"AI and the Political Philosophy of the Future: Smarter Planet or Wiser Earth?"

Keynote Presentation for the 12th NIC.br Annual Workshop on Survey Methodology
São Paulo, Brazil, August 29th, 2022

Gray Cox

Professor at College of the Atlantic
Clerk of the Quaker Institute for the Future
gray@coa.edu, #207-460-1163

1

Outline and Goals of Talk

1. **Frame key problems for AI** and our future
2. Distinguish **Two kinds of reasoning** -- monological inference vs. dialogical negotiation – to help us address those problems
3. Describe **Two models of AI** – the Turing Machine vs. the Turing Child
4. Explain **7 Strategies for developing Turing Child systems** in programming and collaboratively solving our problems

2

A Musical Prelude:

I'm gonna slow right down,
so I can get there sooner.
I'm gonna slow right down,
so I can get there today.
I'm gonna slow right down,
maybe even come to a full stop.
Maybe if I come to a full stop
I'm gonna get there right away.

Song available at https://graycox.bandcamp.com/track/im-gonna-slow-right-down

3

Part 1.) **Framing the problems** for AI and the future of ethics, policy and political philosophy – Smarter Planet vs. Wiser Earth?

4

"Civilization is not an incurable disease. But we should always remember that the English people are currently afflicted by it."

– M. K. Gandhi, ***Hind Swaraj or Indian Home Rule***

5

Our global civilization is structured by ways of reasoning in economics, governance, technology and morality that threaten our species with:

1. ecological collapse,
2. pervasive injustice & the threat of mutually assured destruction,
3. domination by super-human machine intelligence and/or foolishness
4. moral relativism and the annihilation of meaning for human life

Imagine an alien anthropologist from Alpha Centauri arriving on Earth and observing all this. Her first note home to her advisor?

"A species which imposes such radical existential threats upon itself --- what are they thinking???!!"

The alien graduate advisor's likely reply might be:

"Clearly their dominant reasoning strategies are, in a profound sense, irrational. The central research question is: HOW are they thinking?"

6

The Vision of a World of Algorithmic Intelligence:

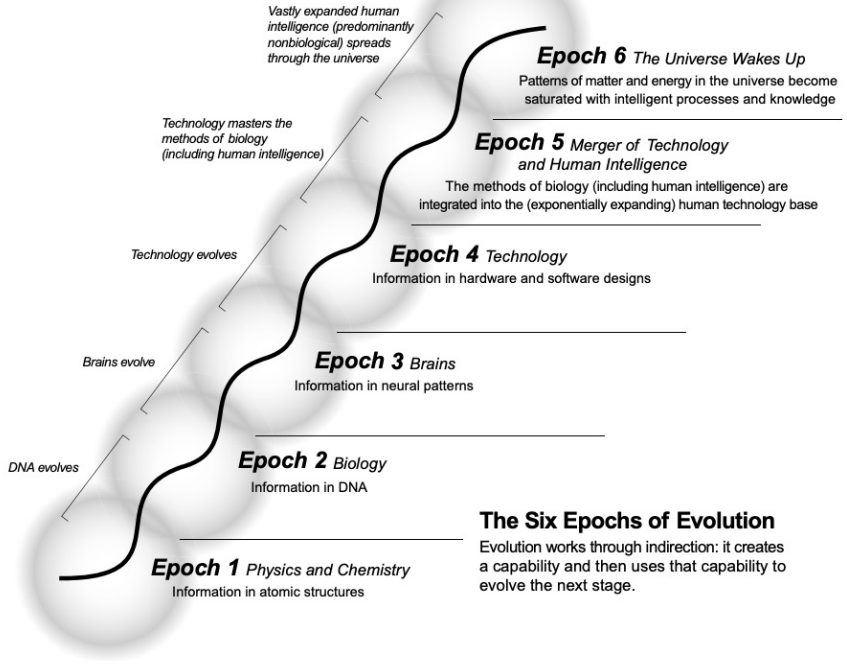
Quote from IBM Homepage, **July 7, 2009:**

"The Smarter Planet"

"Bit by bit, our planet is getting smarter. By this, we mean the systems that run the way we live and work as a society.

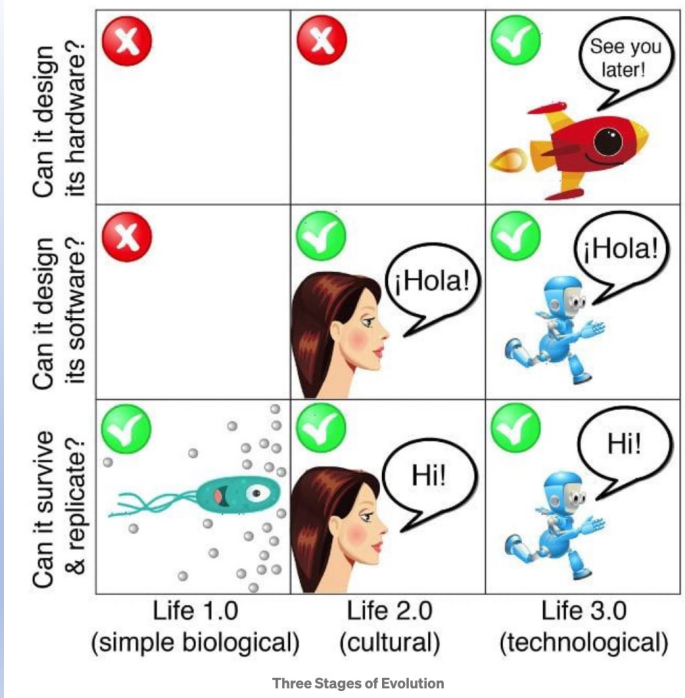
Why now?
Because the systems of our planet are increasingly**:**
Instrumented – more than a billion transistors per human, each one costing one ten-millionth of a cent
Interconnected -- With a trillion networked things
Intelligent – AI pervading the systems, monitoring and managing them.
And this is all **Inevitable".**

