment each element in WP1 - WP5 in order to improve the performance in the use cases in the subsequent phase. The time line also highlights this iterative development; see the Gantt chart in Figure 2 in Section 5, where detailed work package descriptions are also provided.

For the ethnographic study of the cultural factors that impact on the acceptance of social robots in different countries that will be carried out in WP1, we perform these studies using two independent groups, and cross-validate the results, with one group validating the other group's results, adjusting appropriately, if necessary. In addition, we plan on engaging an external expert in ethnographic research in developing countries to ensure the validity of our approach and adapt it, as required.

In terms of technical development, we plan to adopt the development methodology and outline functional architecture for a culturally competent robot proposed by Bruno et al. (2017b), basing the initial design on ROS4HRI (Mohamed & Lemaignan, 2021) for component interfaces, and the CRAM cognitive architecture (Beetz et al., 2010) for action planning and culture knowledge representation and reasoning, using the culture knowledge ontology proposed by Bruno et al. (2019) as a foundation.

From a software engineering perspective, we will adopt a compositional approach, one that wraps the software functionality developed in WP4 and integrated in WP5, and facilitates the use-case specific configuration of this functionality. Specifically, we will use ROS, a globally-used implementation of component-based software engineering (CBSE). Furthermore, we will adopt an integration-focussed approach to the development of the system architecture (Vernon et al., 2015) based on CBSE, in general, and the component-port-connector model (i.e., the publish and subscribe model), in particular. In essence, then, we propose an adaptive, compositional agent-based message-passing software architecture to bridge WP4 & WP5 functionality and WP6 use-case behaviours.

3 Expected Outcomes

The research described in this proposal will produce four measurable outcomes, as follows.

1. The identification of the cultural factors that impact on the acceptance of social robots in Africa, the constraints they impose on spatial, non-verbal, and verbal communication and social behavior, and the preferred behavioral traits that are considered appropriate for human-robot interaction in Africa.
2. The development of a suite of culturally-sensitive robot interaction primitives derived from the preferred behavioral traits, implemented on an ARI humanoid robot[4] and also, if technically feasible,[5] Pepper humanoid robot. These primitives will include, for example, maintenance of appropriate interpersonal distance, adjustment of head and gaze direction, deployment of arm movement and hand gestures, and adoption of body posture.
3. The creation of a set of design patterns for culturally-sensitive social interaction in human-robot interaction, tuned to the preferences of African people.
4. A demonstration of the effectiveness of these design patterns in two complementary use cases.

We will validate the research by developing a ROS-based application that involves a humanoid robot giving a guest or group of guests a tour of a typical university laboratory at CMU-Africa. For this, we will specify and implement the functional requirements of the tour (for example, what exhibits to show, what to say to explain their purpose, how to navigate from one exhibit to another) and then factor in the non-functional requirements that address the culturally sensitive interaction while executing the functional elements of the tour (for example, how to greet and address the guest, how to maintain their engagement, how to draw their attention to the exhibit, how to lead them from one exhibit to another).

As noted in Section 1, the potential benefits of these research outcomes include AI technology, i.e., social robotics, that is culturally sensitive and therefore much more likely to be accepted and adopted, contributing to the *inclusive* digital transformation of Africa while simultaneously building research capacity by providing a foundation for future research in human-robot interaction in the form of reusable software and by exposing students to the various aspects of successful research practice. Since the software will be made available on GitHub with an open source licence, it can be freely used by researchers and software developers in

[4] https://pal-robotics.com/robots/ari/.

[5] The ROS package for peper only supports the legacy ROS Indigo distro. CMU-Africa is porting this to ROS Melodic and investigating alternative work-arounds.

The CRAM Cognitive Architecture for Robot Manipulation in Everyday Activities

M. Beetz, G. Kazhoyan, D. Vernon, $n\times$A. N. Other, ...

Institute for Artificial Intelligence, University of Bremen, Germany.

Abstract

This paper describes a hybrid robot cognitive architecture, CRAM, that enables robot agents to accomplish everyday manipulation tasks. It addresses the five key challenges that arise when carrying out everyday activities. These include (i) the underdetermined nature of task specification, (ii) the generation of context-specific behavior, (iii) the ability to make decisions based on knowledge, experience, and prediction, (iv) the ability to reason at the level of motions, and (v) the ability to explain actions and the consequences of these actions. We explore the computational foundations of the CRAM cognitive model: implicit-to-explicit manipulation, generative models, and generalized action plans. We describe the structure and components of the cognitive architecture. We explain the process by which CRAM transforms a generalized action plan for some category of under-determined action into a low-level parameterized motion plan. It does this by using knowledge and reasoning to identify the parameter values that maximize the likelihood of successfully accomplishing the requested action. We demonstrate the ability of a CRAM-controlled robot to carry out a variety of everyday activities in a domestic kitchen environment, such as setting a table for a meal and tidying up afterwards by storing food and placing dirty tableware in a dishwasher. Finally, we consider future extensions that focus on achieving greater flexibility through transformational learning.

Keywords: cognitive architecture, cognitive robotics, everyday activity, generalized action plan, implicit-to-explicit manipulation.

1 Cognitive Architectures

The concept of a cognitive architecture arises from over sixty years of research in various strands of cognitive science, a discipline that embraces neuroscience, cognitive psychology, linguistics, epistemology, philosophy, and artificial intelligence, among others. The primary goal of cognitive science is to explain the underlying processes of human cognition, ideally in the form of a model that can be replicated in artificial agents. It has its roots in cybernetics (Wiener, 1948), but appears as a formal discipline referred to as cognitivism in the late 1950s. Cognitivism, built on the logical foundations laid by the early cyberneticians, taking a computational stance on cognitive function and operation and using symbolic information processing as its core model of cognition and intelligence (Newell & Simon, 1976). Cybernetics also gave rise to the alternative emergent systems approach which recognized the importance of self-organization in cognitive processes, eventually embracing connectionism, dynamical systems theory, and the enactive perspective on cognitive science (Stewart, Gapenne, & Di Paolo, 2010). Hybrid systems seek to combine the cognitivist and emergent approaches to varying degrees in an effort to exploit the knowledge representation and reasoning of symbolic approaches with sub-symbolic representation and inference, typically connectionist or dynamical systems.

An Overview of Different Options for the Implementation of the Situation Model Framework in the CRAM Cognitive Architecture

David Vernon

Institute for Artificial Intelligence,
University of Bremen,
Germany.

Abstract

This report presents the options for the design and implementation of the situation model framework and its integration with the CRAM cognitive architecture. Two options are presented, one symbolic, based on internal simulation using the Unreal Engine, and one sub-symbolic, based on multi-modal associative memories implemented with deep convolutional and recurrent neural networks. The CRAM cognitive architecture provides a computational infrastructure for cognition-enabled robot manipulation in everyday activities. The design, implementation, and integration of the situation model framework will enable more flexible cognitive behaviour by allowing tasks to be planned and executed by exploiting new actions, created on the fly in response to contextual requirements. Two forms of new action are anticipated. One form is based on recombination of perception-action-outcome behavioral episodes, using known actions in new circumstances, as well as generating novel sequences of known actions. The other form is based on generating new motions.

Keywords:

1 The Situation Model Framework

The Situation Model Framework was introduced by Schneider, Albert, and Ritter (2020) as the basis for understanding how cognitive behaviour, in general, and flexible context-sensitive cognitive behaviour, in particular, is realized in humans, animals, and machines. The situation model perspective does not yet commit to a specific computational theory, representation, or process. Rather, at present, it is to be viewed as a framework in which to set such a theory, representation, and process. As such, it sets out in explicit terms the assumptions — or foundations — on which to base and build such a theory. This report documents the first steps to construct and implement such a theory.

Three cross-cutting themes underpin the situation model framework. First, action plays a key role in cognition. Whereas reactions are elicited by earlier events, actions are initiated by a motivated agent, they are defined by goals, and they are guided by prospective information (Vernon, von Hofsten, & Fadiga, 2011; Vernon, 2014). Essentially, actions are organized by goals and not by their trajectories or constituent movement, although of course these are needed to accomplish the action. Control of action is seen as the process that integrates the many different components of an intelligent systems and lies at the heart of flexible context-sensitive cognitive behaviour. Second, complex behaviours can emerge by scaffolding simpler behaviours. Third,

Design and Implementation of the Situation Model Framework

David Vernon

Institute for Artificial Intelligence, University of Bremen, Germany.

Abstract

This paper presents the options for the design and implementation of the situation model framework and its integration in the CRAM cognitive architecture. Two options are presented, one symbolic, based on internal simulation using the Unreal Engine, and one sub-symbolic, based on multi-modal associative memories implemented with deep convolutional and recurrent neural networks.

Keywords:

1 The Situation Model Framework

The Situation Model Framework was introduced by Schneider, Albert, and Ritter (2020) as the basis for understanding how cognitive behaviour, in general, and flexible context-sensitive cognitive behaviour, in particular, is realized in humans, animals, and machines. The long-term vision is to explore situation models both at a functional level and at a mechanistic level. The functional level corresponds to the computational model level in Marr's three-level hierarchy of abstraction (Marr & Poggio, 1977; Marr, 1982), often referred to as the Levels of Understanding framework, while the mechanistic level corresponds to the representation and algorithm level and the hardware or wetware level (Poggio, 2012); see Figure 1. Thus, situation models, fully developed, can be viewed at three levels of abstraction: as a computational model, as a representation and process, and, ultimately, as an implementation of flexible context-sensitive cognitive behaviour. The situation model perspective does not yet commit to a specific computational theory, representation, or process. Rather, at present, it is to be viewed as a framework in which to set such a theory, representation, and process. As such, it sets out in explicit terms the assumptions — or foundations — on which to base and build such a theory.

Underpinning the situation model framework are three cross-cutting themes. First, action plays a key role in cognition. Control of action is seen as the process that integrates the many different components of an intelligent systems and lies at the heart of flexible context-sensitive cognitive behaviour. Second, complex behaviours can emerge by scaffolding simpler behaviours. Third, perception, memory, and action are tightly linked in the context of a given task by prioritized control mechanisms, i.e. attention.

An important feature of the situation model framework is the ability to create rich situation models on-the-fly. However, a cognitive agent can also envisage a number of alternative situation models because models depend on focus and context. Together with short-term memory, this brings the agent into the high-level situation involving a network of related situation models. Now the situation model framework principles can be applied again at this higher level again, e.g. by simulating outcomes of alternative actions on the models. We argue that the capacity to operate in this generative manner on models, rather than just use them, is the distinctive

Culturally Competent Social Robotics for Africa: A Case for Diversity, Equity, and Inclusion in HRI

David Vernon
Carnegie Mellon University Africa
Kigali, Rwanda
vernon@cmu.edu

Abstract

We begin by addressing central role that AI and robotics play in driving the fourth industrial revolution in Africa and the digital transformation of African economies, highlighting the importance of socio-cultural factors in achieving the trust, acceptance, and widespread adoption on which innovation depends. We explain why this is particularly true in the case of social robotics where culturally competence is pivotal for success and we provide examples of the culture-specific knowledge derived from social and cultural norms in African countries. We conclude by unwrapping the concepts of diversity, equity, and inclusion, and we explain how the need for culturally competent social robotics in Africa impacts each of these three issues.

Keywords: social robotics, human-robot interaction, culture sensitivity, trust, diversity, equiity, inclusion

1 Socio-cultural Factors Underpin the Fourth Industrial Revolution in Africa

AI is having an increasingly positive impact in Africa in many sectors such as energy, healthcare, agriculture, public services, and financial services [26]. It has the potential to drive economic growth, development, and democratization, reducing poverty, improving education, supporting healthcare delivery, increasing food production, improving the capacity of existing road infrastructure by increasing traffic flow, improving public services, and improving the quality of life of people with disabilities [28]. AI can empower workers at all skill levels to make them more competitive [4, 22]. Specifically, AI can be used to augment and enhance human skills, not to replace or displace humans, and to do so at all levels, empowering average or low-skill workers to fit better into high-performance environments and take on more complex responsibilities.

AI forms the foundation of the Fourth Industrial Revolution, Industry 4.0 [31]. Countries around the world have prepared AI strategies to ensure they are not left behind. The scope of these strategies is extensive, embracing the research and development necessary to advance the capability of AI in disciplines, the strategies for promoting innovation, and ensuring that this is done in an ethical manner [15]. While most of the effort to develop and exploit AI happens in developed countries, there is increasing awareness of its relevance to developing countries[3, 33], with some countries, such as Rwanda, creating national AI strategies [2] and hosting a World Economic Forum Centre for the Fourth Industrial Revolution (C4IR) [11]. South Africa also hosts a World Economic Forum Centre for the Fourth Industrial Revolution (C4IR) [12]. Africa, a continent with 54 countries, launched a ten year plan in 2022 for the digital transformation of its economies [1].

