Chapter XX

Intelligence and Consciousness in Natural and Artificial Systems

1. Introduction

Humans are highly intelligent, and their brains are associated with rich states of consciousness. We typically assume that animals have different levels of consciousness, and this might be correlated with their intelligence. Very little is known about the relationships between intelligence and consciousness in artificial systems.

Intelligence is a complex multifaceted term and many overlapping definitions have been put forward [Legg & Hutter, 2007a]. Most of these definitions were developed to describe human intelligence and they have severe limitations as definitions of non-human and artificial intelligence.[a] To address this issue, I have developed a new interpretation that links intelligence to a system's ability to generate accurate predictions (see Sec.2.2).

Many theories have been put forward about the nature of consciousness and its relationship to the physical world. There is less diversity in the definitions of consciousness: many people agree that consciousness is the stream of colorful moving noisy sensations that starts when we wake up

[a] See Burkart et al. [2017] for a discussion of intelligence in animals.

and ceases when we fall asleep at night. My version of this definition is given in Sec. 4.1.

These definitions of intelligence and consciousness enable us to carry out preliminary non-scientific work on the relationships between intelligence and consciousness. We can use our knowledge of the domains and our own imagination, intelligence and consciousness to picture possible relationships between intelligence and consciousness in natural and artificial systems. Some of the insights that can be gained from this approach are covered in Sec. 5.

A scientific understanding of the relationships between intelligence and consciousness can be developed when we have accurate ways of measuring intelligence and consciousness in natural and artificial systems. Previous work on the measurement of intelligence is covered in the first half of Sec. 3, and Sec. 3.4 describes a new method that I have developed for measuring predictive intelligence. Sec. 4.3 explains how we can use mathematical theories of consciousness to make deductions about the consciousness of non-human animals and artificial systems. Accurate measurements of intelligence and consciousness will eventually lead to a more systematic understanding of the relationships between intelligence and consciousness in natural and artificial systems.

2. Definitions of Intelligence

2.1 Previous Definitions of Intelligence

Intelligence is a complex multifaceted term and many overlapping definitions have been put forward. These include cognitive ability, rational thinking, problem-solving and goal-directed adaptive behavior [Neisser et al., 1996; Bartholomew, 2004; Legg & Hutter, 2007a]. Most of these definitions are based on factors that are linked to intelligence in humans. They often generalize poorly and generate many counterexamples when we try to apply them to non-human animals and to artificial systems.

The problems with defining intelligence have led some people to view intelligence as a collection of abilities. For example, Thurstone [1938] claims that intelligence consists of verbal comprehension, word fluency,

number facility, spatial visualization, associative memory, perceptual speed and reasoning. Sternberg [1985] identifies analytical, creative and practical components of intelligence, and Gardner [2006] suggests that there are multiple types of intelligence, including musical intelligence, linguistic intelligence and emotional intelligence. Warwick [2000] frames this more generally with his idea that intelligence is a high-dimensional space of abilities.

A distinction is often made between crystallized and fluid intelligence [Cattell, 1971]. Crystallized intelligence is a stored ability to solve problems. For example, older intelligence tests included factual questions, such as "Who is the president of the USA?". The answers to this type of question must be remembered – they cannot be deduced by reasoning. Crystallized intelligence also includes rules that can be used to solve problems, known as heuristics. For example, it is theoretically possible to deduce how to solve a Rubik's cube from scratch. However, most people use heuristics to solve different parts of the problem - for example, a method for moving a color to a different face - and then sequence the heuristics together to complete the puzzle. Heuristics also exist for some of the problems that appear in intelligence tests.

Fluid intelligence is the ability to generalize knowledge and solve problems that have not been seen before. For example, someone with high fluid intelligence might be able to generalize what they have learnt from solving the Rubik's cube to similar puzzles. Modern intelligence tests are mostly designed to measure fluid intelligence. In humans there is a constant interaction between fluid and crystallized intelligence. A solution to a problem might be discovered through fluid intelligence and then stored for rapid recall at a later date.

If we want to understand intelligence in non-human animals and artificial systems, then we need a non-anthropocentric general definition of intelligence that can be applied to any system at all. The next section outlines a number of reasons for thinking that intelligence is closely linked to prediction.

2.2 Prediction and Intelligence

Prediction and the Brain

In recent years there has been a surge of interest in the idea that the primary function of the brain is the generation of predictions [Clark, 2016]. According to these theories, each layer in the mammalian cortex[b] generates predictions about activity in the layer below. The layers compare the predictions from higher layers with their own activity and pass information about the prediction errors back up to the layers above. This explains why there are more top-down than bottom-up connections in the brain.

Predictive brain theories typically treat the brain's predictions as probability distributions. This accommodates situations in which we are certain about something, as well as more common scenarios in which we assign probabilities to different events. People working on the Bayesian brain investigate the extent to which the probability distributions of the brain's predictions match the probability distributions of the environment [Knill & Pouget, 2004; Doya et al., 2007]

Bayesian and predictive theories are plausible and attractive interpretations of the brain that are consistent with our subjective experiences. If these hypotheses are partly or wholly true, then the generation of probabilistic predictions is a core function of the brain, and we would expect there to be a strong correlation between a brain's predictive ability and its intelligence.

At the present time there is little direct evidence for predictive interpretations of the brain. Bayesian brain theories are supported by more experiments, but these are controversial – for example, Bowers and Davis [2012] claim that Bayesian models are frequently adjusted to match the data, leading to unfalsifiable theories that are rarely compared with alternatives. There are also ongoing issues with small sample sizes and the reproducibility of experiments in psychology [Collaboration, 2015; Baker, 2016].