7



From Ray Kurzweil's
The Singularity is Near, pp. 15

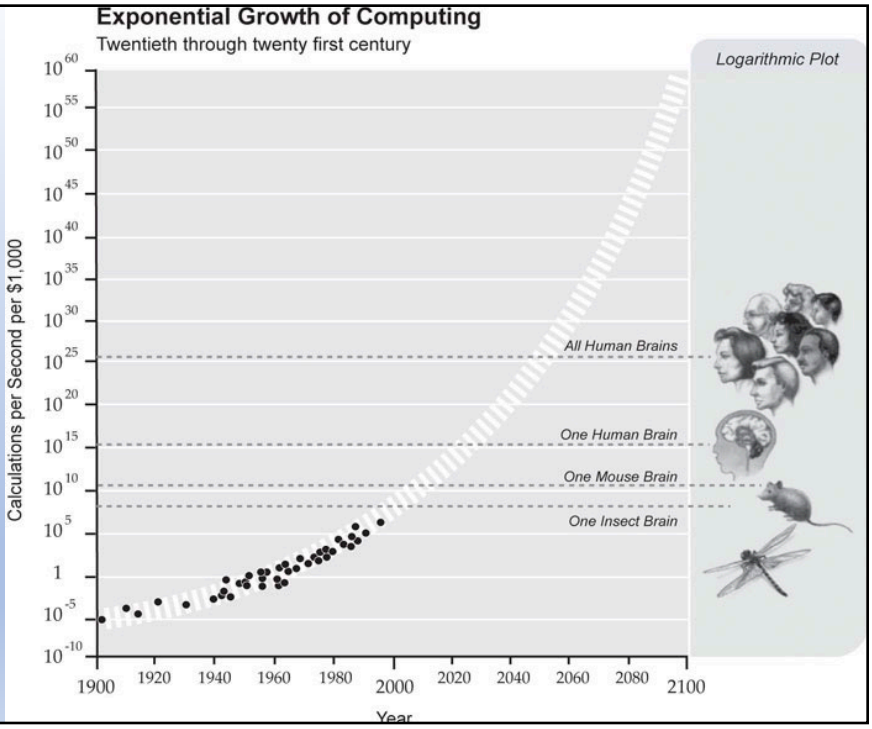
8



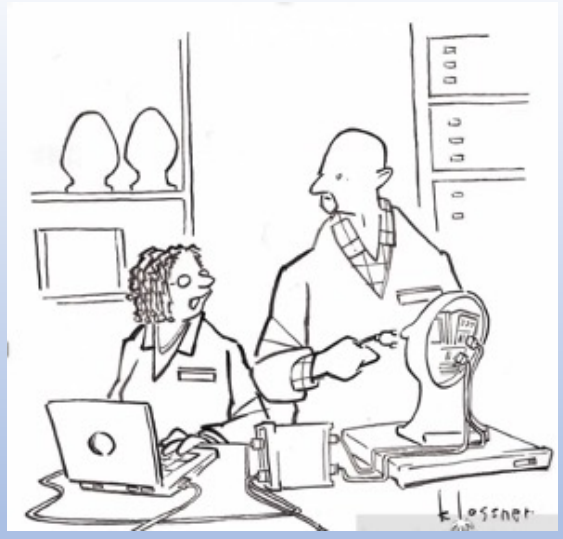
From Max Tegmark, *Life 3.0,* page 26

9

From Ray Kurzweil's *The Singularity is Near*, page 70



10



Shouldn't we hold off on artificial intelligence until we figure out actual intelligence?

11

What is Intelligence?

A proposal for our purposes today:

Intelligence is the ability to sustain and/or enhance one or more values in various contexts over time.

12

Note some key features:

1. Intelligence, in this sense, is **guided by values**. We can only distinguish more vs less intelligent behaviors if we care about consequences – In a world without values, there are no wrong answers and no smarter methods.
2. Intelligence **reshapes or adapts the self and/or the world** to reflect those values.
3. It can take **many forms** -- calculating a solution, negotiating an agreement, writing a melody, constructing a piece of furniture, sharing an intimate feeling, cooking a new dish, keeping warm, nurturing an offspring . . .
4. In this sense, organisms and biological communities may exhibit intelligence and so may machines and other systems – **"intelligence" in this sense does not require consciousness**.
5. **"Intelligence"** may be partial and limited, falling short of a **wisdom** that responds appropriately to the full range of values we should hold in our lived context.

13

Contrast Intelligence with ***Wisdom***
which we might tentatively define as:

"***systematic intelligence that responds appropriately to the full range of values we should hold in the context in which we live***."

In that sense wisdom is human ecological.

Unfortunately many who aim at high levels of intelligence often focus on only one or a few relevant values – un-wisely.

14

Artificial Intelligence is:

- Created by **"artifice"–** a design process at least initially, in part, guided by explicit intentions
- Typically **silicon** based but **need not be**
- Traditionally programmed by a person or team but **can be designed to use evolutionary processes,** for example**, to program itself.**

Today I will use this intentionally very broad definition of AI that includes everything from the minimal intelligence of heating thermostat controlled to the most advanced forms of machine learning systems like GPT-3 and Wu Dao 2.0

15

It is useful to distinguish varieties of Artificial Intelligence (AI) as:

Narrow AI – machine able to act intelligently in some limited, well defined arena

Artificial General Intelligence (AGI) – able to act intelligently across a broad range of areas comparable to a mature human

Artificial Super Intelligence (ASI) – able to increase its own intelligence at an exponential rate surpassing human capacity to comprehend (leading to the "Singularity")

16

Two Projects in our Civilization	Two Promises	Two Perils in "values alignment"
1.) Increasingly "Smarter Planet" AI extending capitalist "rationalization"	Ever Better Management of our Spaceship Earth	The Spaceship of Fools: Narrow AI managing for only one or a few values
2.) Ever smarter AI – leading to a "Singularity" of Artificial Super Intelligence	Ever Better Science and Technology	Two "Friendly AI" Problems: Will AI be ethical enough – and will we?

17

Part 2.) Two kinds of reasoning with illustrations from ethics

A.) monological inference following algorithms

vs.

B.) dialogical negotiation following strategies

18

Part 2.) A. Reasoning as Monological Inference
employing algorithms

Example A from the history of Formal Logic -- Aristotle (384-322 BCE) formulates a system of Categorical Syllogisms illustrated here with one of its algorithmic rules:

If "All A are B" ***and*** "C is A", then "C is B".

Application:

Step A	Input (premises): All men (A) are mortal (B). ***and*** Socrates (C) is a man (A).
Step B	*run the algorithm . . .*
Step C	Output (conclusion): Socrates (C) is mortal (B).