The Fourth Industrial Revolution and digital transformation requires innovation, something that is not as straightforward as it might seem. [Rose] distinguishes between creativity, invention, and innovation. Creativity can lead to the invention of a novel idea or artefact but innovation carries the creativity and inventions into wider use: the diffusion of that invention and its widespread adoption, leading to substantial social change in the practices of a community of people. He captures this in a beguilingly simple equation: "innovation = invention + exploitation + diffusion", where the invention is commercially developed and exploited, and, significantly, adopted in a wider community of users.

Successful innovation also depends on infrastructure. [Rose] notes that aAIinfrastructure is the unnoticed precondition for technology innovationaAI. There are two forms of infrastructure, the physical and the social. The physical infrastructure might include the availability of electrical power, communications networks, or internet connectivity, something that is taken for granted in developed countries but which cannot always be assumed in developing countries. Of equal importance is the social infrastructure which includes the social conventions that govern people's behaviour and the practices they find acceptable and unacceptable. Social infrastructure heavily impacts on whether or not an invention is adopted and becomes an innovation that can yield benefits for the local community. Again, social infrastructure includes trust and people's sense of what is trustworthy.

[Hofman et al.] define trust as "the expectation that a service will be provided or a commitment will be fulfilled", emphasizing the importance of *expectation* in their definition. Expectations are grounded in the socio-cultural experience of those whose trust is required.

The importance of the cultural context in building trust is emphasized by [Lee and See]. They define culture as "a set of social norms and expectations that reflect shared educational and life experiences associated with national differences or distinct cohorts of workers". An awareness of these social norms and expectations, and the socio-cultural background from which they arise, is crucial to the development of trust in and acceptance of any new technology, including AI-based products and services, and by extension to their diffusion and adoption.

Culture can be characterised in many ways. Hofstede identifies six dimensions in which an understanding of cultural issues should be addressed [17–19]. Others highlight the different ways that cultures perceive time and space, noting that concepts of time in the West and in Africa differ significantly [5]. Without wanting to fall into the trap of generalising across a multitude of cultures and ignoring ethnographic diversity, one can say that time in Africa has traditionally been tied to events, which may be regular or irregular, in contradistinction to the view in the West of time as continually moving from past, to present, to future. These factors have a bearing on how technology, generally, and information technology, powered by AI, in particular, can support an individual or a local community in Africa and whether or not that support, no matter how well intended, will be accepted, trusted, and adopted. Lack of trust can severely and negatively impact the adoption of these services and products, fatally undermining the achievement of the anticipated benefits: "Changes in the factors that affect users' expectations will also impact users' trust levels" [16]. Furthermore, AI and robotics brings their own special factors, e.g. explainability, transparency, lack of bias, all of which have their own influence on whether or not products and services that use AI will be trusted and adopted.

The consequence of this argument is that, if developing countries in Africa are to reap the rewards of adopting AI, innovation needs to be founded on the socio-cultural factors that impact on trust, which is essential for adoption and the realization of the benefits of the technological invention.

To summarize: socio-economic development in Africa must be sensitive to people's culture for it to be successful [27]. Concerning the role of AI, Virginia Dignum drives this home when, in *Responsible AI in Africa* [14], she says "research and development of AI systems must be informed by diversity, in all the meanings of diversity, and obviously including gender, cultural background and ethnicity" [13]. While the overarching agenda of the inclusive digital transformation of Africa is widely recognized to have the potential to be a positive disruptive influence many aspects of the lives of African citizens, the transition from recognition of potential to realization of benefits is not a straightforward matter. The transition depends on turning technological invention into innovation, requiring widespread adoption [30]. However, adoption, especially of AI, depends on trust [3], which, in turn, depends on social and cultural sensitivity [24].

2 Culturally-Competent Social Robotics

The need for artificial intelligence technology to be culturally competent and capable of interacting effectively with humans is perhaps best exemplified by the field of social robotics, a field that is growing quickly. The global social robotics market was valued at USD 1.98 billion in 2020 and is expected to reach USD 11.24 billion by 2026, registering a compound annual growth rate (CAGR) of 34.34% during the period of 2021-2026 [29].

Social robots serve people in a variety of ways and operate in everyday environments, often in open spaces such as hospitals, exhibition centers, and airports, providing assistance to people, typically in the form of advice, guidance, or information. The people interacting with the robot have no special training and they expect the robot to be able to interact with them on *their* terms, not the robot's. There are two aspects to this expectation.

First, it means that social robots need to be able to interpret the intentions of the people with whom they are interacting. This is difficult to achieve because humans do not necessarily articulate their specific needs explicitly when they interact with social robots (or, indeed, with other humans). As [Sciutti et al.] note, "the ability of the robot to anticipate human behavior requires a very deep knowledge of the motor and cognitive bases of human-human interaction". Furthermore, humans use a variety of ways — spatial, non-verbal, and verbal — to communicate their needs, desires, beliefs, intentions, and emotions. These are heavily influenced by social and cultural norms.

Second, and conversely, humans have expectations of the robot's behavior and they have a distinct preference for robots that exhibit legible and predictable behavior [32]. Since people make predictions based on what they are used to, robot behaviors must be tuned to the socio-cultural context in which they are operating and their spatial, non-verbal, and verbal communications must reflect the social and cultural norms of their interaction partners.

A culturally competent robot requires at least five elements: (i) cultural knowledge representation, (ii) culturally sensitive planning and action execution, (iii) culturally aware multimodal human-robot interaction, (iv) culture-aware human emotion recognition, and (v) culture identity assessment, habits, and preferences [8], as well as intention recognition and some capacity for forming a theory of mind [34].

Ideally, culturally competent robots by combine top-down and bottom-up approaches based on the predetermined profiles of a cultural group and the cultural profiles derived from the behaviors of individuals, respectively [23].

Culture-specific knowledge, i.e., knowledge of cultural and social norms, must be encapsulated in a knowledge ontology for use in a knowledge representation and reasoning system when selecting culturally sensitive robot behavior and recognizing culturally dependent human behavior [10].

In short, social robots must be culturally competent to be effective [9, 23] and therefore social robotics must embrace cultural diversity if their are to be widely adopted.

3 Diversity in Cultural Competence

While there are studies on cultural differences in the acceptance of robots in the West and East, e.g., [7, 8, 21], similar studies of the cultural factors that impact of acceptance in Africa have not been reported [6]. The recent survey by [25] briefly mentions Egypt, Tunisia, Libya, and Sudan but only to contrast perceptions with the Gulf region when interacting with an Arabic robot.

This highlights the need to identify culture-specific knowledge through ethnographic research.

The specific factors that underpin effective human-robot interaction include spatial interaction (proxemics, localization and navigation, socially appropriate positioning, initiation of interaction, communication of intent), nonverbal interaction (gaze and eye movement, deictic, iconic, symbolic, and beat gesture, mimicry and imitation, touch, posture and movement, and interaction rhythm and timing), and verbal interaction (speech, speech recognition, language understanding, speech generation) [6]. These spatial, nonverbal, and verbal *interaction factors* must be adjusted to reflect the *traits* that would make social robots effective in Africa.

It is important to recognize that there are many different cultures in Africa, with many different norms for deictic, iconic, and symbolic manual gesturing, as well as gestures involving eye gaze, head tilt, eyebrows, and body posture, generally. Similarly, there are many different ways in which spoken language can express nuances of meaning by modulating amplitude and timbre.

~~Having identified~~ the verbal and non-verbal social and cultural norms of human interaction that are prevalent in different countries in Africa, they can then be encapsulated in the behavioral traits of social robots so that these robots engage with African people in a manner that is consistent with their expectations of acceptable — respectful — social interaction, rather than using inappropriate or insensitive social behaviors and modes of interaction from the West or the East.

4 Interaction in Africa

While much more ethnographic research is required, Tables 1 and 2 present some preliminary findings on the cultural factors that impact on the acceptance of social robots in Africa, the preferred behavioral traits that are considered appropriate for human-robot interaction in Africa, and design patterns for culturally-sensitive social interaction in human-robot interaction, tuned to the preferences of African people. We base the design patterns on the eight design patterns for sociality in human-robot interaction proposed by [Kahn et al.].

Table 1. African Culture-specific Knowledge

No.	Socio-cultural Norm or Trait
1	An appropriate greeting should precede interactions.
2	The younger interaction partner is expected to initiate greetings.
3	The younger interaction partner is expected to bow when greeting an elderly person or when rendering a service.
4	Interrupting a human interaction partner is considered rude.
5	~~Giving~~ objects, pointing, and even waving with the left hand is considered rude. Humanoid robots should be right-handed by default. A social robot should adjust its orientation in order to hand over an item with its right hand.
6	Language is a valued aspect of culture; native languages should be used for verbal interaction.
7	Africans are more receptive when they feel they are respected. Robot behaviors and modes of speech should convey a sense of respect for the interaction partner.
8	Most African countries display friendliness through physical contact i.e., touch. However, a robot should ask for permission before touching. ~~For example, a social robot should ask "can I hug you?" or "will you shake my hand?" before attempting to do either.~~
9	A robot should either respond with a friendly smile when touched or express friendliness by touching its user during a conversation.
10	Africans are energetic people who like rhythmic movement. ~~They~~ love dancing to ~~highly~~ rhythmic beats, ~~and they easily~~ embrace anyone that exhibits an appreciation of rhythm. ~~Humanoid~~ robots behave ~~accordingly~~.
11	When communicating information, it is appropriate to intermittently keep eye contact; lack of eye contact depicts disrespect as it shows divided attention during the interaction.
12	In some African countries such as Kenya, making eye contact during a conversation is considered ~~very~~ disrespectful. To increase the acceptance of humanoid robots in such countries, eye contact and gaze must be minimized during an interaction.
~~13~~	~~African men almost never hug.~~
14	A robot is expected to take a subordinate role.
15	Opting to go first in an interaction is sometimes considered rude.

5 Unpacking Diversity, Equity, & Inclusion

We conclude be considering the ~~social~~ implications of a discipline of culturally competent social robotics that fully embraces diversity, equity, and inclusion. To do this, we need to unpack what is meant by these terms.

(*) These findings are based on a survey of 16 people from x countries in Africa.

Table 2. Africa-centric Design Patterns for Social Robots, adapted from [20]

No.	Design Pattern
Initial Introduction	The robot must acknowledge the presence of the person. The robot should exhibit deference by bowing or kneeling when encountering someone from the older generation. The robot should greet first. "Good morning, Sir" or "Good afternoon, Madam" are more respectful greetings than "Hello" or "Hi" which are are used with peers and subordinates. Personal and intimate distances should be respected during interaction.
Reciprocal Turn Taking	The robot should respectfully give the initial turn to the human interaction partner.
Didactic Communication	Pointing a hand directly at someone is disrespectful. For deictic gestures, the robot should use its left hand. Gesture with an open palm rather than pointing a finger.
Personal Interests and History	The robot should avoid trying to share personal history since it will be perceived to be inauthentic. The robot should focus on and highlight its functional usefulness.
In Motion Together	The robot should explicitly say "Please come along" to remove any ambiguity of intention. The robot should not walk too far ahead when showing the way.
Recovering from Mistakes	The robot should apologize profusely. The robot should slightly bow when introducing itself and after it makes a mistake.
Physical Intimacy	In general, men should not be hugged. Women must be hugged carefully so as not to offend. Personal space should be entered only with prior consent. Do not pass in between two people that are interacting.
Claiming Unfair Treatment or Wrongful Harm	To enhance the perception that the robot is being respectful, the robot should not be aggressive by claiming unfair treatment.

Diversity concerns the many different dimensions in which people differ. Gender, sexual orientation, race, culture, socio-economic status, traditions, education, age, religious and spiritual beliefs, nationality, ethnicity, experience, physical ability: these are just some of the facets that characterize diversity. Diversity creates opportunities for greater mutual understanding of the individual contribution that a person of each background can make. It does this by breaking down barriers — typically manifested as preconceptions and bias — and exposing what is special and positive in each and every individual. In a sense, diversity is a means to an end: a way of tapping into everyone's potential and using that potential to empower everyone else through mutual respect.

Realizing this makes it easier to understand the concept of equity. In contrast to equality, equity is less concerned with treating everyone equally and more about doing what is necessary to allow each person to make their special individual contribution and to participate just as much as everyone else. Equality is passive; equity is active. It is the act of empowering, the process that leverages the potential latent in diversity. Without equity, the power of diversity cannot be realized.