[b] Predictive brain theories also apply to animals with different brain architectures, such as cephalopods and birds.

A probabilistic and predictive interpretation of intelligence is likely to be attractive to people who already believe that Bayesian and predictive theories of the brain are true. In the future, better evidence might emerge for Bayesian and predictive brain theories that would support a link between prediction and intelligence.

Prediction and Action

As I interact with the world, I am constantly predicting the results of different possible actions and selecting the ones that lead to my goals. For example, when I am hungry, I consider the location of different shops and plan how I can get to the best one, considering traffic, petrol, crime, and so on. A system that cannot predict cannot plan – it can only react to changes in its environment as they occur. On the other hand, a system with perfect predictive ability would have god-like omniscience. It would know what would happen under all possible permutations of its environment and could plan sequences of actions that have the highest probability of achieving its goals.

As animals increase in intelligence and complexity there is a shift from hard-wired reactions to planned behaviors based on prediction. Snails follow chemical trails and retreat when danger threatens. The world does something to the snail and it responds in an evolutionarily determined way that, on average, leads to the future survival of the species. More sophisticated animals, such as sheep, can classify features of their environment (food, enemies, mates, etc.) and they have a limited ability to predict how their environment will respond to their actions [Gamez, 2019; Marino & Merskin, 2019]. Corvidae (crow family) and cephalopods (for example, octopi and squid), combine reactive behaviors with actions based on richer predictions about their environment, which enables them to solve more complex problems and build tools. Humans combine their reactive behaviors with planning based on complex predictions on multiple time scales.

Prediction and Artificial Intelligence

Systems that are classified as artificially intelligent replicate behaviors that typically require intelligence in humans. Self-driving cars and chess-playing programs are regarded as intelligent because human intelligence is required to drive cars and play chess. Dialysis machines, that replicate the functions of the human kidneys, are not regarded as artificially intelligent because blood filtration does not require intelligence in humans.

The problem with this definition of artificial intelligence is that computers can imitate human behaviors in simple ways that do not require intelligence. For example, natural language conversation requires intelligence in humans, but it can be reproduced to a limited extent in chatbots using simple pattern-matching algorithms [Neff & Nagy, 2016]. AI systems that are more plausibly intelligent include game-playing systems, such as AlphaGo, which predict the consequences of different actions in the space of the game. Robots and self-driving cars contain intelligence that enables them to predict the consequences of different actions in the world. AI systems that generate predictions about the future (for instance, climate models) are also regarded as intelligent.

Prediction and Compression

Some people have suggested that intelligence is linked to a system's ability to compress knowledge. This has led to the development of universal measures of intelligence based on compression [Hutter, 2021]. It has been shown that there is a close connection between compression and prediction [Bell et al., 1990], so compression-based theories of intelligence support a link between prediction and intelligence.

Retrodiction / Postdiction

Humans use their intelligence to discover facts about the past as well as the future. For example, historians debate the economic and social consequences of the plague; physicists develop theories about the origins of the universe. This work clearly requires intelligence, and it is typically called retrodiction or postdiction.

Spatial 'Prediction'

While prediction is typically thought of as something that occurs across time, we can also make 'predictions' about events that are happening simultaneously at inaccessible points in space. For example, at 4pm in London, I 'predict' that people are sleeping in Japan. Here I am using my intelligence to reach out beyond the spatial boundaries of my senses.

Predictive Intelligence and Environments

Some people think that intelligence is completely independent of the environment. According to this interpretation, a person has a certain amount of *general* intelligence regardless of whether they are working in the natural world or studying a genomics database.

The most convincing evidence for general intelligence in humans is a correlation between the results of tests of different human abilities, which is usually referred to as *g*. [Humphreys, 1979; Haier, 2017]. Experimental work on *g* shows that performance on numerical, spatial, memory and natural language tasks that are *designed for humans* is correlated. However, this research does not show that our ability to perform human-oriented tests generalizes to tasks that are difficult for humans, such as spatial reasoning in high dimensions or identifying patterns in large data sets. The correlation, *g*, only exists because we are comparing tasks that are reasonably easy for humans to complete. No one has demonstrated that human intelligence is general enough to solve all possible types of problem.

Human brains consume a lot of energy and evolution has made many compromises between performance, size, working memory and sensory resolution. Modern human brains are highly capable of solving typical problems in a hunter-gather environment, but they have a limited ability to generalize beyond these problems. This point is nicely made by Chollet [2019, p.22-23]:

> We argue that human cognition follows strictly the same pattern as human physical capabilities: both emerged as evolutionary solutions to specific problems in specific environments (commonly known as "the four

> Fs"). Both were, importantly, optimized for adaptability, and as a result they turn out to be applicable for a surprisingly greater range of tasks and environments beyond those that guided their evolution (e.g. piano-playing, solving linear algebra problems, or swimming across the Channel) … Both are multi-dimensional concepts that can be modeled as a hierarchy of broad abilities leading up to a "general" factor at the top. And crucially, both are still ultimately highly specialized (which should be unsurprising given the context of their development): much like human bodies are unfit for the quasi-totality of the universe by volume, human intellect is not adapted for the large majority of conceivable tasks.
>
> This includes obvious categories of problems such as those requiring long-term planning beyond a few years, or requiring large working memory (e.g. multiplying 10-digit numbers). This also includes problems for which our innate cognitive priors are unadapted; for instance, humans can be highly efficient in solving certain NP-hard problems of small size when these problems present cognitive overlap with evolutionarily familiar tasks such as navigation (e.g. the Euclidean Traveling Salesman Problem (TSP) with low point count can be solved by humans near-optimally in near-linear optimal time, using perceptual strategies), but perform poorly – often no better than random search – for problem instances of very large size or problems with less cognitive overlap with evolutionarily familiar tasks (e.g. certain non-Euclidean problems).