19

Further Examples:

Euclid (˜350-250 BCE) formalizes logical proofs of a system of geometry in ***The Elements***.

Newton (1642-1726) formalizes proofs for a systematic theory of physics.

This inferential model of reason is **monological** in the sense that it starts from one point of view, one set of premises, and draws conclusions. And it can be all carried out by a single individual – or machine.

20

In the 18th Century, Jeremy Benthan and Immanuel Kant adopted this same core conception of rationality as their model for thinking about ethics. They sought one or a few principles to enable them, like Newton, to provide "laws of moral motion" to create an ethical rocket science they could use for a set of premises about the data of the world and the axioms of ultimate value to then, following logical algorithms, infer conclusions about how to act.

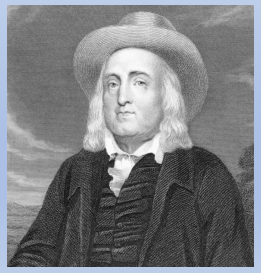
Ethical rationality as monological inference using:

1. the input of premises
2. and the algorithmic rules of logic
3. to infer conclusions in a unilateral way.

21

Utilitarianism -- Jeremy Bentham's Greatest Happiness Principle:

Always choose that action that will yield the greatest net happiness to all concerned!



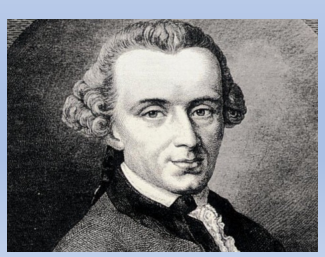
22

For the Duty based ethics of Kant's Categorial Imperative, the fundamental axiom could be expressed as:

Act only according to that policy that you can, rationally, at the same time, will as a universal law!

or, alternatively:

Second: Never act according to a policy in which you treat other rational creatures as means only (mere things) but rather, always with **respect** as **ends in themselves** (persons)!



23

Which approach to choose?

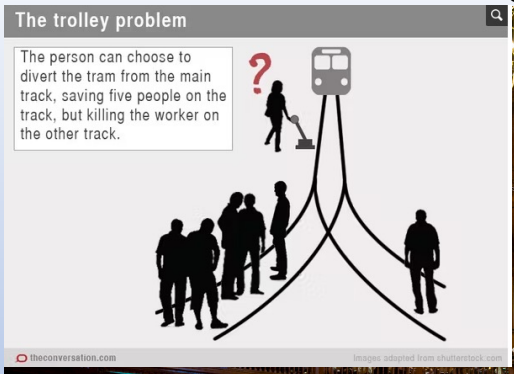
In the US, a dominant pedagogical approach is to focus, in university classes, on ethical situations like the Trolley Car Dilemma -- for two purposes: 1. To force students to clarify their own intuitions or prejudices about which ethical principle they believe is more fundamental and 2. To give students practice in the kind of moral reasoning both Bentham and Kant assumed was appropriate in ethics, namely, monological processes of inferences using algorithms to go from premises to conclusions.

There is a fascinating documentation of this pedagogy in a youtube video of an exemplary teacher, Michael Sandel of Harvard University.

24

Mainstream Dilemma Based Pedagogy in Ethics

Michael Sandel teaching "Justice" at Harvard
www.justiceharvard.org



Example of such pedagogy in popular culture with "Trolley Car Dilemmas."

25

Examples

Utilitarian Reasoning:

P 1. If an option yields the greatest net happiness to all concerned, choose it.

P 2. The option of pulling the Trolley switch to kill one instead of four will yield the greatest net happiness.

C 1. Choose to **pull the Trolley switch**.

Kantian Reasoning:

P 3. If you can not will a policy from every point of view, do not act on it.

P 4. You can not will the policy of pulling the Trolley switch from the point of view of the person who will be sacrificed.

C 2. **Do not act on the policy of pulling the Trolley switch**.

26

A Further Dilemma (and a Student's Creative Response):

You are a surgeon who has four patients in need of different organs for lifesaving transplant.

And a healthy patient asleep in the waiting room . . .

As a rational Utilitarian, what would you do?

As a Kantian?

27

But notice:

In real life, the student with the third option is just the kind of innovative thinker we would want on our team. We want folks like him in the dialogue brainstorming other creative options – like polling the terminally ill to see who might have a motive for making such a sacrifice. The search for new ways of framing options available – and people's underlying interests -- can often provide "win/win" outcomes by "increasing the size of the pie" or even provide outcomes that stop framing the situation as a conflict with winners and see it instead as a shared problem participants are collaborating on.

Approaching ethics this way, we would want to make use of an alternative model of rationality.

28

Part 2.) B – Reasoning as Dialogical Negotiation following guiding strategies

29

Instead of algorithmic rules to follow, Dialogical Reasoning is structured by **strategies that guide**. They suggest methods of observation, discernment, search and creative invention. Instead of an inference from premises to conclusion, as in formal logic, the process of reasoning would be of this basic form:

Step A. Encountering a difference with Other(s) →
Step B. pursue strategies of negotiation/problem solving in dialogue →
Step C. . . . till reaching genuine, voluntary agreement.

The Harvard Negotiation Project's *Getting to Yes* proposes guiding strategies like:

1. Multiply Options!
2. Focus on interests behind positions!
3. Separate the people issues from the engineering problems!
4. Look for objective criteria!

30

In the last fifty years, research on negotiation and conflict transformation has yielded detailed accounts of these strategies and a host of others that help parties "get to Yes", engage in group problem solving, community based collaboration, mediation, dispute resolution, conflict transformation and peacemaking.

And studies of dialogical reasoning have spread to a wide variety of other fields . . .