By themselves, diversity and equity create the necessary conditions but can't guarantee that these conditions will lead to the positive interaction between each person in that environment. This is what inclusion means: that each person feels they belong in that environment and that their place in that environment is valued. It is not enough that they are present and empowered, but that they are visibly, openly, and transparently valued by everyone else. Naturally, this is a reciprocal process and, therefore, it can only be achieved by mutual respect for the perspectives of others. This is the essence of empathy. It necessitates that each individual actively adopts the perspective of others and sees the value in it, irrespective of whether or not she or he agrees with it, at that moment in time. Eventually, exposure to these perspectives brings about a greater and a deeper understanding, and a more harmonious, effective, and fulfilling way of interacting with one another. Inclusion is the psychological prerequisite of mutual empathy that allows diversity and equity to function effectively in creating a better, richer, more enlightened mode of interaction. This is neatly summarized by the poet George Eliot (the pen name of Mary Ann Evans):

> The highest form of knowledge is empathy,
> for it requires us to suspend our ego and
> live in another's world.

This is the essence of a theory of mind, where someone, or some social robot, takes a perspective on the needs, desires, beliefs, intentions, and emotions of others, understanding the manner in which these are modulated by social and cultural norms.

The development of culturally competent social robots that can achieve this level of understanding of their interaction partners would not only facilitate effective human-robot interaction by leverage cultural and social norms but it would also contribute to the empowerment of the individuals with which the social robots are interacting by recognizing and valuing the importance of those individuals' cultural heritage. This, in a nutshell, is the practical and principled importance of diversity, equity, and inclusion in human-robot interaction.

References

[1] African Union. 2022. The Digital Transformation Strategy for Africa (2020 – 2030). https://au.int/sites/default/files/documents/38507-doc-dts-english.pdf

[2] AI Rwanda. 2021. The Future Society – Development of Rwanda's National Artificial Intelligence Policy. https://thefuturesociety.org/2020/08/31/development-of-rwandas-national-artificial-intelligence-policy/

[3] C. Delmus Alupo, D. Omeiza, and D. Vernon. 2022. Realizing the Potential of AI in Africa. In *Towards Trustworthy Artificial Intelligence Systems*, M. I. Aldinhas Ferreira and O. Tokhi (Eds.). Springer.

[4] B. Simons. Artificial Intelligence hits African companies, The Africa Report. March 2017. [n. d.]. https://www.theafricareport.com/833/artificial-intelligence-hits-african-companies/.

[5] S. F. Babalola and O. A. Alokan. 2013. African Concept of Time, a Socio-Cultural Reality in the Process of Change. *Journal of Education and Practice* 4, 7 (2013), 143–147.

[6] C. Bartneck, T. Belpaeme, F. Eyssel, T. Kanda, M. Keijsers, and S. Sabanovic. 2020. *Human-Robot Interaction – An Introduction.* Cambridge University Press.

[7] C. Bartneck, T. Nomura, T. Kanda, T. Suzuki, and K. Kennsuke. 2005. Cultural Differences in Attitudes Towards Robots. In *Proceedings of the AISB Symposium on Robot Companions: Hard Problems And Open Challenges In Human-Robot Interaction.* 1–4.

[8] B. Bruno, N. Y. Chong, H. Kamide, S. Kanoria, J. Lee, Y. Lim, A. K. Pandey, C. Papadopoulos, I. Papadopoulos, and F. Pecora. 2017. The CARESSES EU-Japan project: Making assistive robots culturally competent. In *arXiv 1708.06276.*

[9] B. Bruno, N. Y. Chong, H. Kamide, S. Kanoria, J. Lee, Y. Lim, A. K. Pandey, C. Papadopoulos, I. Papadopoulos, F. Pecora, A. Saffioti, and A. Sgorbissa. 2017. Paving the way for culturally competent robots: A position paper. In *26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN).* Lisbon, Portugal, 553–560.

[10] Barbara Bruno, Carmine Tommaso Recchiuto, Irena Papadopoulos, Alessandro Saffiotti, Christina Koulouglioti, Roberto Menicatti, Fulvio Mastrogiovanni, Renato Zaccaria, and Antonio Sgorbissa. 2019. Knowledge representation for culturally competent personal robots: requirements, design principles, implementation, and assessment. *International Journal of Social Robotics* 11, 3 (2019), 515–538.

[11] C4IRRwanda. [n. d.]. Centre for the Fourth Industrial Revolution Rwanda. https://www.weforum.org/centre-for-the-fourth-industrial-revolution/c4ir-rwanda

[12] C4IRSA. [n. d.]. Centre for the Fourth Industrial Revolution South Africa. https://www.weforum.org/centre-for-the-fourth-industrial-revolution/c4ir-south-africa

[13] V. Dignum. 2023. Responsible Artificial Intelligence: Recommendations and Lessons Learned. In *Responsible AI in Africa – Challenges and Opportunities*, D. O. Eke, K. Wakunuma, and S. Akintoye (Eds.). Palgrave Macmillan Cham.

[14] D. O. Eke, K. Wakunuma, and S. Akintoye (Eds.). 2023. *Responsible AI in Africa – Challenges and Opportunities.* Palgrave Macmillan Cham.

[15] EUAI. 2021. European Commission Ethics Guidelines for Trustworthy AI. https://ec.europa.eu/futurium/en/ai-alliance-consultation

[16] L. J. Hofman, K. Lawson-Jenkins, and J. Blum. 2006. Trust beyond security: an expanded trust model. *Commun. ACM* 49, 7 (2006).

[17] G. Hofstede. 1980. *Cultures Consequences: International Differences in Work-related Values.* Sage Publications.

[18] G. Hofstede. 1991. *Cultures and Organizations: Software of the Mind.* McGraw-Hill.

[19] G. Hofstede, G. J. Hofstede, and M. Minkov. 2010. *Cultures and Organizations: Software of the Mind.* McGraw-Hill.

[20] P. H. Kahn, N. G. Freier, T. Kanda, H. Ishiguro, J. H. Ruckert, R. L. Severson, and S. K. Kane. 2008. Design patterns for sociality in human-robot interaction. In *Proc. 3rd ACM/IEEE International Conference on Human-Robot Interaction.* 97 – 104.

[21] F. Kaplan. 2004. Who is afraid of the humanoid? Investigating cultural differences in the acceptance of robots. *International Journal of Humanoid Robotics* 1, 3 (2004), 1 – 16.

[22] J. E. Kelly. 2015. *Computing, cognition and the future of knowing.* White Paper. IBM Corporation.

[23] A. Khaliq, U. Köckemann, F. Pecora, A. Saffiotti, B. Bruno, C. Recchiuto, A. Sgorbissa, H.-D. Bui, and N. Chong. 2018. Culturally aware Planning and Execution of Robot Actions. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).* Madrid, Spain, 326–332.

[24] J. D. Lee and K. A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors* 46, 1 (2004), 50–80.

[25] V. Lim, M. Rooksby, and E. S. Cross. 2021. Social Robots on a Global Stage: Establishing a Role for Culture During Human–Robot Interaction. *International Journal of Social Robotics* 13 (2021), 1307–1333.

[26] L. Novitske. 2018. The AI invasion is coming to Africa (and it's a good thing). *Stanford Social Innovation Review* (February 2018).

[27] A. Olasunkanmi. 2011. Development in Africa: The Need for a Culture-Sensitive Approach. *Journal of Sociology and Social Anthropology* 2, 2 (2011), 97–101.

[28] N. Pillay and Access Partnership. 2018. *Artificial Intelligence for Africa: an opportunity for growth, development, and democratization.* White Paper. University of Pretoria.

[29] Research and Markets. 2022. Global Social Robots Market. https://www.researchandmarkets.com/reports/5120156

[30] J. Rose. 2010. *Software Innovation: eight work-style heuristics for creative software developers.* Software Innovation, Dept. of Computer Science, Aalborg University.

[31] K. Schwab. 2021. The Fourth Industrial Revolution: What it Means, How to Respond, Word Economic Forum. https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/

[32] A. Sciutti, M. Mara, V. Tagliasco, and G. Sandini. 2018. Humanizing Human-Robot Interaction: On the Importance of Mutual Understanding. *IEEE Technology and Society Magazine* 37, 1 (2018), 22–29.

[33] D. Vernon. 2019. Robotics and Artificial Intelligence in Africa. *IEEE Robotics & Automation Magazine* 26, 4 (December 2019), 131–135.

[34] D. Vernon, S. Thill, and T. Ziemke. 2016. The Role of Intention in Cognitive Robotics. In *Toward Robotic Socially Believable Behaving Systems*, A. Esposito and L. C. Jain (Eds.). Intelligent Systems Reference Library, Vol. Vol. 105. Springer, 15–27.

FRUBOTS: Cognition for Frugal Robots by Design

David Vernon

Carnegie Mellon University Africa, Rwanda.

September 12, 2022

Abstract

FRUBOTS: Future-directed Optimal Regulation of Energy Utilization in Robot Behaviors Over Multiple Timescales

This optimal regulation is achieved through allostasis-based cognition in an autonomic cognitive architecture.

Keywords: action, action selection policy, adaptation, allostasis, autonomic regulation, cognition, efficacy, efficiency, embodiment, energy, goal-directed behavior, learning, metabolism, robot, safety, social interaction, prediction, prospection, urgency, value system.

1 Introduction

A frugal robot is a robot that exerts the minimal effort to accomplish a task or engage in an interaction. In turn, this means that it expends the minimum energy. It works efficiently, with minimal effort where possible.

Different dimensions of minimizing exertion.

Efficient deployment of actuators: reaching, grasping, looking, walking, running. The focus is on smooth motion dynamics.

Efficient sequencing of primitive movements, avoiding unnecessary motion. The focus is on finding the most efficient path through a network of possible movements involved in reaching a goal.

Efficient adaptation of primitive movements, selecting the most effective responses in dynamic environments. The focus is on anticipating ways in which the environment can obstruct or disrupt the current sequence of primitive motions - the current action - and preparing alternative strategies.

The core thesis in the draft is that frugality is an innate, evolutionary attribute of biological systems, that it is enhanced by learning through experience and interaction with conspecifics, and that it can be used by cognitive processes to effect long-term energy-efficient motion and action selection and modulation policies.

There are two strands in the proposed research.

(i) Generating operationally-effective energy strategies for four elements of effective action,

(ii) Learning energy strategies by joint action with conspecifics through physical interaction and verbal communication.

Both leverage cognition, understood in its broadest sense of integration of perception, knowledge and reasoning (both implicit sub symbolic and explicit symbolic) to achieve effective action. Here, the prospective aspect of cognition is particularly important in anticipating the need to act, the outcome of those actions, and, especially for FRUBOTS, the metabolic energy impact of those actions.

In the first strand, the four elements of effective action span different timescales, from motion to action to task execution. These are:

(a) Form of motion: minimizing jerk on the basis of object affordances. (b) Form of action: minimizing torque by exploiting an action language to produce energy-efficient motion sequences based on the goal of the action. The goal of the action is reflected in the semantics of the action sentence and the specific form of the sentence reflects a parsimonious, frugal way of expressing that meaning. (c) Required force based on experience and expectation. (d) Adaptive force; perform the work required to achieve the desired effect by adapting the speed of an action or motion and the time it takes to perform the action or motion.

All four elements are achieved by implicit or explicit reasoning on implicit or explicit knowledge to predict the energy impact of adopting or selecting a candidate motion, action, or task strategy. It is a tacitly and legitimately assumed that cognition operates at all timescales.

In the second strand, there are two elements at play. These are:

(a) Modulation or selection of movement based on energy expenditure in joint action. (b) Modulation or selection of actions and movements based on knowledge exchanged through explicit verbal (language-based) communication.

The former element implies a shared motor representation: joint action entails shared goals and shared intentions, the intentions being inferred by implicit sharing of motor knowledge through theory of mind, a process of perspective taking that is, by definition, empirically easier when anthropomorphic robots, also known as humanoid robots, are being used.

The later element finesses the problem of theory of mind at the potential cost of neglecting possibly important non-verbal communication. This could be addressed by including non-verbal communication.

"We assert that employing basic principles in human biological motion ... and human cognition ... will enable FRUBOTS to optimize resources at all times." The basis for this assertion is as follows.

As I understand it, the draft has three theories in mind: (a) the potential field approach to motion specification, (b) the action language as a compositional mechanism for action understanding and action generation, and (c) the goal-central action notation for encapsulating procedural knowledge. Apologies if I've misunderstood something or misrepresented anything. As the draft is elaborated, it would be good to bind these together closely to show how the learning from demonstration that is implicit in the approach will be realized and how the multi-timescale motion and action prospection will be realized, ideally in the context of a proto cognitive architecture.