Non-human animals and artificial systems have less generalizable intelligence than humans [Burkart et al., 2017]. Dogs can use their social intelligence to understand humans; they cannot learn to build tools or solve advanced mathematical problems. Birds can adapt nest-building skills to construct wire tools; they cannot learn to play chess. We have built AI systems that can play Go, drive cars and predict protein folding; no one has developed a completely general AI that can solve any problem and outperform systems built for specific environments.

Humans, non-human animals and AI systems have, to a greater or lesser extent, limited abilities to solve problems that they have not encountered before. None of the known forms of intelligence are completely general - they can only solve problems within specific types of environment.

Fluid and Crystallized Predictive Intelligence

When a system encounters a problem or situation that it has encountered before, it can use its memory (crystallized intelligence) to predict what will happen next. For example, the first time that I see a cat crouch down and wiggle its bottom, I might not understand its behavior. When I see the cat pounce on its prey the meaning of the behavior becomes clear. The next time I see a cat crouching down and wiggling its bottom, I can predict, with reasonable certainty, that it will pounce. I have remembered the sequence of events and can use my memory of this sequence to map earlier events onto later events. This is one form of crystallized predictive intelligence. Crystallized predictive intelligence also includes stored heuristics that we use to solve problems and predict future events.

Fluid predictive intelligence is the ability to make predictions in situations that we have not encountered before. When I first see the cat crouching and wiggling, I might be able to *deduce* from the cat's attitude, preferences and environment that it is about to pounce. I can do this without ever having seen a cat pounce before. In the future I can use my memory to map from the cat's current state to its future behavior – the prediction has moved from fluid predictive intelligence to crystallized predictive intelligence. This transition depends on our ability to *remember* a new prediction or solution so that we can use it more rapidly next time. Without memory our crystallized predictive intelligence remains constant. This is illustrated in Fig. 1.

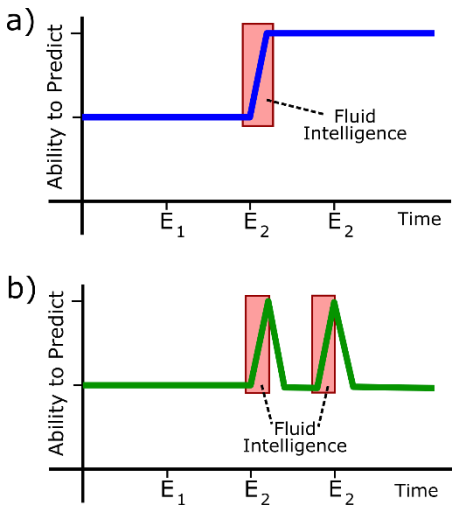

Fig. 1. Illustration of possible relationship between fluid and crystallized predictive intelligence. E_1 is a known event; E_2 has not been experienced by the system before. a) *System with long term memory*. The system predicts the consequences of E_1 using crystallized predictive intelligence. When E_2 first occurs the system uses fluid intelligence to make the predictions. It stores the results, leading to an increase in its crystallized intelligence. The second time E_2 occurs the system uses the stored solution to predict the consequences of E_2. b) *System without long term memory*. When E_2 first occurs it uses fluid intelligence to predict the consequences. This new predictive skill is temporarily held in working memory and is soon forgotten. The next time E_2 occurs it uses fluid intelligence again to make the predictions.

The graphs shown in Fig. 1 suggest that fluid intelligence is linked to an *increase* in our ability to make predictions.

2.3 Four Hypotheses about Intelligence

The discussion in the previous section leads to four hypotheses about natural and artificial intelligence:

- **H1**. Prediction is the most important component of intelligence.
- **H2**. Prediction and intelligence are relative to sets of environments.
- **H3**. The amount of a system's crystallized intelligence varies with the number of accurate predictions that it can make in a set of environments.

- **H4**. The amount of a system's fluid intelligence varies with the positive rate of change of its crystallized intelligence.

This interpretation of intelligence fits in well with many of the previous definitions of intelligence. We can use accurate predictions to plan, achieve goals and receive rewards from the environment. Crystallized predictive intelligence corresponds to our understanding of how things work in the world. However, in this interpretation, knowledge is only linked to intelligence to the extent that it helps us to make accurate predictions. The construction and manipulation of tools also depends on accurate predictions about how manipulations will change the material and how the finished tool will enable us to change the environment. As systems learn, their crystallized predictive intelligence increases. During the time of learning the system will have positive values of fluid predictive intelligence.

When we accept that intelligence is relative to a set of environments (hypothesis H2), it becomes clear that there are many different forms of natural and artificial intelligence, which are specific to their environments. We don't have to broaden our concept of intelligence to handle this (embracing Gardner's multiple intelligences or Warwick's high dimensional space of abilities). Instead, we can say, for example, that system A has a high level of predictive intelligence in a musical environment and system B has a high level of predictive intelligence in a chess environment. These environments can be natural, simulated, data, and so on.