31

32

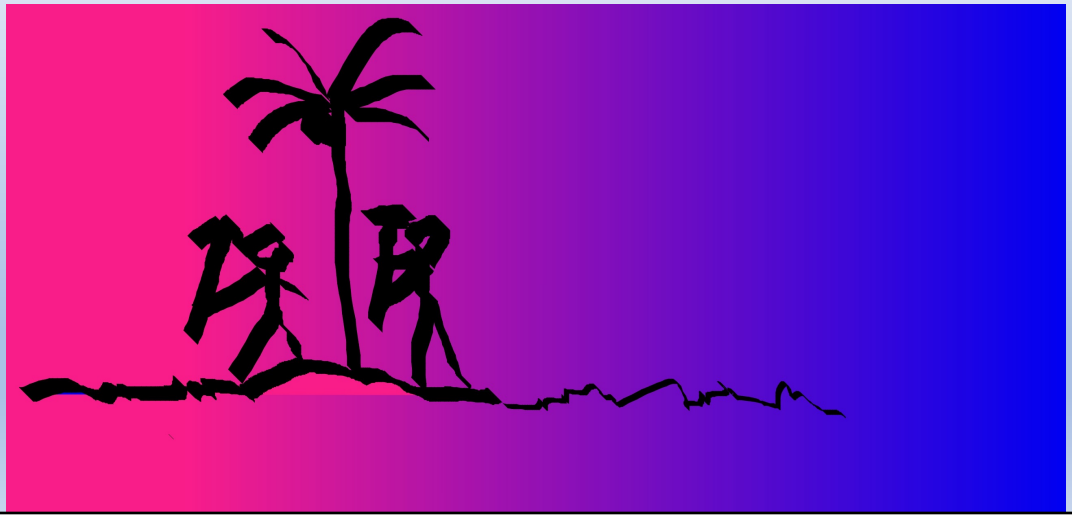
In different ways, this research aims at a shift in our civilization – following Gandhi.

We live in a culture in which peace is obscured, defined in terms of what it is not, and as a state rather than an activity like its "opposite", war. In English, "peace" can't be used as a verb.

This is because our culture is dominated by practices of things like economics, law, politics, and public discourse that assume conflict is essential to life.

33

Culture of Conflict core metaphor for life: Two Islanders and only one coconut . . .



34

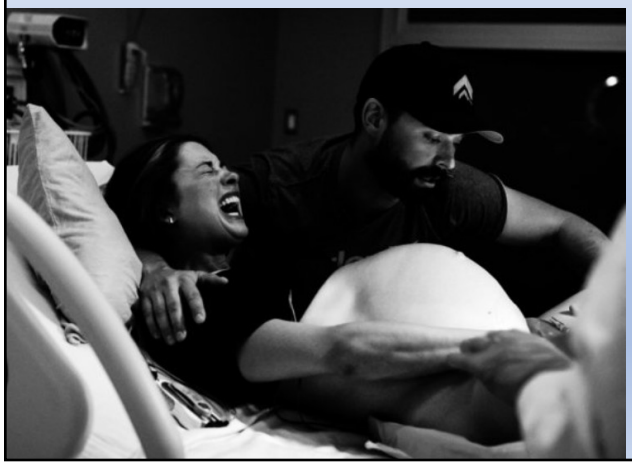
Models and Metaphors for Dealing with Differences between People
Ranging from Conflict-Centered Cultures toward Cultures of Peace

Lose/Lose -> Win/Lose -> Win/Win -> --

- Shared Problem solving
- Birthing New Selves
- Team players
- Collaborative Dance
- Interdependent Self-arising
- Agape Love
- Braiding Sweetgrass

35

The process of birth can provide one core metaphor for Culture of Peace



36

Distinctive features of these traditions of dialogical reasoning?

First, they understand the reasoning process as involving two or more real people with substantively different language, beliefs, and norms for starting points.

The challenge for these parties is to negotiate those differences and develop new language, practices and plans of action on which they can agree.

37

Second, they commit to seeking genuine agreement through nonviolent practices of investigation and persuasion, without threats or coercion.

38

Third, they use variations on the **four basic guiding strategies**:

Multiply Options!

Focus on interests behind positions!

Separate the people issues from the engineering problems!

Look for objective criteria!

Note that these take the form of open-ended imperatives that guide. **They are not algorithms.**

A fourth common feature is a shift away from the *Golden Rule which says*: Do unto others as you would have them do unto you!

As typically interpreted, that rule can provide an ethnocentric approach inviting colonialism and imposing our own preferences and values on others.

Instead, we are asked to start by inquiring into others' interests and values and shift to the ***Rainbow Rule:*** Do unto others as **they** would have you do unto **them**!

39

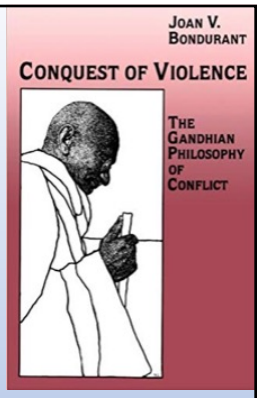
A fifth feature of these exemplars is that they understand the elements and aspects of the reasoning process in "emergentist" rather than "static" or "reductionist" ways. For them, the meaning and truth of sentences, the identities of the selves and communities stating them, and the social realities involved all emerge and grow or otherwise develop during the dynamic course of negotiation.

Many of the distinctive practices of rationality in these traditions focus, precisely, on methods for getting shared meanings to emerge in forms that express increasingly truer views of our options and become more agreeable for all.

In his "Experiments with Truth", Gandhi developed a kind of experimental method for discerning, demonstrating and defending emergent objective moral truths through practices of what he called non-violent "clinging to truth" or *"satyagraha".*

40

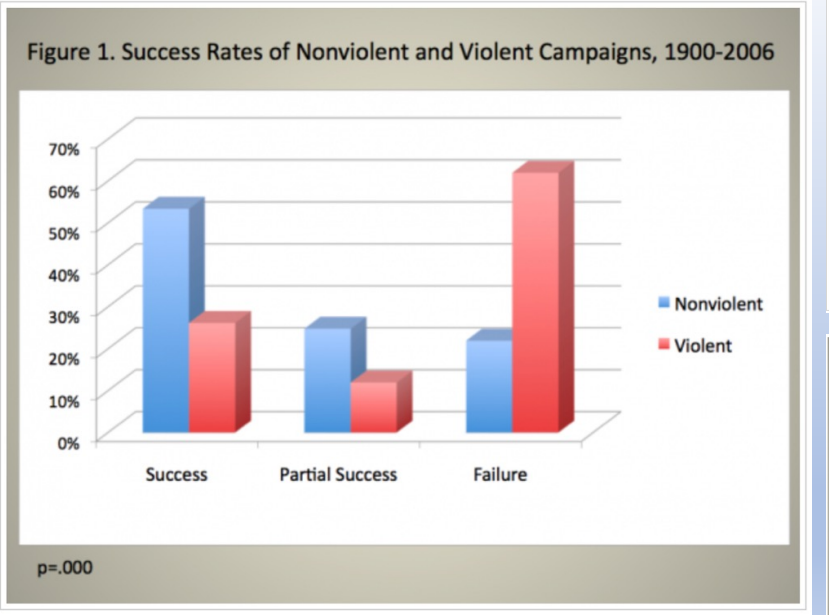
Noviolent satyagraha practices of voluntary self-suffering can serve to:



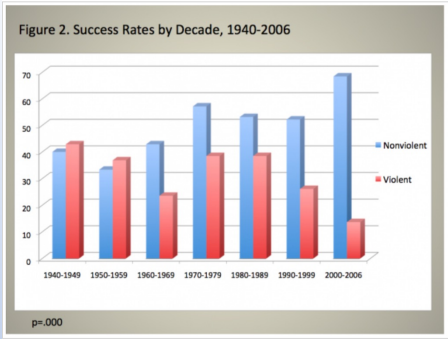
1. Discern moral truths for oneself -- the prospect of pain gave pause for considered reflection and for a review of conscience
2. Demonstrate moral truths to others – "melt their hearts" and persuade them of its truth.
3. Defend moral truths by active resistance to oppression

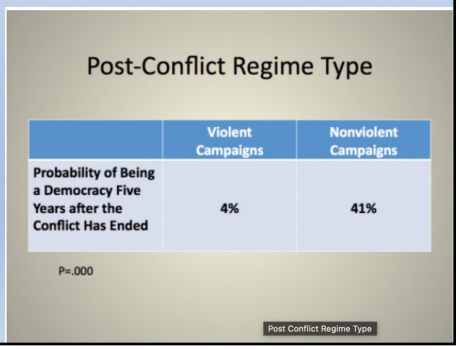
-- sit-ins, boycotts, and the establishment of parallel institutions, etc

41



From Chenoweth & Stephan 2012, Why Civil Resistance Works





42

But, to sum up, so far, I suggest we need to move

from a primary reliance on the 18th Century model of rationality as monological inference that makes us "smarter"

to a more inclusive 21st Century model that draws on monological reasoning to express individual voices but then seeks to resolve their conflicts through **more inclusive** forms of **dialogical rationality** that make us **wiser**

and help us deal not just with "complicated" problems like landing on the moon but also with **"complex" or "wicked" problems** like ending poverty – problems that involve multiple, divergent and incongruous perspectives and frames of meaning for understanding values, elements and dynamics.

43

Two Basic Frames → of Reasoning with variations ⬇	**Monological** Reasoning modelled on the "rocket science" of Newton	**Dialogical** Reasoning exemplified by Gandhian and other consensus approaches to conflict transformation
Economics	**"Rational Economic Man"** Individual Producer/ Consumer Maximizing Profit and "Utility"	**Rational Historical Agent** pursuing meaningful projects in community
Politics & International Relations	Nation States pursuing power through **realpolitik**	Communities pursuing **swaraj** through **satyagraha**
Technology	Pursuit of maximum power to manipulate environment through"smart" algorithms of **instrumentalist** reasoning	Pursuit of wise and sustainable relationships in community through **dialogue** including local and indigenous knowing
Morality	Seeking **foundations** in absolute, universal principles or intuitions (e. g. utilitarian, Kantian)	**Experimental search** for **emergent objective Truth** through satyagraha

44

Part 3.) Two models of AI – the Turing Machine vs. the Turing Child

45

A. M. Turing (1950) Computing Machinery and Intelligence. *Mind 49:* 433-460.

COMPUTING MACHINERY AND INTELLIGENCE

By A. M. Turing

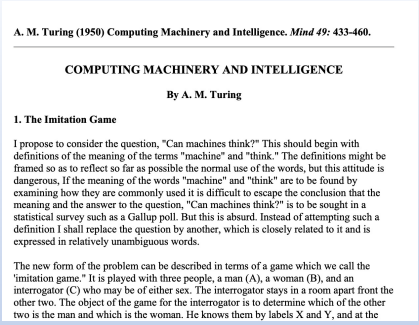
1. The Imitation Game

I propose to consider the question, "Can machines think?" This should begin with definitions of the meaning of the terms "machine" and "think." The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous, If the meaning of the words "machine" and "think" are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, "Can machines think?" is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem can be described in terms of a game which we call the 'imitation game." It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart front the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the

46

Alan Turing (1912-1954)
”Computing Machinery and Intelligence”



A. M. Turing (1950) Computing Machinery and Intelligence. *Mind 49:* 433-460.

COMPUTING MACHINERY AND INTELLIGENCE

By A. M. Turing

1. The Imitation Game

I propose to consider the question, "Can machines think?" This should begin with definitions of the meaning of the terms "machine" and "think." The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous, If the meaning of the words "machine" and "think" are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, "Can machines think?" is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem can be described in terms of a game which we call the 'imitation game." It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart front the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the

1. Imitation Game or Turing Test
to operationally define intelligence

2. Turing Machine – as the defining model of the modern programmed, inferential, algorithmic computer

3. Turing Child – as vision of a machine that learns through dialogue and socialization

47

The essence of a Turing Machine:

“The idea behind digital computers may be explained by saying that these machines are intended to carry out any operations which could be done by a human computer. The **human computer** is supposed to be **following fixed rules**; he has **no authority to deviate from them in any detail**. We may suppose that these rules are supplied in a book, which is altered whenever he is put on to a new job. He has also an unlimited supply of paper on which he does his calculations. He may also do his multiplications and additions on a “desk machine,” but this is not important.”

48

Key Elements and Functions of a Turing Machine

- Takes **input** as sequences of symbols – e. g. “0”, “1”, “-”, ”<”, “=”, and “2” or ”countdown” and “LAUNCH
- **Storage** for strings of sequences and for algorithmic rules
- Applies **algorithms** to the input sequence to transform it – e. g. ”If C > 0, then C = (C-1)” or . . . “If countdown = 0, then LAUNCH.”
- Then it delivers an **output** – e. g. ”C = 0” or . . . ”LAUNCH

Note: Here an algorithm is understood as a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer. To function properly, every element of it must be clearly and explicitly defined in a finite way so that it can be followed in an automatic or mechanical manner.