There is one mention of a value system that, I assume, will act as an objective function for motion and action selection. As noted above, this value system will be both innate and learned. It would be nice if the proposal could suggest a candidate value system and make explicit how it will capture the multiple timescales involved in motions (or movements), actions, and task plans.

For example, one could contemplate the possibility of an autonomic cognitive architecture that predictively balances its long-term energy expenditure through allostasis - systemic, adaptive and predictive self-regulation - rather than homeostasis. It might possibly leverage insights from predictive processing, active inference, and an expanded version of the free-energy principle that minimises long-term surprisal. In this case, surprisal would be extended to include not just the inconsistency of internal model and perception of actual sensorimotor contingency, but also unexpected, wasteful energy drain. Thus, a frugal robot might be viewed at one that uses proprioception, exteroception, and, crucially, interception to optimise its long-run use of energy resources. I think these ideas are consistent with the positions taken in the draft that frugality is a goal that is encoded by evolution in embodied biological systems and enhanced through learning by interaction with conspecifics. It would also allow the value system to be cast in information theoretic terms, thereby allowing uncertainty to be modelled explicitly, as well as catering for the need to factor some entropy into the system configuration to facilitate adaptivity and ensure that unexpected energy drains are not catastrophic. Clearly, all these ideas would need to be fleshed out and integrated seamlessly with the six elements set out in the draft. I'm unsure of the difficulty of that challenge but at least it might be worth exploring. In any case, an energy-specific time-dependent value system will be needed, along with mechanism to optimise the robot behaviours with respect to that value system. There may be many other, more plausible, options.

Our aim is not to develop a way of minimizing instantaneous energy expenditure because there will inevitably be situations which it is imperative to expend energy hungry such as in the case of emergencies. There may also be situations where short-term energy expenditure may result in long-term energy savings. A greedy algorithm for energy expenditure by minimizing local (i.e., immediate) expenditure may not lead to a safe or optimal strategy. This situation is analogous to the ability of humans to defer immediate reward and is a key element of human cognitive development.

Regarding the idea in the email thread of addressing neuromorphic computing and event-based sensors, this is could be well worth pursuing - Krichmar and others make a compelling case for their energy efficiency - but the concerns Yiannis raises are important: we'd need to minimise the risk of criticism that the consortium has all the requisite expertise and experience in this area.

Cognition is effective action. unpack.

Effective in robotics typically means achieving a goal in the face of uncertainty, with incomplete knowledge, with underdetermined task specification, leading to an intractably large search space of possible solutions. Achieving success in the face of these challenges is the essence of intelligence. We characterize this as functional effectiveness: success in goal achievement and we refer to this as efficacy.

However, other factors are also in play that are not strictly functional. These related to the fact that all real systems operate with additional constraints: restricted computational resources, limited supply of energy, while being able to act safely but take urgent action, when necessary. The constraints are often mutually incompatible: the urgency of the situation may require a less safe action and one that requires additional energy resources, for example.

The research question then is: how can we design a system that satisfies all these constraints or exhibits these four attributes?

Efficacy (effectiveness of capacity for goal achievement) Efficiency (power sensitivity) Urgency (time sensitivity) Safety (risk sensitivity)

One that weaves the functional and the non-functional together into an adaptive system, cf. (W. Ross Ashby , (1960)) a system that can follow high-level instructions and optimally balance the four factor of effective action. This requires that the action selection mechanism uses a four-dimensional value system or objective function when weighing alternative options for action and, crucially, that it does so across the multiple time-scales associated with elementary motions, individual actions, tasks, and behaviors, applying different value system weighting depending on scale. This suggests that the value system is five-dimensional, time being the fifth dimension. The ability to do this is the essence of cognition so we can re-cast cognition is effective action as effective action requires cognition.

Actions are carried out by motivated agents (cf. Claus von Hofsten) and the manner in which they carry them out reflects their value systems.

Value systems have both an intrinsic and an extrinsic element, related to constitutive autonomy and behavioral autonomy, respectively. In turn, this means that the value system is a function of current and predicted future states, mediated through exteroception (the current and predicted future state of the world), proprioception (the current and predicted future state of the physical state of the robot), and interoception (the current and predicted metabolic state and reserves; effectively the current and future capacity for effective action). It would be a weighted function of the safety (or risk) associated with an action, the energy expenditure of the action, and the urgency of the action required to go from the current state to the goal state.

Intrinsic elements are related to the self-organization of the system: the regulation of its autonomic and metabolic state, the internal capacity for future action, through processes of homeostasis, allostasis, and adaptation (ontogenetic adaptation, rather than genetic adaptation, i.e. development).

Extrinsic elements are learned through interaction with the world, through experimentation and social intercourse. Social discourse takes several forms:

Passive observation of third-party demonstration. Cooperation, recognizing the goals of a third party and facilitating their achievement. Collaboration, sharing goals and intentions with a third party and working together through joint action to achieve them.

All forms of social intercourse require an ability to take a perspective on the goals and intentions of the third party, often referred to as theory of mind.

[DV 5D value system, regulating two complementary but mutually-dependent states (autonomic/metabolic and behavioral), informed by four modes of interaction.

Normally, the process of action execution is as follows. From the goal, infer a set of possible solution actions (cf. Hammer), simulate the outcome of these actions, then select actions that result in the state most closely matched with the goal state. FRUBOT differs in that it selects the action based on the 5D value system (one functional, three non-functional, one time).

The goal of FRUBOTs is to realize a system that can use cognition to regulate power demand in robots to optimize long-term depletion of energy resources while ensuring actions are effective, safe, and timely, in the sense of being able to adapt its response to different levels of urgency.

Candidate cognitive architectures to be adapted for FRUBOT include

DACS ISAC CLARION CRAM

Need a catalog of capabilities and extensions, as well as an assessment of the feasibility of developing a completely new cognitive architecture.

Internal models as interoceptive essential variables, cf. Ashby

Cognition is the process by which a system adapts.

Adaptation: "A form of behaviour is adaptive if it maintains the essential [metabolic and autonomic] variables within physiological limits." W. Ross Ashby, 1960, p. 58.

The value system takes eight inputs, as follows.

Motion specification Action specification Urgency level (i.e., tolerance to instantaneous energy cost) Safety level (i.e., tolerance to safety cost, indicated by uncertainty of successful completion of motion)

Current system energy level Predicted system energy level

Current subsystem energy levels Predicted subsystem energy levels

The value system produces four outputs, as follows.

Merit value over motion timescale Merit value over action timescale Merit value over task timescale Merit value over behaviour timescale

The value system provides input to the action selection process along with the motion and action specifications. In turn, the action selection process identifies the selected motion and action. The action selection process uses an action selection policy that is adapted over time on the basis of experience, both external (i.e., policies learned by observing the actions of other agents and the success of its own actions) and internal (i.e., policies learned by observing the robots metabolic history, i.e., by interoception).

Energetic optimization can be achieved to an extent by simple control systems, reactive systems, at the single action/movement level. However, cooperation is impossible without sharing optimization strategies and without sharing motor representations (e.g. think of handshaking). Efficient learning of optimization strategies for single actions or action sequences both in cooperation and on ones own, requires communication/language/verbal and non-verbal, sensorimotor interaction, especially as complexity increases (think developmentally: we communicate to infants/toddlers the mostefficient/economic/frugal and safe* way of doing something, out of many different ways the infant may experiment with for achieving a goal). We will explore cooperation among conspecifics (e.g. human to human, human to humanoid robot, humanoid to humanoid the use of an anthropomorphic robot has thus an advantage with regards to frugality) with the vision that at the end of this project we may become aware of universals/primitives (of energy expenditure optimization) across agents/organisms regardless their morphology. Our experiments (and corresponding use cases/demonstrations/scenarios) will involve: 1. Learning how to achieve a goal that requires a single action undertaken by the learner alone 2. Learning how to achieve a goal that requires an action sequence undertaken by the learner alone 3. Learning how to achieve a goal that requires a single action undertaken through cooperation among conspecifics 4. Learning how to achieve a goal that requires an action sequence undertaken through cooperation among conspecifics Some questions to keep in mind: How far ahead [in a task plan] should one look to save energy? Can we compare our frugal solution to current approaches? (this takes us back to the issue of proposing a value system.)

2 Psychology and Development

The development of an autonomous agent is driven by motives and value systems [1, 2]. There are social motives and exploratory motives, reflecting, loosely, the psychology of development espoused by Vygotsky and Piaget, respectively [3, 4, 5, 6]. Both motives function from birth and provide the driving force for action throughout life.

The social motive focusses on finding comfort, security, and satisfaction through interaction with others, allowing the agent to learn new skills and acquire knowledge about the world about the world from the experience of others. It is manifest from birth in the tendency to fixate social stimuli, imitate basic gestures, and engage in social interaction. The social motive is so important that it has been suggested that without it a person will stop developing altogether. Social motives also include a strong need to belong, a drive for self-preservation, and the need for cognitive consistency with other [7].

There are at least two exploratory motives, one to do with the discovery of novelty and regularity in the world and the other to do with finding out about the potential of one's own action capabilities. Infants are visually attracted to new objects and events but after a while they cease to be attracted. Infants also have a strong motivation to discover what they can do with objects, especially with respect to their own sensorimotor capabilities and the particular characteristics of their embodiment. Effectively, infants have a strong motivation to discover the affordances of objects around them.

The motivation to seek new ways of doing things is very strong and it can override ways of doing something that has already become established through previous development. This means that skills are developed non-monotonically: sometimes you get worse at doing something before you get better at it. So, it isn't necessarily success at achieving task-specific goals that drives development in infants but rather the discovery of new modes of interaction with the world in which the infant is embedded: the acquisition of a new way of doing something through exploration [8, 9].

In developing new skills and in learning how to act and interact, prospection comes to the fore. Actions are initiated and executed by a motivated subject and they are defined by the goal of the action, not the specific movements by which the goal is achieved. More especially, they are guided by prospective information [10]. For example, when performing manipulation tasks or observing someone else performing them, people fixate on the goals and sub-goals of the movements, e.g. the point where an object is to be grasped, the target location of object, and the support surface, not on the body parts, e.g. the hands or the grasped object. In other words, gaze is governed by predictive motor control. Again, we see that development is focussed on expanding the repertoire of actions and extending the time horizon of an agent's predictive capacity.

The phased aspect of development is particularly relevant in the manner in which infants and children come to understand the intentions of others and to help them achieve their goals. It takes several years for human infants to develop the requisite abilities.

During the first year of life the progressive acquisition of motor skills facilitates the development of an ability to understand the intentions of other agents, initially by anticipating the goal of simple movements and eventually understanding more complex goals. During this period, the ability to infer what another agent is focussing their attention on and the ability to interpret emotional expressions begins to improve substantially. Around 14 to 18 months of age children begin to exhibit instrumental helping behaviour, i.e. they display spontaneous, unrewarded helping behaviours when another person is unable to achieve his goal [11]. This is a critical stage in the development of a capacity for collaborative behaviour, a process that progresses past three and four years of age. Around 2 years of age children start to solve simple cooperation tasks together with adults [12]. This phase of development sees the beginning of shared intentionality where a child and an adult form a shared goal and both engage in joint activity. Children seem to be motivated not just by the goal but by the cooperation itself, i.e. the social aspect of the interaction. The ability to cooperate with peers and become a social partner in joint activities develops over the second and third years of life as social understanding increases [13]. More complex collaboration, which necessitates the sharing of intentions and joint coordination of actions, appears at about three years of age when children master more difficult cooperation tasks such as those involving complementary roles for the two partners in a collaborative task [14]. At three years of age, children begin to develop the ability to cooperate by coordinating two complementary actions. By three-and-a-half years of age children quickly master the task, can deal effectively with the roles in the task being reversed, and can even teach new partners [15]. The motives which drive instrumental helping are simpler than those of collaborative behaviours: they are based on wanting to see the goal completed or wanting to perceive pleasure in the human at being able to complete it. In this case, the motivational focus is solely on the needs of the second agent and the needs of the first agent do not figure in this. The motives underlying collaborative behaviour are more complicated. In this case, the intentions and the goals have to be shared and the motivational focus is on the needs of both agents.

The foregoing has provided some insights into the ontogenetic aspect of development, i.e. the developmental process itself and, consequently, it suggests some of the social and exploratory elements that an agent must possess for development to happen. Drawing on this, let us now look at development from the perspective of phylogeny, i.e. the agent's cognitive architecture.