The relativization of intelligence to sets of environments helps us to understand and appreciate the many different forms of human intelligence. IQ tests measure intelligence within an academic environment of mathematical symbols, abstract shapes, etc. This is why IQ tests are correlated with measures of academic success (school grades, advanced degrees, publication of papers, professional careers, etc.). But the academic environment is just one area where human intelligence operates. A successful plumber has a high level of intelligence within the environment of pipes, fittings, water flow, etc. and can make many accurate predictions within this environment. The same is true of other trades and professions. A predictive approach leads to a much broader

interpretation of intelligence than the academic intelligence measured by IQ tests.

3 Measurement of Intelligence

3.1 Measurement of Human Intelligence

Most of the previous work on the measurement of human intelligence has been based on sets of tests that measure behavioral characteristics judged to be linked to intelligence. In the early days these tests included significant numbers of questions based on factual knowledge. Modern human intelligence tests are now mostly based on verbal reasoning, spatial manipulation and mathematics. The results from these tests are typically converted into values of intelligence quotient (IQ) or g-score. To calculate IQ you take the test results from a sample of the population and calculate the mean and standard deviation. The mean score is assigned an IQ of 100 and each standard deviation above and below the mean corresponds to 15 IQ points. The resulting IQ score can be used to rank individuals according to how well they perform on a battery of intelligence tests. IQ is a population derived measure that does not correspond to a property of a particular individual. Measures of IQ and g-score are controversial and they have often been misused. However, they have played a valuable role in scientific research on intelligence, and they can be an effective way of pre-processing large numbers of applicants for jobs, education, or the military.

Critics of intelligence testing have claimed that intelligence tests only measure the ability of people to complete intelligence tests – they do not actually measure intelligence in the test subjects. In humans this argument is not particularly convincing because human intelligence test scores are correlated with other measures of intelligence. For example, people who score highly in intelligence tests are more likely to achieve advanced educational degrees and pursue careers in areas, such as science, that are generally regarded as requiring intelligence [Robertson et al., 2010; Haier, 2017].

3.2 Measurement of Intelligence in Non-human Animals

Animals cannot take human intelligence tests, so there has been a lot of work on the development of cognitive test batteries for animals [Shaw & Schmelz, 2017]. While it might be possible to come up with a plausible set of tests that could be applied to similar animals, this approach is likely to neglect the different types of intelligence that animals develop to survive in their ecological niches. A measure of intelligence that is designed for sheep or fish, for example, cannot easily be transferred to birds or bees. Suppose we want to develop a test that compares human and pigeon intelligence. We could include mathematical abilities and spatial reasoning in our tests, which might be common to both. But pigeons have a greater capacity to map and navigate through their environment, so should this be included in the test as well? As our test battery expands with each species we will end up with a very ad-hoc collection, with each animal scoring well on the tests that are specific to their own set of abilities. It seems highly unlikely that we will be able to design a single set of cognitive tests that would enable us to meaningfully compare intelligence across all species.

A second problem with the measurement of non-human animal intelligence is that we do not have a way of connecting an animal's test results to other indicators of intelligence for that species. Most people would agree that a person who gets top grades in school, gets a first at MIT and publishes ground-breaking physics research is likely to be intelligent. If an intelligence test gives this person a low score, then this is a failure of the test, not an indicator of low intelligence. But how could we ground the results of intelligence tests in octopi, bees or dogs? Animals do not take advanced degrees or write papers on quantum theory. It is far from clear how we could prove that intelligence tests in animals measure anything more than the ability to perform the test itself.[c]

These problems are often addressed by giving simplified human tests to animals– for example, tests of spatial reasoning or mathematical ability [Boysen & Capaldi, 1992]. These measure the extent to which non-human animals exhibit human intelligence. They are not a meaningful measure of

[c] These problems are discussed by Legg and Hutter [2007c, p.5].

non-human animal intelligence and they do not enable us to compare general intelligence across species.

3.3 Measurement of Artificial Intelligence

Turing testing is often used to measure intelligence in artificial systems. The Turing test was originally proposed by Turing [1950] as a way of answering the question whether a machine could think. He described a thought experiment in which a human and a machine were connected to an electronic typing system and placed in a separate room. A human tester asked the two systems questions and tried to decide which was the human and which was the machine. If the human tester could not reliably identify the machine, then the machine would be judged to be capable of thinking. This test is extremely challenging for machines to pass because the interrogator can ask questions about any topic. While claims have been made about AI systems passing constrained versions of the Turing test, our current AI systems are extremely far from passing the full version. Many variants of the Turing test have been proposed. These include embodied Turing tests [Harnad, 1994], behavior in game environments [Hingston, 2009] and the Animal-AI Olympics [Crosby et al., 2019], which provides an environment in which artificial systems can attempt tasks that are believed to require intelligence in animals.

One problem with Turing testing is that as machines improve they are likely to exhaust the possibilities of human tasks. For example, they might eventually map out and completely understand all the possibilities of Go, which would become for them what Tic Tac Toe is for humans – a trivial game whose possibilities can be easily comprehended. To rank AIs according to their intelligence we need tasks that challenge them and which they can complete to different degrees. If they all completely solve a task that is challenging for humans and get the same score, then we can, at most, say that they have super-human intelligence on that task.

A second limitation of Turing testing is that it relies on a clear definition of the human behaviors that require intelligence. In the past it was thought that chess playing was a paradigmatic example of intelligent behavior and that any system that could play chess well would be highly intelligent. Computers can already get an average score on an IQ test

[Sanghi & Dowe, 2003] and we now know that low and medium ability chess systems can be built without much intelligence. We are also coming to realize how much of our intelligence is linked to our ability to understand and interact with the natural environment.