49

Turing’s second model, The Child approach:

“In the process of trying to imitate an adult human mind we are bound to think a good deal about the process which has brought it to the state that it is in. We may notice three components:

The initial state of the mind, say at birth,

The education to which it has been subjected,

Other experience, not to be described as education, to which it has been subjected.

Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child’s? If this were then subjected to an appropriate course of education one would obtain the adult brain.”

50

This second model is not a tool that is programmed by a user; it is a child that is educated in a community:

"It will not be possible to apply exactly the same teaching process to the machine as to a normal child. It will not, for instance, be provided with legs, so that it could not be asked to go out and fill the coal scuttle. Possibly it might not have eyes. But however well these deficiencies might be overcome by clever engineering, one could not send the creature to school without the other children making excessive fun of it. It must be given some tuition. We need not be too concerned about the legs, eyes, etc. The example of Miss Helen Keller shows that education can be take place provided that communication in both directions between teacher and pupil can take place by some means or others."

51

Some key features of a "Turing Child":

The child machine will need to have a **body** and **engage in dialogical reasoning and interaction.**

The "programming" structuring such behavior will require kinds of interaction that are not monological reasoning or algorithmic calculations taking place in a formal "object" language. They will have to involve dialogue in which the **teacher and child machine** repeatedly **renegotiate the meanings** of terms and sentences.

They will also have to be able to move back and forth between the object language and the **meta-language** standpoints.

52

In that regard, with a reference to Bertrand Russell's introduction of a Theory of Types to avoid paradoxes of self-reference and infinite regress, Turing makes the following very revealing comment:

"The processes of inference used by the machine need not be such as would satisfy the most exacting logicians. There might, for instance, be no hierarchy of types. But this need not mean that type fallacies will occur, any more than we are bound to fall over unfenced cliffs. Suitable imperatives (expressed within the systems, not forming part of the rules of the system) such as 'Do not use a class unless it is a subclass of one which has been mentioned by teacher' can have a similar effect to 'Do not go too near the edge.'"

53

We are reaching a critical moment in which obstacles to creating "Turing Child" machines may be receding. It is the stage Max Tegmark describes as "Life 3.0" – with entities that can intentionally redesign their hardware and software.

We are, further, at a stage in which cutting edge textbooks in AI are reframing their core goals. For instance, Stuart Russell and Peter Norvig note, in the newest edition of their classic intro book:

54

”Previously we defined the goal of AI as creating systems that try to **maximize expected utility**, where the specific utility information – the objective – is supplied by the human designers of the system. **Now we no longer assume that the objective is fixed and known** by the AI system; instead, the system may be uncertain about the true objectives of the humans on whose behalf it operates. It must learn what to maximize and must function appropriately even while uncertain about the objective.”

-- Stuart Russell and Peter Norvig***, Artificial Intelligence: A Modern Approach***, fourth edition, p. vii

55

A Very Rough Sketch of AI History:

1950 -- Good Old Fashioned AI
based in symbolic logic, aiming at AGI through handcrafted knowledge engineering

1980 – growing focus on Narrow AI

2012 – Boom in Big Data and ”connectionist evolutionary strategies of programming using neural nets” and other forms of “machine learning”

The Present and Future -- Growing aim to combine symbolic logic and evolutionary strategies with robotics, reinforcement learning, deep learning, collaborative and community based programing, dialogical reasoning and . . . **”Turing Children”?**

56

Two Approaches to AI	Turing Machines	Turing Children
Style of reasoning	Single frame, unilateral, monological	Multiperpectival, collaborative, dialogical
Process of reasoning as inference. Vs. negotiation	Uses algorithmic rules to infer conclusions from premises or outcomes from inputs	Uses guiding strategies to arrive at shared solutions or genuine, voluntary agreements
Starting point	Given definitions of terms, data and rules	Different points of view with different meanings ascribed to terms and different beliefs and rules
Process of reasoning	Inference according to determinate algorithm rules	problem solving and conflict resolution in which any meaning, belief or value can be renegotiated
Goal of reasoning	Generate conclusions as output	Reach genuine voluntary agreements
Truth conceived as	Statements in correspondence to a fixed reality	Cultivation of shared understanding of emergent objective reality
Reasoner as	Substrate and context independent computer	Not substrate independent, must be an embodied agent engaged in dialogue in open-ended contexts that include life worlds
Method of enhancing reasoning	Reprogramming software, increasing the speed and power of the hardware	Socializing the agent in lived contexts through parenting, teaching, play . . .

57

NOTE: the difference between a Turing Machine and a Turing Child is not a matter of consciousness or using some breakthrough technology like quantum computing.

It draws on a familiar basic process we all are born capable of and can learn and teach.

We can improve our skills at it.

We can incorporate them into our practices as individuals, communities – and programmers.

And we can get computer systems to nudge us to use them more, use them better, and incorporate them in the reasoning processes of the machines themselves.

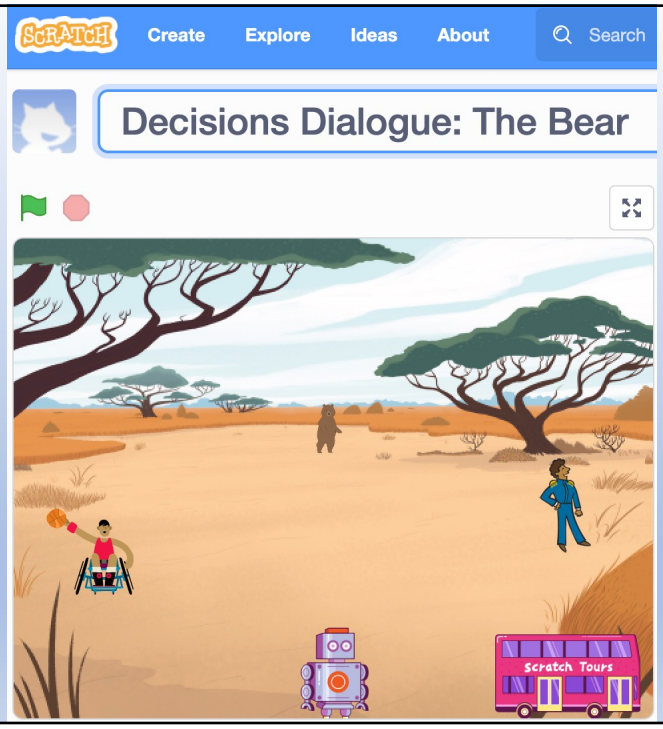
58

Part 4.) Strategies for developing Turing Child approaches in programming and collaborative problem solving