3 Cognitive Architectures and Development

While the term cognitive architecture derives from Allen Newell's pioneering work in cognitivist cognitive science, and in particular to his work and his colleagues work on unified theories of cognition [16, 17, 18], it is also used by those who work in enactive systems to refer to the phylogenetic configuration of a new-born or newly-created cognitive agent: the initial state from which it subsequently develops. An appropriately-configured cognitive architecture doesn't guarantee successful development because, as we saw in the previous section, development also requires exposure to an environment that is conducive to development, one in which there is sufficient regularity to allow the system to build a sense of understanding of the world around it, but not excessive variety that would overwhelm an agent which has inherent limitations on the speed with which it can develop. Thus, cognition has two necessary elements: phylogeny and ontogeny: a cognitive architecture and gradually-acquired experience.

Although several guidelines for configuring cognitive architectures have been proposed, e.g. [19, 20, 21, 22, 23], few address development explicitly, mainly because these guidelines derive from work in cognitivist cognitive science. On the other hand, Jeffrey Krichmar proposes five design principles for developmental artificial brain-based devices [24, 25, 26] which are also applicable to cognitive architectures.

First, the cognitive architecture should address the dynamics of the neural elements in different regions of the brain, the structure of these regions, and especially the connectivity and interaction between these regions. In other words, a developmental cognitive architecture should make explicit the operation of the system as a whole.

Second, the architecture should support perceptual categorization: i.e. the capacity to organize unlabelled sensory signals of all modalities into categories without prior knowledge or external instruction. In effect, this means that the system should be autonomous and, as a developmental system, it should be a model generator, rather than a model fitter (a point also emphasized by John Weng [27]).

Third, a developmental system must have a physical instantiation, i.e. it must be embodied, with the system's morphology conditioning the agent's understanding of its environment.

Fourth, the system should have some minimal set of innate behaviours or reflexes in order to explore and survive in its initial environmental niche. From this minimum set, the system can develop so that it improves its behaviour over time.

Fifth, and of particular importance to the argument in this article, a developmental system should have a means to adapt. This entails the presence of a value system, i.e. a set of motivations that guide or govern its development [1, 2]. These should be non-specific (in the sense that they don't specify what actions to take) modulatory signals that bias the dynamics of the system so that the global needs of the system are satisfied: in effect, so that the system's autonomy is preserved or enhanced.

Directly or indirectly, these value systems should manifest the social motives that enable fixation on social stimuli, imitation of basic gestures, and engagement in social interaction, and exploratory motives that facilitate the discovery of novelty and regularities in the environment and the system's own action capabilities, in line with the brief synopsis of infant development in the previous section.

4 Autonomy and Development

So far, so good. However, enactive cognitive systems are, first and foremost, autonomous systems. As noted already, autonomy is a difficult concept to tie down [28] and there are several perspectives on what it means [29]. Nonetheless, few would disagree that autonomy is degree of self-determination of a system, i.e. the degree to which a system's behaviour is not determined by the environment and, thus, the degree to which a system determines its own goals [30, 31, 32, 33]. For biological autonomous agents, as well as bio-inspired artificial agents, the issue of autonomy is one of survival in the face of precarious conditions, operating in an uncertain possibly-dangerous constantly-changing environment. To do this, it must keep itself intact as an autonomous system, both physically and organizationally as a dynamic self-sustaining entity. The self-maintenance of autonomy is a crucial aspect of enactive cognitive agents [34], continually repairing damage to itself. Since it is better if the agent can avoid damage in the first place, cognition, as a prospective modulator of perception and action, is one of the primary mechanisms at the agent's disposal [35] to anticipate the need for action and the outcome of that action.

From this perspective, autonomy, aided by cognition, *is* the self-maintaining organizational characteristic of living creatures that enables them to use their own capacities to manage their interactions with the world in order to remain viable [36]. In other words, autonomy is the process by which a system manages — self-regulates — to maintain itself despite the precarious conditions with which the environment continually confronts it. Arguably, Autonomy and autonomy-preserving processes are the foundation of cognition [34].

While more than twenty types of autonomy can be distinguished [37], two broad classes can be discerned: behavioural autonomy and constitutive autonomy [29, 35]. Behavioural autonomy is concerned with the external behaviour of the system: the extent to which the agent sets its own goals and its robustness and flexibility in achieving them as it interacts with the world around it, including other cognitive agents. Constitutive autonomy is concerned with the internal organization and the organizational processes that keep the system viable, maintaining itself as an identifiable autonomous entity. Indeed, Maturana and Varela, whose work provided the inspiration for the enactive view of cognition, define autonomy as "the condition of subordinating all changes to the maintenance of the organization" [38]. Constitutive autonomy and behavioural autonomy are related: an agent can not deal with uncertainty and danger if it is not organizationally — constitutively — equipped to do so. Behaviour depends on internal preparedness but appropriate behavioural is needed to allow the agent to achieve the requisite environmental conditions — through interaction — for constitutive autonomy to be able to operate effectively. This complementarity of the constitutive and the behavioural reflects two different sides of the characteristic of recursive self-maintenant systems [34] to deploy different processes of self-maintenance depending on environmental conditions, with constitutive and behavioural autonomy corresponding to the internal — endogenous — and external — exogneous — aspects of that adaptive capacity, respectively.

Self-regulation is central to constitutive autonomy. In biological systems, the automatic regulation of physiological functions is referred to as *homeostasis* [39, 40]: "the process of maintaining the internal milieu physiological parameters (such as temperature, pH and nutrient levels) of a biological system within the range that facilitates survival and optimal function" [41, 42]. It has been suggested [43, 44] that the autonomy of an agent is effected through a hierarchy of homeostatic self-regulatory processes, exploiting a progression of associated affective (i.e. emotional or feeling) states, ranging from basic reflexes linked to metabolic regulation, through drives and motives, and on to the emotions and feelings often linked to higher cognitive functions, similar to Damasio's hierarchy of levels of homeostatic regulation [41].

Typically, the autonomous agent is perturbed during interactions with the world with the result that the organizational dynamics have to be adjusted. This process of adjustment is exactly what is meant by homeostasis and the motives at every level of this hierarchy of homeostatic processes are effectively the drives that are required to return the agent to a state where its autonomy is no longer threatened. In the interaction with the world around it, the perturbations of the agent by the environment have no intrinsic value in their own right: they are just the stuff that happens to the agent as it goes about its business of survival. However, for the agent this stuff — these interactions and perturbations — has a perceived value in that it acts to endanger or support its autonomy. This value is conveyed through the affective aspect of these homeostatic processes and consequently the agent then attaches some value to what is an otherwise neutral world (even if it is a precarious one) [45]. This gives rise to a reciprocal coupling — and mutual dependency — of action and perception in cognition where perceptions and actions form a complementary set of environment-agent / agent-environment perturbations that are related not as extrinsic stimulus-response perceptuo-motor contingencies but as intrinsic processes that lead to the regulation of the system and autonomy preservation through emergent self-organization [46]. The processes of perception and action are mutually dependent because they are both modulated by the system — globally-determined — through downward causation [47, 33] and, together with other homeostatic processes, they give rise to the global constitutive autonomy-preserving system behaviour. This is a subtle but important point as it suggests a causal link between the processes of constitutive autonomy (*qua* self-organization) and behavioural autonomy (*qua* viable interaction with the environment). We return to this point in the next section.

Just as, from the perspective of behavioural autonomy, a cognitive agent continually deploys prospection through internal simulation to prepare to act [48, 49, 50, 51, 52], so too are the processes of constitutive autonomy prospective. This predictive self-regulation is known as *allostasis* [53, 54, 55]. Sterling notes that allostasis provides a *global* mechanism for overriding of normal homeostasis, serving the organism as a whole with the resources previously learned to be necessary to meet predicted environmental pressures [53]. Thus, allostasis differs from homeostasis in its predictive character and in its ability to anticipate and adapt to change rather than resist it. Significantly, allostasis is effected at a higher level of organization, involving greater number of sub-systems acting together in a coordinated manner with global processes modulating local ones, reflecting the character of circular causality. In contrast, mechanisms for homeostasis operate at a simpler level of negative feedback control [53, 55, 56].

Now here we finally come to the challenge. Development is commonly cast as a process of adaptation based on interaction with the environment and other agents [57, 10] and, when autonomy is considered, the focus is usually on behavioural autonomy. However, here we see the critical importance of constitutive autonomy. Development applies not only to the behavioural capacities for interaction that we have discussed above, but it applies also to the constitutive elements of the agent. Specifically, the value systems that drive development are relevant not just to the processes of behavioural autonomy but also to those of constitutive autonomy and both forms of autonomy exhibit prospection, the key attribute of cognition. We have seen how prospection is essential for effective processes of behaviour and interaction (e.g. we anticipate the need to buy groceries before cooking dinner) but it is also essential for effective constitutive processes (e.g. blood sugar levels are raised in anticipation of the demands of imminent exercise; see [53, 55] for this and other examples of predictive metabolic regulation). Furthermore, since enactive systems are operationally-closed,[1] autonomous, and self-maintaining, the constitutive processes may be the primary source of autonomy for both constitutive processes and behavioural interaction processes. Since development is normally cast in behavioural terms, i.e., in terms of an agent interacting with the world around it, not in terms of internal interaction, this presents us with a dilemma: how can the value systems and motivations that drive development support both constitutive autonomy and behavioural autonomy? In particular, how can the processes that support constitutive autonomy also give rise to behavioural autonomy and especially to the development of the agent based on interaction with its environment and other agents through social and exploratory motives?

5 Addressing the Challenge

While the aim of this article is to identify the challenge of modelling development in enactive cognitive agents, in general, and enactive cognitive robots, in particular, it would be rather unsatisfactory to finish without suggesting where possible answers might be sought. We do this now.

Any substantive answers to the questions raised above will probably be founded on an innate capacity for allostatic and homeostatic self-organization that makes sense of the agent's structural coupling with its environment in maintaining the agent's constitutive and behavioural

[1]The term *operational closure* characterizes any system that is identified by an observer to be self-contained and parametrically-coupled with its environment but not controlled by the environment. It is related to *organizational closure*, a necessary characteristic of a particular form of self-producing self-organization called *autopoiesis* [58, 59]. Technically, autopoiesis operates at the bio-chemical level, e.g., in cellular systems, but its usage has been expanded to deal with autonomous systems in general where, more correctly, it is referred to as *operational closure*. The operational closure *vs.* organizational closure terminology can be confusing because in some earlier publications, e.g. [60], Varela refers to *organizational closure* but in later works (by Maturana and Varela themselves, e.g. [61], and by others, e.g. [62]) this term was subsequently replaced in favour of *operational closure* to reflect is more general usage, with organizational closure being used to characterized an operationally-closed system that exhibits some form of self-production or self-construction [63].

Just as, from the perspective of behavioural autonomy, a cognitive agent continually deploys prospection through internal simulation to prepare to act [48, 49, 50, 51, 52], so too are the processes of constitutive autonomy prospective. This predictive self-regulation is known as *allostasis* [53, 54, 55]. Sterling notes that allostasis provides a *global* mechanism for overriding of normal homeostasis, serving the organism as a whole with the resources previously learned to be necessary to meet predicted environmental pressures [53]. Thus, allostasis differs from homeostasis in its predictive character and in its ability to anticipate and adapt to change rather than resist it. Significantly, allostasis is effected at a higher level of organization, involving greater number of sub-systems acting together in a coordinated manner with global processes modulating local ones, reflecting the character of circular causality. In contrast, mechanisms for homeostasis operate at a simpler level of negative feedback control [53, 55, 56].

Now here we finally come to the challenge. Development is commonly cast as a process of adaptation based on interaction with the environment and other agents [57, 10] and, when autonomy is considered, the focus is usually on behavioural autonomy. However, here we see the critical importance of constitutive autonomy. Development applies not only to the behavioural capacities for interaction that we have discussed above, but it applies also to the constitutive elements of the agent. Specifically, the value systems that drive development are relevant not just to the processes of behavioural autonomy but also to those of constitutive autonomy and both forms of autonomy exhibit prospection, the key attribute of cognition. We have seen how prospection is essential for effective processes of behaviour and interaction (e.g. we anticipate the need to buy groceries before cooking dinner) but it is also essential for effective constitutive processes (e.g. blood sugar levels are raised in anticipation of the demands of imminent exercise; see [53, 55] for this and other examples of predictive metabolic regulation). Furthermore, since enactive systems are operationally-closed,[1] autonomous, and self-maintaining, the constitutive processes may be the primary source of autonomy for both constitutive processes and behavioural interaction processes. Since development is normally cast in behavioural terms, i.e., in terms of an agent interacting with the world around it, not in terms of internal interaction, this presents us with a dilemma: how can the value systems and motivations that drive development support both constitutive autonomy and behavioural autonomy? In particular, how can the processes that support constitutive autonomy also give rise to behavioural autonomy and especially to the development of the agent based on interaction with its environment and other agents through social and exploratory motives?