Turing testing also cannot measure non-human forms of intelligence. For example, computers are much better at processing vast amounts of data, so they could have much higher levels of intelligence in bioinformatics, while being incapable of solving a Raven's Matrix. It would be extremely anthropocentric to declare that a machine is not intelligent because it cannot solve the narrow range of problems that can be tackled by human intelligence.

To address the limitations of Turing testing, people have developed *universal* measures of intelligence that, in theory, can be applied to any system at all. For example, Legg and Hutter [2007c] developed an algorithm that sums the rewards that an agent receives across all possible environments, with some adjustment for the complexity of different environments. This measure has some intuitive plausibility, but it is not practically calculable because it sums across all possible actions of the agent in all possible environments.

A more practical universal measure of intelligence was put forward by Hernández-Orallo and Dowe [2010], which is based on inductive inference, prediction, compression and randomness. The algorithm is designed to be 'anytime,' which means that whenever it is halted the result should approximate the system's level of intelligence. One issue with Hernández-Orallo and Dowe's approach is that they test the agent in balanced environments in which random actions lead, on average, to zero reward. In practice this is unrealistic because most environments in which intelligence operates are not balanced. If we try to sum intelligence across many unbalanced environments, then the intelligence will be drowned out by random noise. The time limitation of their anytime measure also fails to capture the intelligence of systems that operate on non-human time scales, such as trees and human societies.

Other people have developed universal measures of intelligence based on compression. In the Hutter prize people compete to compress 1GB of Wikipedia data [Hutter, 2021]. Hernández-Orallo's C-test measures the ability of a system to find the best explanation for sequences of increasing

complexity in a fixed time [Hernández-Orallo, 2000]. The best explanation is usually a compressed version of the sequences that enables the subject to predict more sequences of the same type. As explained in Section 2.2, good compressors are good at prediction, so these kinds of tests go some way towards measuring predictive intelligence. However, predictive intelligences are typically specialized for the different areas that the operate in (motor movement, psychology, stock prices, protein folding, etc.), so a single type of compression test only provide one way of estimating predictive intelligence in a limited class of systems.

More detailed summaries of some of these measures of artificial intelligence are given by Legg and Hutter [2007b]. Hernández-Orallo [2017] discusses other ways in which the performance of AI systems can be evaluated.

3.4 Λ: A New Universal Measure of Predictive Intelligence

This section describes a new universal algorithm that I have developed to measure predictive intelligence. This algorithm works by comparing the probability distributions of a system's predictions with the probability distributions that actually occur (see Fig. 2).

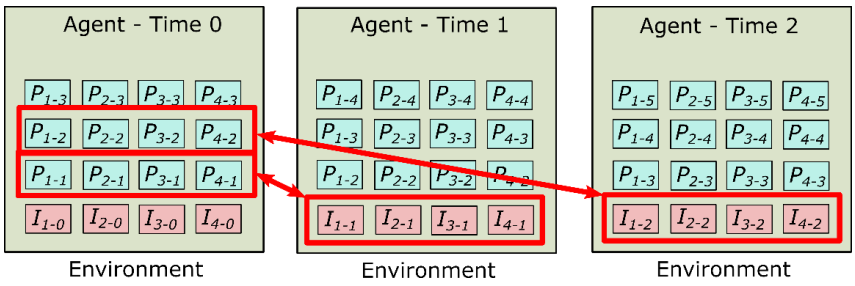

Fig. 2. Agent's predictions about its internal states. The agent has internal states I_1, I_2, I_3 and I_4. $I_{1\text{-}0}$, $I_{2\text{-}0}$, $I_{3\text{-}0}$ and $I_{4\text{-}0}$ are the probability distributions of the internal states at time 0. $P_{1\text{-}1}$, $P_{1\text{-}2}$, $P_{1\text{-}3}$ are predictions that the agent makes about the values of I_1 at times 1, 2 and 3. As the spatial and temporal properties of the environment change, the future states of I_1, I_2, I_3 and I_4 are compared with earlier predictions to evaluate their accuracy.

Prediction Accuracy

To measure the accuracy of a system's predictions we need to compare the probability distributions of the predictions with the later probability distributions of the internal states. For example, in Fig. 2, we compare prediction $P_{1\text{-}2}$ made at time 0 with $I_{1\text{-}2}$ at time 2. For discrete probability distributions the prediction accuracy is measured using Hellinger distance:

$$H(P,Q) = \frac{1}{\sqrt{2}}\sqrt{\sum_{i=1}^{k}\left(\sqrt{p_i} - \sqrt{q_i}\right)^2} \quad (1)$$

Hellinger distance is 0 when there is an exact match between two probability distributions, and 1 when there is a complete mismatch. So 1-*H(P,Q)* gives the *degree of match* between two discrete probability distributions, *P* and *Q*, expressed as a number between 0 and 1.

For continuous probability distributions, the accuracy is measured as the intersection between the probability distribution of the prediction with the probability distribution of the actual value with error bounds (see Fig. 3).