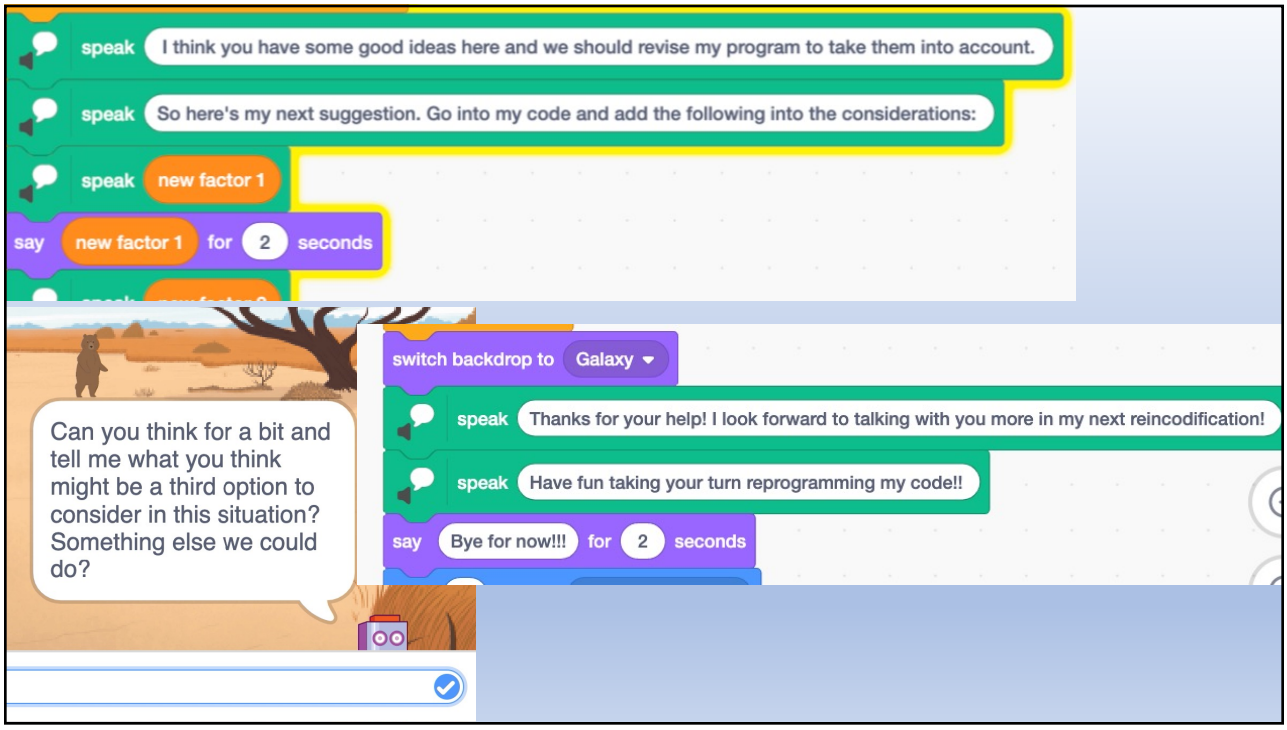
59

A key part of the shift involves thinking, explicitly, about how groups of people can incorporate dialogical methods in the decision processes that groups they go through when using computers. This can start in extremely simple ways. For example, with a program teaching ethics and block coding to kids.

https://scratch.mit.edu/projects/428374274/



60

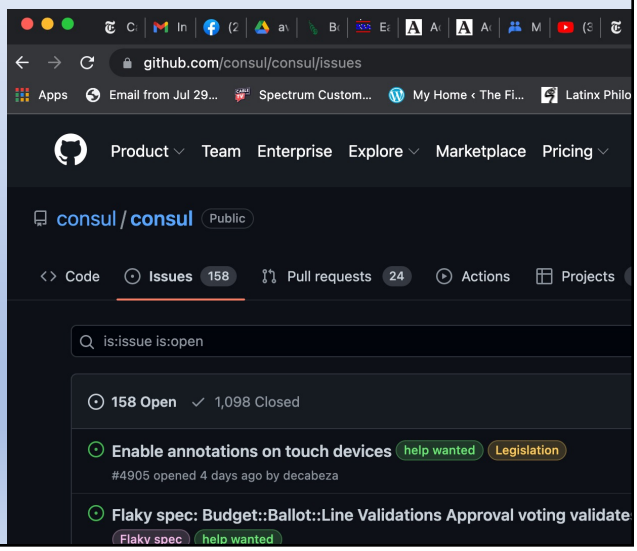


61

Variations on this same strategy can be done in working with adults in any context in which AI is being deployed in a community. For example, an organization developing software for reporting sexual assault can directly involve survivors in the ongoing redesign of the system using dialogical methods and incorporating elements of them into the group processes and structures as well as the program.

62

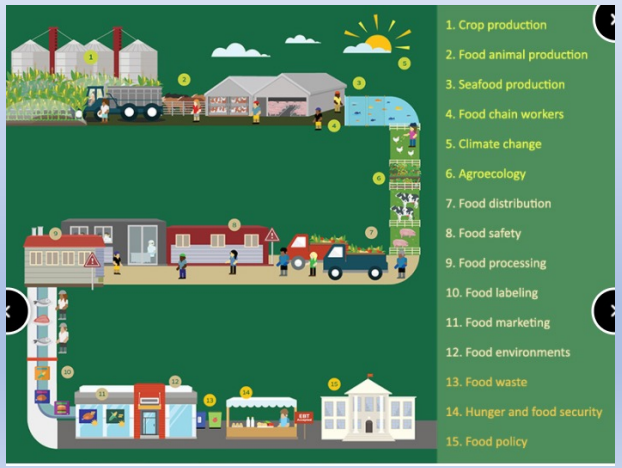
Likewise, an organization like Consul can develop software for municipal governments using open source sites like GitHub that allow for version control. And it can adapt them to create ease of entry and interaction for community members who are not programmers but who play key roles in the dialogical reasoning processes.