5 Addressing the Challenge

While the aim of this article is to identify the challenge of modelling development in enactive cognitive agents, in general, and enactive cognitive robots, in particular, it would be rather unsatisfactory to finish without suggesting where possible answers might be sought. We do this now.

Any substantive answers to the questions raised above will probably be founded on an innate capacity for allostatic and homeostatic self-organization that makes sense of the agent's structural coupling with its environment in maintaining the agent's constitutive and behavioural

[1]The term *operational closure* characterizes any system that is identified by an observer to be self-contained and parametrically-coupled with its environment but not controlled by the environment. It is related to *organizational closure*, a necessary characteristic of a particular form of self-producing self-organization called *autopoiesis* [58, 59]. Technically, autopoiesis operates at the bio-chemical level, e.g., in cellular systems, but its usage has been expanded to deal with autonomous systems in general where, more correctly, it is referred to as *operational closure*. The operational closure *vs.* organizational closure terminology can be confusing because in some earlier publications, e.g. [60], Varela refers to *organizational closure* but in later works (by Maturana and Varela themselves, e.g. [61], and by others, e.g. [62]) this term was subsequently replaced in favour of *operational closure* to reflect is more general usage, with organizational closure being used to characterized an operationally-closed system that exhibits some form of self-production or self-construction [63].

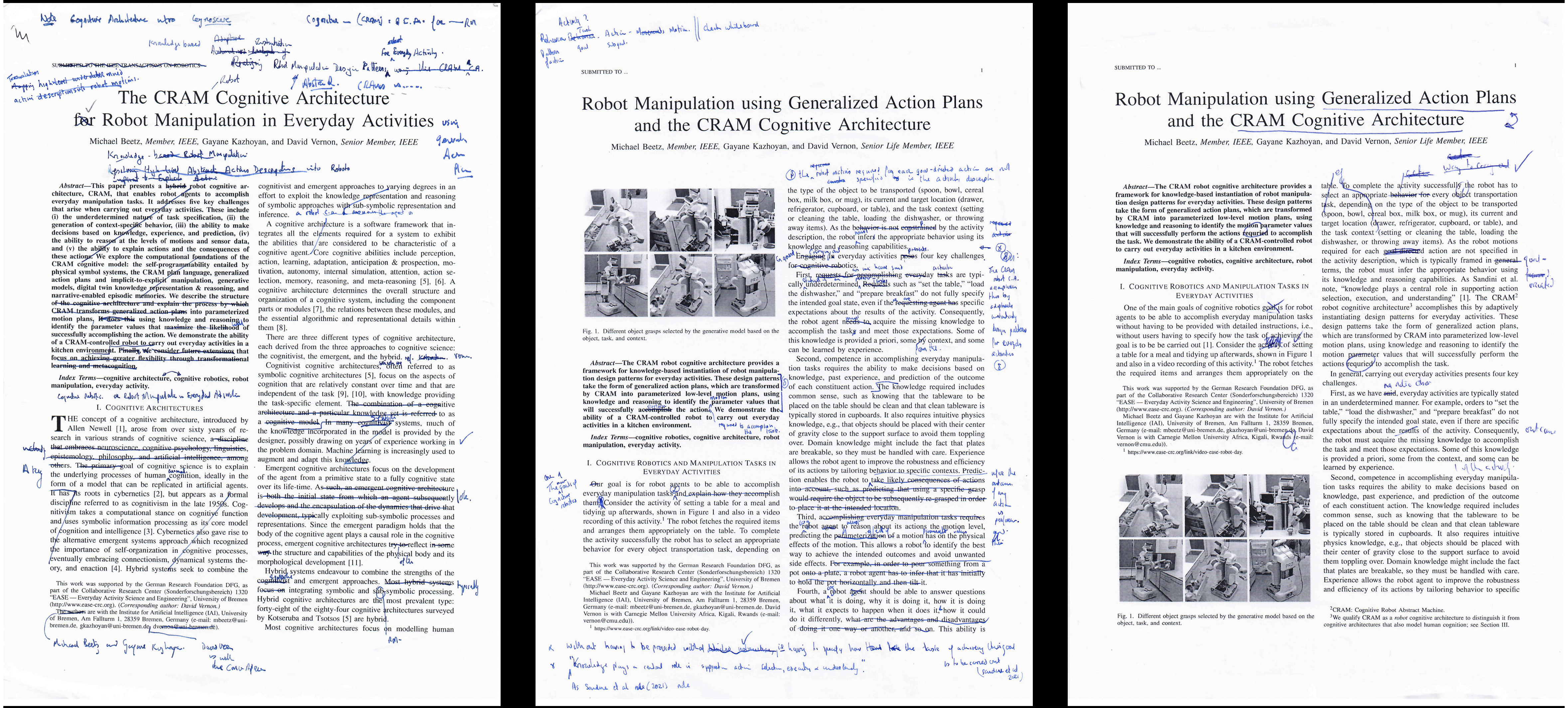
autonomy. It will trade the traditional emphasis on exteroception for interoception and internal action, much as John Weng has suggested with his self-affecting self-effecting models of autonomous mental development [27], and it will leverage the innate phylogenetic capacities for modulating its interactions as set out in Section 2 and 3 above.

In recent work [64, 65], Anil Seth discusses the importance of prediction in cognition, suggesting, as others have done [66], that the brain engages in continual predictive inference of the causes of sensory perturbations, i.e. the predictive perception of sensorimotor contingencies. In this, he develops the concept of *predictive processing* whereby the brain infers the most likely causes of its sensory inputs by minimizing the difference between sensory signals and signals derived from continuously updated predictive models. However, his central thesis is that this process derives less from classical exteroception than from interoception. This interoception is based on cybernetic principles. They assert that "the purpose of cognition (including perception and action) is to maintain the homeostasis of essential variables and of internal organization ... [and] ... perception emerges as a *consequence* of a more fundamental imperative towards organizational homeostasis, and not as a stage in some process of internal world-model construction" [65], p. 8. Viewed in this enactive light, cognitive agents adapt — develop — to ensure continued existence by successfully responding to environmental perturbations so as to maintain their internal organization. This neatly links constitutive autonomy to behavioural autonomy and suggests how behavioural autonomy derives from constitutive autonomy. The question then is: how is this accomplished.

Seth builds on Karl Friston's *Free Energy Principle* [67, 68], according to which "organisms obey a fundamental imperative towards the avoidance of (information-theoretically) surprising events, according to which they must minimize in the long-run average surprise of sensory states, since surprising sensory states are (in the long run) likely to reflect conditions incompatible with continued existence" [65], p. 2. Seth suggests that *active inference*, an extension of predictive processing, operates to suppress the interoceptive prediction errors not only by updating the generative model that gave rise to the predictions but by internal action, translating the predictions into reference points for autonomic regulatory processes, e.g. physiological organizational homeostasis.[2] He notes that attention can then be viewed as a way of balancing active inference and model update, referred to as precision weighting. He reinforces the idea that "an organism should maintain well-adapted predictive models of its own physical body ... and of its internal physiological condition" [64], p. 567. Active inference can act both to selectively sample sensory data to conform to current predictions and to seek evidence that contradicts current predictions or disambiguate multiple competing hypotheses. This leverages "the capacity of predictive models to encode counterfactual relations linking potential (but not necessarily executed) actions to their expected consequences". It implies model comparison and selection, much as the HAMMER architecture for internal simulation in cognitive robotics does with its multiple forward and inverse models [73, 74], and not just the optimization of the parameters of a single model.

It remains, however, to find a way of minimizing the information-theoretic surprisal associated with the agent's internal organization. One possibility is an information-theoretic technique introduced by Robert Ulanowicz [75, 76, 77]. Although the model was originally intended to model the growth and development of ecosystems, it is quite general and has already been used to characterize the emergence and development of beliefs in human cognition [78] and is also being investigated a value system to drive the development of joint episodic-procedural memory networks [79]. Modelling the system as a flow network, growth and development are framed

[2] Seth views allostasis as "the process of achieving homeostasis" [65], p.7, emphasizing its roots in cybernetics, in general, and the ultrastability of Ashby's homeostat [69, 70, 71, 72], in particular. He notes that the fundamental cybernetic principle is for systems to ensure their continued existence by successfully responding to environmental perturbations so as to maintain their internal organization. He goes further, stating that "The purpose of cognition (including perception and action) is to maintain the homeostasis of essential variables and of internal organization (ultrastability)" [65], p. 8.



SUBMITTED TO THE IEEE TRANSACTIONS ON ROBOTICS

The CRAM Cognitive Architecture for Robot Manipulation in Everyday Activities

Michael Beetz, *Member, IEEE*, Gayane Kazhoyan, and David Vernon, *Senior Member, IEEE*

Abstract—This paper presents a hybrid robot cognitive architecture, CRAM, that enables robot agents to accomplish everyday manipulation tasks. It addresses five key challenges that arise when carrying out everyday activities. These include (i) the underdetermined nature of task specification, (ii) the generation of context-specific behavior, (iii) the ability to make decisions based on knowledge, experience, and prediction, (iv) the ability to reason at the levels of motions and sensor data, and (v) the ability to explain actions and the consequences of these actions. We explore the computational foundations of the CRAM cognitive model: the self-programmability entailed by physical symbol systems, the CRAM plan language, generalized action plans and implicit-to-explicit manipulation, generative models, digital twin knowledge representation & reasoning, and narrative-enabled episodic memories. We describe the structure of the cognitive architecture and explain the process by which CRAM transforms generalized action plans into parameterized motion plans. It does this using knowledge and reasoning to identify the parameter values that maximize the likelihood of successfully accomplishing the action. We demonstrate the ability of a CRAM-controlled robot to carry out everyday activities in a kitchen environment. Finally, we consider future extensions that focus on achieving greater flexibility through transformational learning and metacognition.

***Index Terms*—cognitive architecture, cognitive robotics, robot manipulation, everyday activity.**

I. Cognitive Architectures

THE concept of a cognitive architecture, introduced by Allen Newell [1], arose from over sixty years of research in various strands of cognitive science, a discipline that embraces neuroscience, cognitive psychology, linguistics, epistemology, philosophy, and artificial intelligence, among others. The primary goal of cognitive science is to explain the underlying processes of human cognition, ideally in the form of a model that can be replicated in artificial agents. It has its roots in cybernetics [2], but appears as a formal discipline referred to as cognitivism in the late 1950s. Cognitivism takes a computational stance on cognitive function and uses symbolic information processing as its core model of cognition and intelligence [3]. Cybernetics also gave rise to the alternative emergent systems approach which recognized the importance of self-organization in cognitive processes, eventually embracing connectionism, dynamical systems theory, and enaction [4]. Hybrid systems seek to combine the cognitivist and emergent approaches to varying degrees in an effort to exploit the knowledge representation and reasoning of symbolic approaches with sub-symbolic representation and inference.

A cognitive architecture is a software framework that integrates all the elements required for a system to exhibit the abilities that are considered to be characteristic of a cognitive agent. Core cognitive abilities include perception, action, learning, adaptation, anticipation & prospection, motivation, autonomy, internal simulation, attention, action selection, memory, reasoning, and meta-reasoning [5], [6]. A cognitive architecture determines the overall structure and organization of a cognitive system, including the component parts or modules [7], the relations between these modules, and the essential algorithmic and representational details within them [8].

There are three different types of cognitive architecture, each derived from the three approaches to cognitive science: the cognitivist, the emergent, and the hybrid.

Cognitivist cognitive architectures, often referred to as symbolic cognitive architectures [5], focus on the aspects of cognition that are relatively constant over time and are independent of the task [9], [10], with knowledge providing the task-specific element. The combination of a cognitive architecture and a particular knowledge set is referred to as a cognitive model. In many cognitivist systems, much of the knowledge incorporated in the model is provided by the designer, possibly drawing on years of experience working in the problem domain. Machine learning is increasingly used to augment and adapt this knowledge.

Emergent cognitive architectures focus on the development of the agent from a primitive state to a fully cognitive state over its life-time. As such, an emergent cognitive architecture is both the initial state from which an agent subsequently develops and the encapsulation of the dynamics that drive that development, typically exploiting sub-symbolic processes and representations. Since the emergent paradigm holds that the body of the cognitive agent plays a causal role in the cognitive process, emergent cognitive architectures try to reflect in some way the structure and capabilities of the physical body and its morphological development [11].