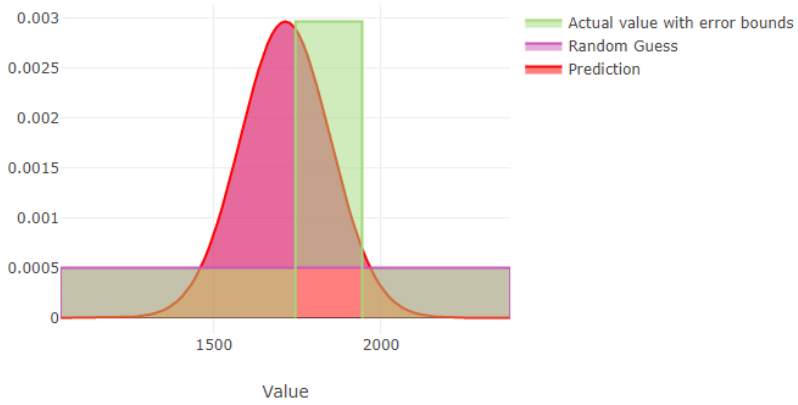

Fig. 3. Prediction match for continuous probability distributions. Here the prediction is a normal distribution. The actual value is discrete and error bounds are added so that there is a non-zero intersection with the prediction. The random guess is an equal distribution across the known range of values.

Eliminating Random Guesses

Systems can make random guesses about their future states. For example, suppose that I_1 in Fig. 2 can have discrete values 1, 2 or 3. In this case, a random guess would be $p(I_1 = 1) = 0.3333$, $p(I_1 = 2) = 0.3333$ and $p(I_1 = 3) = 0.3333$. This guess does not require a significant amount of intelligence because it is generated without any data from the environment. However, it will still have some match with the states that actually occur at the next point in time. In discrete probability distributions this issue is addressed by subtracting the random guesses from the prediction match, as shown in Eq. 2.

$$PM(P,I) = \left|\left(1 - H(P,I)\right) - \left(1 - H(R,I)\right)\right| \tag{2}$$

PM is the match between a prediction, *P*, and an internal sensory state *I*. *R* is an equal distribution across possible sensor values. With continuous probability distributions, the random guess is interpreted as an equal distribution across the range of previously seen values, as illustrated in Fig. 3. In both cases the absolute value of the result is taken to prevent negative values of intelligence. The prediction match is summed for all predictions that are made at each unique state of the environment.

Trivial and Non-trivial Predictions

Systems could artificially increase their intelligence by generating large numbers of trivial predictions. This is addressed by multiplying the sum of the prediction matches by the Kolmogorov complexity of the predictions, as shown in Eq. 3:

$$PM_e = \frac{K(l)}{L(l)} \sum_{s=1}^{p} \sum_{i=1}^{q} \sum_{t=1}^{r} PM(P_{i-t}, I_{i-t}) \tag{3}$$

PM_e is the total prediction match for a single environment, *l* is a string that describes all the predictions in the environment, *K(l)* is the Kolmogorov complexity of *l*, and *L(l)* is the length of *l*. The predictions are summed for all unique states of the environment (*s=1… s=p*), for all internal states

(*i*=1… *i*=*q*) and for all times covered by the predictions (*t*=1…*t*=*r*). Kolmogorov complexity cannot be directly calculated, so it is often approximated by compression algorithms.

Comparing Simple and Complex Systems

To make it easier to compare systems with large differences in intelligence, the log is taken of the total prediction match for the environment (PM_e). This makes it easier to compare highly complex systems, such as humans, with trivial AI systems on the same scale. In some cases PM_e will be less than 1. This corresponds to a low level of intelligence, but the log of a very small number is a large negative number, which makes no sense for intelligence. So the log is only taken when PM_e is greater than 1. This leads to the final equation for calculating the predictive intelligence, PI_e, of an agent in environment *e*:

$$PI_e = \begin{cases} log_2(PM_e) & if\ PM_e > 1 \\ 0 & otherwise \end{cases} \quad (4)$$

Summing Intelligence across Environments

Predictive intelligence is relative to a *set* of environments (hypothesis H2), so we need to sum PI_e for all the environments in the set. This raises the problem that we don't have a clear definition of what constitutes a distinct environment. Some environments are very different from each other; others only have trivial differences between them. If we simply add the PI_e values from different environments together, an agent will double its intelligence across two environments that are almost identical. To address this issue, the sum of *PI* across environments $E_1, E_2, \ldots E_n$ is multiplied by the Kolmogorov complexity of the combined environments divided by the sum of the Kolmogorov complexity of the environments considered independently:

$$\aleph_c^{1.0} = \frac{K(E_1 + E_2 + \cdots + E_n)}{K(E_1) + K(E_2) + \cdots + K(E_n)} \sum_{e=1}^{p} PI_e \quad (5)$$

Environments appear very differently to systems with unique senses and diverse ways of processing sensory data. So the complexity of environments has to be considered from the perspective of the agents, not from some fictional standpoint of complete objectivity. If two environments are very similar, then the joint complexity will be approximately the same as the individual complexity and the first half of Eq. 5 will be approximately 1/2. On the other hand, if two environments are very different, then the shortest program describing both will be approximately equal to the sum of the lengths of the shortest programs describing the environments individually. In this case the first half of Eq. 5 will be approximately 1. In practice, Kolmogorov complexity is calculated using compression algorithms, which need to be carefully chosen to take the nature of the environments into account.

The letter that I have chosen to represent this measure of intelligence is ᛈ, which is the Old Norse letter (rune) that corresponds to our modern 'p' sound. ᛈ is pronounced 'peorth', 'perth' or 'pertho'. The ᛈ rune is associated with the dice cup, chance, secrets, destiny and the future, which is appropriate for a measure that is based on a system's ability to make predictions. In Eq. 5, the subscript, *c*, indicates that it is crystallized predictive intelligence. The superscript is the version of the algorithm.