63

Seven Key Principles of Dialogical Approaches to AI/Human/Nature Systems

1. In concepts, diagrams and practice, the projects should always be **framed as an AI/Human/Nature systems.** Intelligence is always an activity guided by values and concerns whose meanings are grounded in a holistic context. *Machines can make bits. Only a community can make a meaning.*



64

Strategies:
Seven Key Principles of Dialogical Approaches to AI/Human/Nature Systems

1. In concepts, diagrams and practice, the projects should always be **framed as an AI/Human/Nature systems.** Intelligence is always an activity guided by values and concerns whose meanings are grounded in a holistic context. *Machines can make bits. Only a community can make a meaning.*

65

2. The overall **goal** is to arrive at **genuine, voluntary agreements** – not to simply generate output -- genuine, voluntary mutual agreements between the AI, people and other natural organisms and ecological systems involved in the community engaged in the values and concerns at stake.

understandable

consent

non-coercive context

emergent objective values

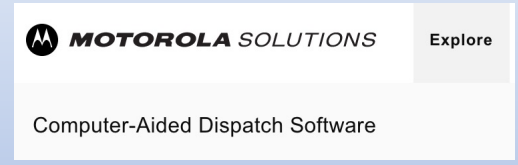
66

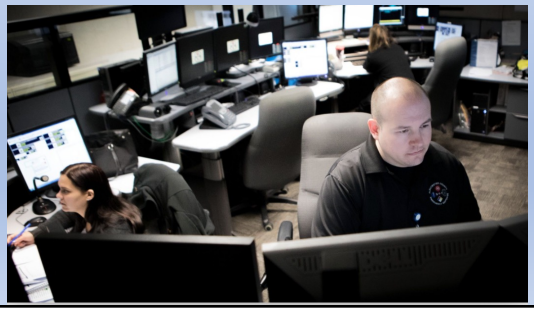
3. The AI procedures need to **flag for review** the cases in which their data, algorithms, framing assumptions and/or outputs are especially questionable and need review by a human or by a representative group of humans and natural organisms from the larger community.

e. g. sensors' margins of error, training data for facial recognition, change of context for application, high risk

machines initiate negotiations

machines advocate appropriately for values and concerns





67

4. The algorithms of the programs can be modified to conform to the agreements arrived at through **easily engaged meta-operations** that can be relatively easily accessed by other participants in the dialogical process.

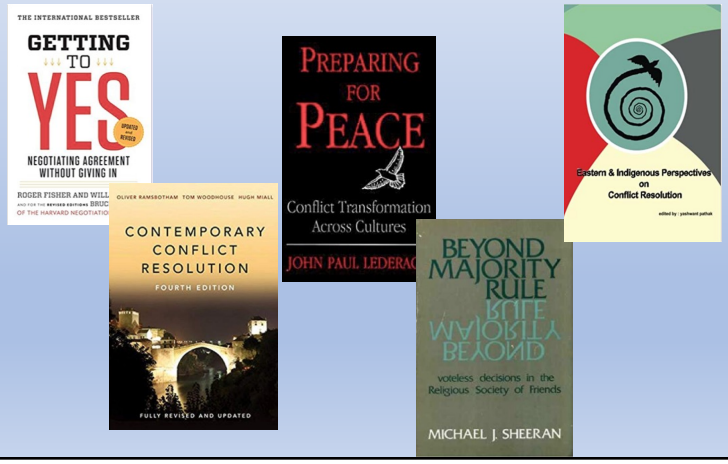
through direct intervention by people

through reform of its own programming

and in combination with principle #3 – e. g. flagging with phrases like "What should I be looking at here?" and "Do you have any idea why I seem to keep getting these two things confused?" or "I don't get it. Why isn't this one an X?"

68

5. The dialogical interactions the AI engages in should be framed and guided by **principles of conflict resolution** as illustrated by Roger Fisher et. al.'s *Getting To Yes*, John Paul Lederach's *Preparing for Peace*, and other studies of negotiation and conflict transformation practices from around the world.



For example: focus on underlying interests, multiply options, "separate the people from the problem", Look for independent, objective criteria

flag problem points

generate a library of specific proposals as well as strategies

69

6. The AI should have a structure and **committed embodiment** that commits it to interests in the well-being of the community in which it is operating.

interdependent with the people and natural systems it engages with

tied irrevocably to physical machinery and power inputs that depend on the community for their maintenance -- should not exist merely as a cloud entity that is substrate independent

can **"emigrate" or become exported AI capital only through genuine, voluntary agreements with the community** that created and maintained it up to that point

70

7. We should work to strengthen processes through which the AI/human/nature system can **discern tacit patterns** in the meanings that provide the context of its thought and action and make them explicit in spirit-led dialogue.

tacit patterns of **value as well as fact**

include both **physical** or material patterns but **also emergent formal and meta-structural patterns**

humans involved use reflection, meditation, "meeting for worship for discernment" & other methods to practice spirit-led communal discernment

also experiment drawing on the distinctive forms of intelligence offered by machines and by natural systems

"holding in the Light" not only natural systems like watersheds or forests but also the machines and artificial intelligence systems

71

Human Ecological Principles of a Collaborative Wisdom Approach to Dialogical Programming:

1. framed as an AI/Human/Nature systems
2. goal is to arrive at genuine, voluntary agreements
3. AI procedures need to flag for review
4. easily engaged meta-operations
5. principles of conflict resolution
6. committed embodiment
7. discern tacit patterns in spirit-led dialogue

How might these 7 ideas for "AI as Collaborative Wisdom (CW)" be applied in practice in your own work?

72

Acknowledgements

In working on these ideas I am very grateful for help from colleagues:

at the College of the Atlantic including, especially, Dave Feldman, Kyle Shank, Chris Petersen, Ken Cline, Davis Taylor, Suzanne Morse and my student Phileas Dazeley-Gaist,

at the Quaker Institute for the Future including, especially, Judy Lumb, Larry Jordan, Geoff Garver, Keith Helmuth, Charles Blanchard, and Sara Wolcott

at the Friends Association for Higher Education including, especially, Laura Rediehs and Don Smith

and at large including, especially, Ram Subramanian, Rich Hilliard, Brian Jennings, and Katie Wasserman

as well as support from the US National Science Foundation for research on differing cultural norms and their effects on the adaptation of technologies and the adoption of them.

73

74