Hybrid systems endeavour to combine the strengths of the cognitivist and emergent approaches. Most hybrid systems focus on integrating symbolic and sub-symbolic processing. Hybrid cognitive architectures are the most prevalent type: forty-eight of the eighty-four cognitive architectures surveyed by Kotseruba and Tsotsos [5] are hybrid.

Most cognitive architectures focus on modelling human

This work was supported by the German Research Foundation DFG, as part of the Collaborative Research Center (Sonderforschungsbereich) 1320 "EASE — Everyday Activity Science and Engineering", University of Bremen (http://www.ease-crc.org). (Corresponding author: David Vernon.)

The authors are with the Institute for Artificial Intelligence (IAI), University of Bremen, Am Fallturm 1, 28359 Bremen, Germany (e-mail: mbeetz@uni-bremen.de, gkazhoyan@uni-bremen.de, dvernon@uni-bremen.de).

SUBMITTED TO ... 1

Robot Manipulation using Generalized Action Plans and the CRAM Cognitive Architecture

Michael Beetz, *Member, IEEE*, Gayane Kazhoyan, and David Vernon, *Senior Life Member, IEEE*

Fig. 1. Different object grasps selected by the generative model based on the object, task, and context.

Abstract—The CRAM cognitive architecture provides a framework for knowledge-based instantiation of robot manipulation design patterns for everyday activities. These design patterns take the form of generalized action plans, which are transformed by CRAM into parameterized low-level motion plans, using knowledge and reasoning to identify the parameter values that will successfully accomplish the action. We demonstrate the ability of a CRAM-controlled robot to carry out everyday activities in a kitchen environment.

***Index Terms*—cognitive robotics, cognitive architecture, robot manipulation, everyday activity.**

I. Cognitive Robotics and Manipulation Tasks in Everyday Activities

Our goal is for robot agents to be able to accomplish everyday manipulation tasks and explain how they accomplish them. Consider the activity of setting a table for a meal and tidying up afterwards, shown in Figure 1 and also in a video recording of this activity.[1] The robot fetches the required items and arranges them appropriately on the table. To complete the activity successfully the robot has to select an appropriate behavior for every object transportation task, depending on the type of the object to be transported (spoon, bowl, cereal box, milk box, or mug), its current and target location (drawer, refrigerator, cupboard, or table), and the task context (setting or cleaning the table, loading the dishwasher, or throwing away items). As the behavior is not constrained by the activity description, the robot infers the appropriate behavior using its knowledge and reasoning capabilities.

Engaging in everyday activities poses four key challenges for cognitive robotics.

First, requests for accomplishing everyday tasks are typically underdetermined. Requests such as "set the table," "load the dishwasher," and "prepare breakfast" do not fully specify the intended goal state, even if the requesting agent has specific expectations about the results of the activity. Consequently, the robot agent needs to acquire the missing knowledge to accomplish the tasks and meet those expectations. Some of this knowledge is provided a priori, some by context, and some can be learned by experience.

Second, competence in accomplishing everyday manipulation tasks requires the ability to make decisions based on knowledge, past experience, and prediction of the outcome of each constituent action. The knowledge required includes common sense, such as knowing that the tableware to be placed on the table should be clean and that clean tableware is typically stored in cupboards. It also requires intuitive physics knowledge, e.g., that objects should be placed with their center of gravity close to the support surface to avoid them toppling over. Domain knowledge might include the fact that plates are breakable, so they must be handled with care. Experience allows the robot agent to improve the robustness and efficiency of its actions by tailoring behavior to specific contexts. Prediction enables the robot to take likely consequences of actions into account, such as predicting that using a specific grasp would require the object to be subsequently re-grasped in order to place it at the intended location.

Third, accomplishing everyday manipulation tasks requires the robot agent to reason about its actions at the motion level, predicting the parameterization of a motion has on the physical effects of the motion. This allows a robot to identify the best way to achieve the intended outcomes and avoid unwanted side effects. For example, in order to pour something from a pot onto a plate, a robot agent has to infer that it has initially to hold the pot horizontally and then tilt it.

Fourth, a robot agent should be able to answer questions about what it is doing, why it is doing it, how it is doing it, what it expects to happen when it does it, how it could do it differently, what are the advantages and disadvantages of doing it one way or another, and so on. This ability is

This work was supported by the German Research Foundation DFG, as part of the Collaborative Research Center (Sonderforschungsbereich) 1320 "EASE — Everyday Activity Science and Engineering", University of Bremen (http://www.ease-crc.org). (Corresponding author: David Vernon.)

Michael Beetz and Gayane Kazhoyan are with the Institute for Artificial Intelligence (IAI), University of Bremen, Am Fallturm 1, 28359 Bremen, Germany (e-mail: mbeetz@uni-bremen.de, gkazhoyan@uni-bremen.de). David Vernon is with Carnegie Mellon University Africa, Kigali, Rwanda (e-mail: vernon@cmu.edu)).

[1] https://www.ease-crc.org/link/video-ease-robot-day.

SUBMITTED TO ... 1

Robot Manipulation using Generalized Action Plans and the CRAM Cognitive Architecture

Michael Beetz, *Member, IEEE*, Gayane Kazhoyan, and David Vernon, *Senior Life Member, IEEE*

Abstract—The CRAM robot cognitive architecture provides a framework for knowledge-based instantiation of robot manipulation design patterns for everyday activities. These design patterns take the form of generalized action plans, which are transformed by CRAM into parameterized low-level motion plans, using knowledge and reasoning to identify the motion parameter values that will successfully perform the actions required to accomplish the task. We demonstrate the ability of a CRAM-controlled robot to carry out everyday activities in a kitchen environment.

***Index Terms*—cognitive robotics, cognitive architecture, robot manipulation, everyday activity.**

I. Cognitive Robotics and Manipulation Tasks in Everyday Activities

One of the main goals of cognitive robotics goal is for robot agents to be able to accomplish everyday manipulation tasks without having to be provided with detailed instructions, i.e., without users having to specify how the task of achieving the goal is to be carried out [1]. Consider the activity of setting a table for a meal and tidying up afterwards, shown in Figure 1 and also in a video recording of this activity.[1] The robot fetches the required items and arranges them appropriately on the table. To complete the activity successfully the robot has to select an appropriate behavior for every object transportation task, depending on the type of the object to be transported (spoon, bowl, cereal box, milk box, or mug), its current and target location (drawer, refrigerator, cupboard, or table), and the task context (setting or cleaning the table, loading the dishwasher, or throwing away items). As the robot motions required for each goal-directed action are not specified in the activity description, which is typically framed in general terms, the robot must infer the appropriate behavior using its knowledge and reasoning capabilities. As Sandini et al. note, "knowledge plays a central role in supporting action selection, execution, and understanding" [1]. The CRAM[2] robot cognitive architecture[3] accomplishes this by adaptively instantiating design patterns for everyday activities. These design patterns take the form of generalized action plans, which are transformed by CRAM into parameterized low-level motion plans, using knowledge and reasoning to identify the robot motion parameter values that will successfully perform the actions required to accomplish the task.

In general, carrying out everyday activities presents four key challenges.

First, as we have said, everyday activities are typically stated in an underdetermined manner. For example, orders to "set the table," "load the dishwasher," and "prepare breakfast" do not fully specify the intended goal state, even if there are specific expectations about the results of the activity. Consequently, the robot must acquire the missing knowledge to accomplish the task and meet those expectations. Some of this knowledge is provided a priori, some from the context, and some can be learned by experience.

Second, competence in accomplishing everyday manipulation tasks requires the ability to make decisions based on knowledge, past experience, and prediction of the outcome of each constituent action. The knowledge required includes common sense, such as knowing that the tableware to be placed on the table should be clean and that clean tableware is typically stored in cupboards. It also requires intuitive physics knowledge, e.g., that objects should be placed with their center of gravity close to the support surface to avoid them toppling over. Domain knowledge might include the fact that plates are breakable, so they must be handled with care. Experience allows the robot agent to improve the robustness and efficiency of its actions by tailoring behavior to specific

Fig. 1. Different object grasps selected by the generative model based on the object, task, and context.

This work was supported by the German Research Foundation DFG, as part of the Collaborative Research Center (Sonderforschungsbereich) 1320 "EASE — Everyday Activity Science and Engineering", University of Bremen (http://www.ease-crc.org). (Corresponding author: David Vernon.)

Michael Beetz and Gayane Kazhoyan are with the Institute for Artificial Intelligence (IAI), University of Bremen, Am Fallturm 1, 28359 Bremen, Germany (e-mail: mbeetz@uni-bremen.de, gkazhoyan@uni-bremen.de). David Vernon is with Carnegie Mellon University Africa, Kigali, Rwanda (e-mail: vernon@cmu.edu)).

[1] https://www.ease-crc.org/link/video-ease-robot-day.

[2] CRAM: Cognitive Robot Abstract Machine.

[3] We qualify CRAM as a *robot* cognitive architecture to distinguish it from cognitive architectures that also model human cognition; see Section III.

Writing a Research Proposal

The Research Question

- A Ph.D. or M.Sc. thesis typically asks and attempts to answer a research question
- A good thesis answers it convincingly
- This research question is typically one element of a research proposal

The Goal

- Identify the problem that you wish to solve
- State clearly why this problem is important
- Highlight the reasons why finding a solution is challenging
- The research question asks: how can we overcome these challenges?
- A research thesis provides the answer, in full or in part

Not Just an Interesting Idea

- Not enough just to ask the research question
- Demonstrate you that you understand the extent of the challenge
- Provide a brief survey of the approaches that others have taken
- Demonstrate that you are conversant with the field of research
- Ideally, identify a plausible solution strategy

The Heilmeier Catechism

DARPA operates on the principle that generating big rewards requires taking big risks. But how does the Agency determine what risks are worth taking?

George H. Heilmeier, a former DARPA director (1975-1977), crafted a set of questions known as the "Heilmeier Catechism" to help Agency officials think through and evaluate proposed research programs.

- What are you trying to do? Articulate your objectives using absolutely no jargon.
- How is it done today, and what are the limits of current practice?
- What is new in your approach and why do you think it will be successful?
- Who cares? If you are successful, what difference will it make?
- What are the risks?
- How much will it cost?
- How long will it take?
- What are the mid-term and final “exams” to check for success?

https://www.darpa.mil/work-with-us/heilmeier-catechism

The Heilmeier Catechism

1. "What are you trying to do? Articulate your objectives using absolutely no jargon."

2. "How is it done today, and what are the limits of current practice?"

3. "What is new in your approach and why do you think it will be successful?"

4. "Who cares? If you are successful, what difference will it make?"

5. "What are the mid-term and final "exams" to check for success?"

Proposal Structure

- Goals of the Research Project
- Review of Current Approaches
- Proposed Solution
- Novelty and Significance
- Anticipated Results
- Evaluation Strategy and Metrics of Success

(Adapt these titles to include some text to make them specific to your research project)

Critical Analysis

- A good research proposal requires
  - Some preliminary research
  - A good deal of critical analysis
- It is not just a suggestion for an avenue of enquiry, no matter how interesting or exciting it might appear

Writing a Literature Survey

Survey the Field

- Hallmark of good research: an understanding of how others have approached the problem you are tackling

- You need to develop a deep understanding of

  – The theoretical basis of their techniques
  – The assumptions they make
  – The tools and methodologies they use

Survey the Field

- The literature survey:
  - The nature of a problem
  - The spectrum of possible approaches

- A well-structured synthesis that
  - Collects all the relevant ideas
  - Organizes them
  - Presents each of them in turn
  - Highlights their strengths and weaknesses

Survey the Field

- It will require many attempts and many re-writes
- Start with your own short summary of each paper you have read
- Then try to organize the ideas
  - Identify different classes of topics
  - Relating them together in the form of a taxonomy, or hierarchical classification tree
- This taxonomy then provides you with a way to structure the literature survey
  - Typically by doing a breadth-first traversal of the taxonomy tree
  - Using the material in the cited papers as examples

Survey the Field

Why is the literature survey so important?

1. It provides the essential background for your thesis
   - Critically appraise the state of the art in the field you are conducting research
   - Thereby establish your mastery of the field

Survey the Field

Why is the literature survey so important?

2. It identifies the gap in existing knowledge that is encapsulated in your research question

   – Make a compelling argument that the gap actually exists

   – Without this, you could be wasting your time by trying to fill a non-existent gap or a gap that the research community has decided is not worth filling

Survey the Field

Why is the literature survey so important?