Fluid Intelligence

Fluid intelligence corresponds to a system's ability to learn and adapt to its environment. It is linked to positive changes in a system's crystallized intelligence (see hypothesis H4), as shown in Eq. 6:

$$ᛈ_f^{1.0} = \begin{cases} \frac{dᛈ_c^{1.0}}{dt} & if \ \frac{dᛈ_c^{1.0}}{dt} > 0 \\ 0 & otherwise \end{cases} \quad (6)$$

Experiments

The feasibility and performance of the algorithm have been tested on an agent in a variety of maze environments, and on a deep neural network that performs time series prediction. The experiments are implemented as a website: www.davidgamez.eu/pi, which is shown in Fig. 4.

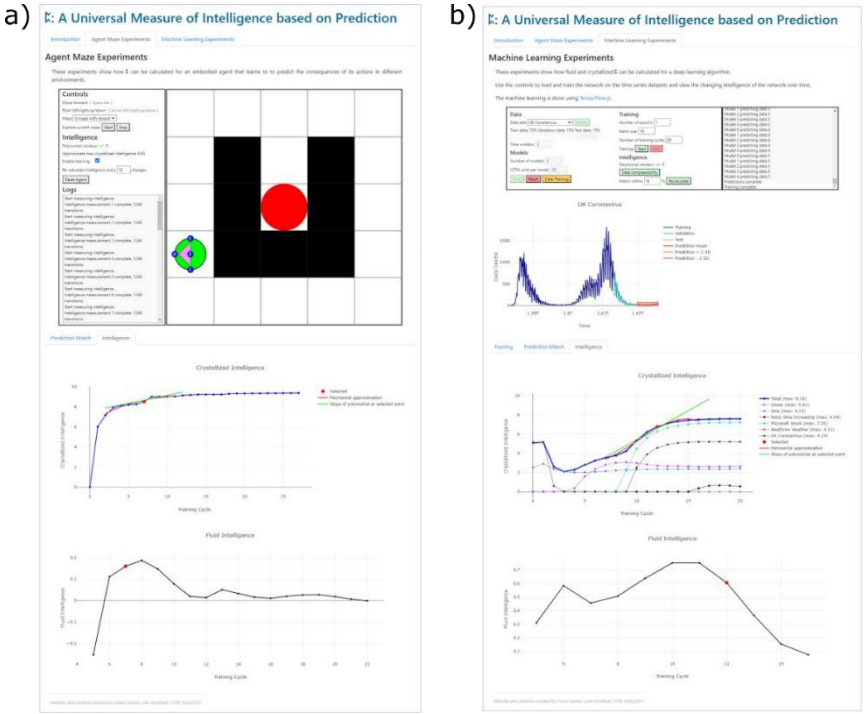

Fig. 4. Website with experiments that test 𝓚 algorithm. a) Fluid and crystallized 𝓚 are calculated for an agent that predicts the consequences of its actions in different maze environments. b) Fluid and crystallized 𝓚 are calculated for a deep network that predicts future values in different time series (synthetic, stock prices, weather and coronavirus).

These experiments show that 𝓚 is straightforward to measure on artificial systems when we have full access to their internal states and the environments can be fully explored. More work is required to develop ways of estimating the predictive intelligence of less accessible artificial systems that partially explore their environments.

We have very limited access to natural systems' internal states. Brain activity can be read non-invasively using fMRI, MEG, electrodes and EEG, but these technologies have very low spatial and/or temporal resolution. Optogenetics can give us close to real time measurements of the entire brain of small transparent organisms, such as the Zebrafish larvae [Portugues et al., 2013], and it might be possible to apply this to

other transparent animals, such as the glass octopus. With non-transparent animals we can only measure ~20,000 neurons on the surface of the brain in real time with current technology. Ways will have to be found to estimate the predictive intelligence of these systems from limited data and from external behavior. We will then be able to use ᛕ to systematically study and compare the intelligence of humans, non-human animals and artificial systems.

4 Consciousness

4.1 Definition of Consciousness

When we are conscious we are immersed in a bubble of space, roughly centered on our bodies, within which objects and non-physical properties, such as color and smell, are distributed. I describe this as a *bubble of experience* [Gamez, 2018]. My bubble of experience currently contains green trees and the smell of coffee. When I am at the beach my bubble of experience contains white sand, blue sea and the taste of tequila. In online perception objects and properties in our bubbles of experience co-vary with the physical world. We can also change our conscious experiences offline, independently of the world, in dreams and imagination.

Bubbles of experience have multiple dimensions of variation. The spatial size of bubbles of experience can vary, there is variation in temporal depth [Husserl, 1964] and there can be more or less objects and properties and more or less types of objects and properties. The contents of bubbles of experience can also appear with different levels of intensity. In dreams, imagination and on the edge of sleep, contents are vague, washed out and unstable. In online perception contents are vivid and stable with rich colors. A person on hallucinogens can have experiences with greater intensity than the normal waking state. The contents of a single experience can have range of intensities. There might be a fleeting impression of a bird at the edge of my field of vision while I am looking at a bright red bus rushing towards me and experiencing intense feelings of fear and panic.

There are challenging philosophical problems with consciousness, such as the hard problem and the relationship between consciousness and the physical world. Elsewhere I have shown how our modern concept of consciousness (and some of its problems) co-evolved with the development of modern scientific theories about the physical world [Gamez, 2018].