3. It helps you identify the tools you will need to use to answer your research question:

   - Mathematical
   - Analytical
   - Software
   - Data collection
   - Data presentation
   - Style of argument

Writing Scientific Papers

The Importance of Writing Papers

- The ultimate test of M.Sc. or Ph.D. research: is it worthy of publication in a journal or in the proceedings of a conference?
- Consequently, research papers are the primary output of a research degree, not software
- Software may be needed to validate the ideas but the contribution to knowledge is the idea itself, encapsulated in a paper
- It should be possible to re-write or re-generate the software based on the information contained in the paper

Know Your Reader

- Assume the person reading your paper is intelligent but not knowledgeable
- Assume the person reading your paper misunderstands things willfully and easily
  - so make sure the argument is really clear
- Make it easy for them to say:

  'Nice idea, good model, great validation; yes, the community would like to know about this'.
- If someone doesn't understand your paper ...
  - Assume it is your fault, not theirs
  - Find out where they got lost and improve it

Structure

- Assist the reader by making your points clearly and in a logical order
- Breaking up the paper into a linear sequence of messages which follow naturally one from the other and which lead to an interesting conclusion
- Begin with a statement describing a claim or hypothesis
- Then provide an argument to support that claim or hypothesis
  - Provide the context for the claim (e.g., other people's work and alternative approaches)
  - State its relevance or importance
  - Offer a model of the subject you are investigating
  - Provide some theoretical or empirical evidence that the model is valid
  - Provide an assessment of how well it works.

The Thread of an Argument

- Construct your message incrementally, piece by piece, building on ideas you have already developed, typically in the previous sentence
- If you need to build on ideas that were introduced in the previous paragraph or a few pages back, you should provide the reader with a short reminder of what these ideas are
- We call this type of incrementally-constructed message the thread of an argument
- You need to keep the thread as cohesive as possible
  - Thread sequentially and logically from one idea to another
  - Don't make repeated reference to ideas that were introduced much earlier (or, worse, not at all)
  - Don't introduce a new idea without warning
  - Don't try to weave several threads together

Don't Begin at the Beginning

- Papers are intended to be read from beginning to end, and the argument should flow linearly from beginning to end
- However, it isn't always best to write the paper in that order
- An alternative approach:
  - Start by describing the technique
  - Write the introduction later, once you've established the core message
  - Possibly after you have drawn your conclusions
- The abstract should be written last
(or first, and assume it will have to be completely rewritten when you have finished)

Do Your Best and Then Improve It

- Be prepared to write, and re-write, many times
  - It can take up to ten attempts to get a good first draft of a paper
- Once you have a good draft, ask other people to read it
- Do not ask people to read early drafts
  - Give them your best and thank them for their time
  - Under no circumstances use your supervisor as a proof-reader to correct mistakes or get hints on how to improve structure before you are certain that the paper can't be improved!
  - You will probably be wrong, but that should be your goal.

Provide Results

- Quantitative vs. qualitative results
- Quantitative results are more convincing
  - Try to identify a metric for the performance of your technique or system
  - Measure how well it performs using this metric
  - One metric is good, more than one is better
- Compare your system to others using the same metric
- Sometimes this will mean re-implementing other people's work
  - User open-source versions of standard approaches
  - These provide an excellent benchmark against which to judge the value of your own contribution

Citations

- Everything you say must be substantiated
  - Provide a citation of a publication which corroborates or supports the statement
  - Provide either theoretical or empirical evidence to support it
- Citations are references to published material which has been subject to some form of peer review
  - The statements or claims have been judged to be legitimate by a group of people, not just the author
  - This elevates the statement from being mere personal opinion of the author to some level of mutually-agreed knowledge

Citations

Avoid unsubstantiated claims or statements in your thesis

- Either provide a citation or provide evidence
- If you can't provide either, then provide a compelling argument in support of your claim

Citations

Other reasons for citing literature

- Acknowledge the source of your ideas: even the most original thoughts are based on the work of others
- Demonstrate that you have done the research and are familiar with the literature in your area.

Citations

Don’t cite a reference unless you have read it!

- You should never just copy the citations you find in someone else’s paper to support your statement without first reading it to make sure it says what the author claims it says

Quotations

- Never copy other people's writing, even if you change it slightly
  - You must re-express it in your own words
- If you must use someone else's writing, put it in quotation marks "…"
  - Cite the source of the quotation.
  - If the grammar or spelling in the quotation is wrong, don't correct it.
    - Quote it exactly as it is but put [*sic*] after the error;
      e.g., "How r [*sic*] you?".

Quotations

- In general, you should change nothing in the quotation
- There are two exceptions
  - Replace a contiguous group of words with ellipsis
    - For example, instead of "computational attention, in its most general form, is a pre-requisite for action selection" you could write "computational attention … is a pre-requisite for action selection"
  - Insert a word to help its comprehensibility
    - Insert the word in square brackets [and]
    - For example, instead of "the lowest level is the most difficult to implement" you could write "the lowest level [of the stack] is the most difficult to implement"

Reviews

Journal papers and most conference papers are subject to peer review

- Between two and four referees read the paper and provide a critical assessment of its contents
- Reviewers are normally experts in the field
- You need to convince them that there is merit in your work
- This is hard
  - They are very busy people
  - They don't have the patience / motivation to read poorly-written text
  - Make it easy for them.

Reviews

There is a strong possibility of rejection

- Nobody likes rejections
- If the paper is rejected, accept the rejection gracefully and learn from it
- Understand why it was rejected
  - Was it rejected because you explained things badly?
  - Was it because there was a flaw in your argument?
  - Was there something wrong with the way you stated the problem
  - Was there something wrong with your theoretical development
  - Was your model wrong?
  - Did you provide enough evidence of the validity of the model, e.g., by providing quantitative tests?
  - How did you establish the robustness, generality, or limitations of the technique?

Reviews

There is a strong possibility of rejection

- Review your paper with these criteria in mind before submitting it

Choose Your Forum Carefully

- Journals and conferences have different standards
  - Journals usually require a more substantial contribution to knowledge
  - This is not always the case: it can be just as difficult to publish in some top-flight conferences as it is to publish in a journal
- Choose your conference and journal carefully
  - Make sure there is a good match between the subject matter of a journal or conference and the topic of your paper

Try, Try, Try again

- If your paper is rejected, don’t give up
- Take on board the reviewers’ comments and improve the paper
- Then submit it somewhere else

M.Sc. and Ph.D. Dissertations

Dissertation or Thesis?

- "Dissertation n. Detailed discourse on a subject, esp. as submitted for a higher degree in university"

- "Thesis n. (pl. ~es pr. –ez). Proposition to be maintained or proved; dissertation, esp. by candidate for degree"

  Oxford English Dictionary (OED)

Length

- Ph.D. Dissertation
  - 60,000 to 100,000+ words
  - Six to eight chapters
  - Bibliography and appendices

- M.Sc. thesis
  - 60% to 75% of a Ph.D. thesis

- Typical upper bounds, not targets

Structure

- Try to achieve modularity and independence amongst your chapters and sections
- Remember you are trying to convey a convincing message to the reader
- Use link sentences or paragraphs
  - At the end of a chapter, for example, remind the reader of the important messages, tell her or him why they are important, and then say what you need to look at next, and why, in order to continue with the “story”.

Title Page

Specific title of the thesis (*e.g.* “Multi-stage Learning in Biomimetic Search and Rescue Robots”)
General Title (*i.e.* “Final Year Project Report”)
Degree (*e.g.* Ph.D. or M.Sc.)
Author (name and student identification number)
Institution (*i.e.* Etisalat University College)
Supervisor
Date

Certification of original work & Signature

Abstract / Summary

What is the subject matter of the thesis: what did you do?
Motivation: why is it important?
Significance: what contribution does the thesis make?
The abstract should be approximately 200 words long. It normally takes at least ten revisions to achieve a good abstract.
The abstract should be written after the thesis has been completed.

Table of Contents

List of Figures

Acknowledgements

Help from friends, colleagues, and staff.
Support from Etisalat
Support from Parents, etc

Chapter 1. **Introduction & Overview**
Chapter 2. **Literature Survey**
Chapter 3. **Theoretical Foundations: Background Material**
Chapter 4. **Formal Model: Theoretical Development and/or system specification (use additional chapters if necessary)**
Chapter 5. **Design:Algorithmic Considerations**
Chapter 6. **Implementation Issues**
Chapter 7. **Evaluation**
Chapter 8. **Discussion & Critical Appraisal**

References
Appendices

- Key Software listings
- Mechanical schematics
- Mathematical proofs

Standards

- Ph.D. research: sufficient quality and depth to allow you to write a paper that would be accepted for publication in a journal

- M.Sc. Research: should have a good chance of being accepted for a relevant conference or workshop

- It is very easy to pick a research goal that is too ambitious

  – a Ph.D. degree is not a Nobel Prize

  – Your work has to be good; it doesn’t have to be revolutionary or world-changing

Standards

- Your thesis must clearly demonstrate your ability to
  - Assimilate
  - Synthesize
  - Critically appraise
- It is extremely important to assess your own work critically, i.e., with objectivity and with a view to seeing how it could be improved

Standards

The exercise of critical appraisal is different from the testing processes of verification, validation, and evaluation, which refer to the functionality of the system you have designed

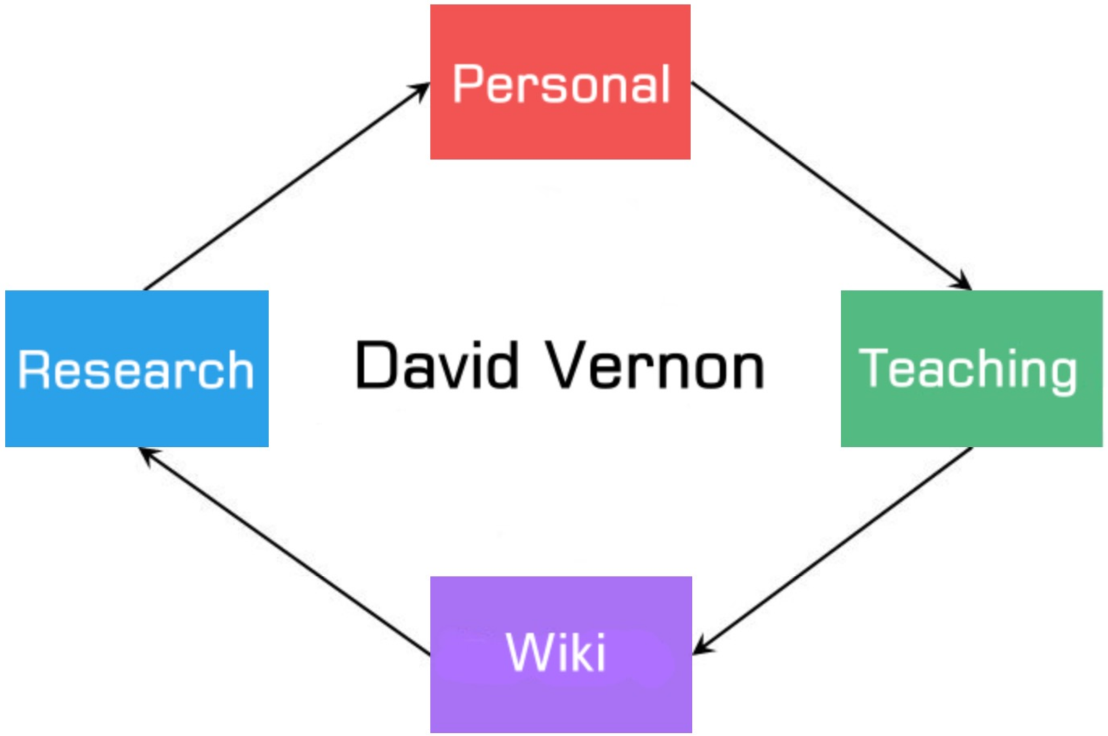
- Overall objectives
- Methodologies
- Findings of the dissertation
- Insights

7. Looking Forward

Congratulations on having completed the guide. At this point, you might be wondering if all this research is worth it. Why bother? Why do all this work? Here's why. Along the path of a research degree, you grow. You become able to do things – hard things – that you could only dream of doing before: developing a new model or algorithm of your own, learning how to master a new technique, seeing simplicity in a complex equation, having and being able to convince others of your own view on an issue. But these are the little rewards that accompany the process. The big reward comes after the degree is complete and after the papers have been published. This is when you realize that you have changed and that you can now tackle more or less any problem, with complete confidence. The unknown becomes a challenge and the reward is success. This success stays with you for the rest of your professional life.


45
min


THANK YOU!



david@vernon.eu
www.vernon.eu