4.2 Physical, Computational, Functional and Informational Theories of Consciousness

Physical theories of consciousness link consciousness to spatiotemporal patterns in particular physical materials. For example, there are neural theories of consciousness [Koch et al., 2016], electromagnetic theories of consciousness [Pockett, 2000] and quantum theories of consciousness [Hameroff & Penrose, 1996]. Physical theories of consciousness are similar to other scientific theories that are based on spatiotemporal physical patterns: moving electrons produce magnetic fields; moving neutrons do not.

Many people believe that consciousness is linked to computations or functions [Cleeremans, 2005]. They claim that consciousness is present wherever a particular computation or function is executed, independently of how the computation or function is implemented. For example, people have connected consciousness with the implementation of a global workspace [Dehaene, 2014]. Information integration theory connects patterns of information to consciousness, independently of the physical implementation of the information [Tononi, 2008].

Physical, computational and functional theories of consciousness have some common ground. It might be the case that global workspace theory, for example, captures a pattern, which is linked to consciousness when it is implemented in a biological brain. However, computational and functional theories of consciousness lose plausibility when the claim is made that a computation or function is linked to consciousness *independently* of the material in which the computation or function is realized. One problem with this claim is that a system executing a computer program is a sequence of physical states, and Putnam [1988] and Bishop [2009] show that any sequence of physical states can be interpreted

as implementing a particular run of a given computation. This leads to an implausible panpsychism and to the untenable result that every brain is associated with an infinite number of different consciousnesses.

A second problem with computational and functional theories of consciousness is that they can only be scientifically tested if we have an objective way of measuring the presence or absence of a computation or function in a system. For example, to prove that global workspace theory is correct, we need to be able to determine whether there is an active global workspace in the conscious brain and show that no global workspaces are being executed in the unconscious brain. Unfortunately we do not have a way of unambiguously measuring the computations or functions that are being executed in a physical system [Gamez, 2014a]. Information integration theory has similar problems with the subjectivity of information and with the measurement of information in a system [Gamez, 2016]. The only reasonable conclusion is that computations, functions and information are subjective - not objectively measurable properties of physical systems. Consciousness must be linked to objective physical properties of a system.

4.3 Measurement of Consciousness

To study the relationships between consciousness and the physical world we need to measure consciousness, measure the physical world and look for connections between these sets of measurements.

Consciousness is measured through first-person reports. For example, when I am eating an apple, I can describe its red color and sweet flavor. We believe our own first-person reports and typically believe the first-person reports of other adult humans (disregarding philosophical problems with zombies, etc.), because other people have similar brains and we assume that there is a close relationship between physical and conscious states.

Infants, animals, and computer programs also generate first-person reports about consciousness. However, these systems have different brains or no brains at all, so the assumption about physical similarity no longer holds. So we cannot completely trust what infants, animals and robots say about their internal conscious states. This makes them unsuitable subjects

for identifying the relationships between consciousness and the physical world [Gamez, 2014b].

First-person reports about consciousness from normal adult humans can be combined with measurements of the brain to identify neural correlates of consciousness [Koch et al., 2016]. We can use this data to develop mathematical descriptions of the relationships between consciousness and the physical world. These theories generate descriptions of consciousness from descriptions of physical states, and vice versa, as shown in Fig. 5.

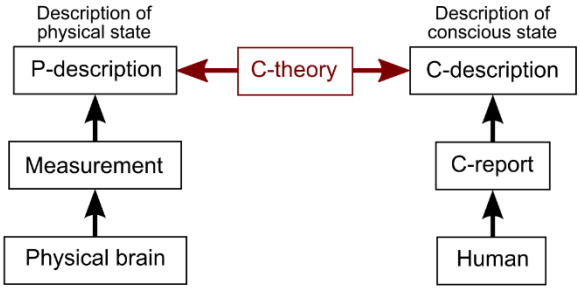

Fig. 5. A mathematical theory of consciousness (c-theory) describes the relationship between physical and conscious states. It can generate a description of consciousness from a description of a physical state and generate a description of a physical state from a description of consciousness.

Tononi's information integration theory (IIT) is an example of a mathematical theory of consciousness [Tononi, 2008]. However, IIT is based on subjective information [Gamez, 2016] and only performs a one way mapping from information to conscious states.

When a mathematical theory of consciousness has been judged to be a reliable way of mapping between physical and conscious states, we can use it to make *deductions* about the consciousness of animals and artificial systems. These are deductions, rather than predictions, because they cannot be confirmed by measuring consciousness through first-person reports (see Gamez [2018]). Fig. 6 illustrates how a mathematical theory of consciousness can be used to make a deduction about the consciousness of an artificial system.

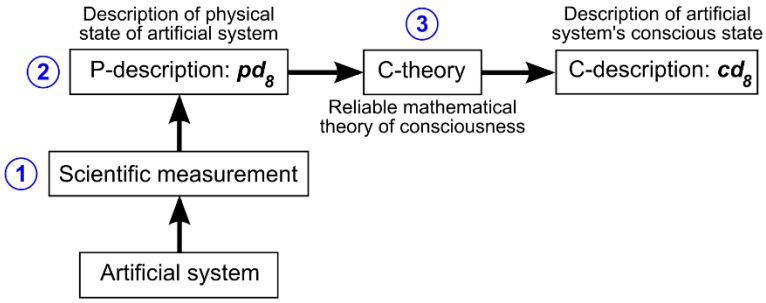

Fig. 6. Deduction of the consciousness of an artificial system. 1) Scientific instruments are used to measure the physical state of the artificial system. 2) Physical measurements are converted into a description of the artificial system's physical state. 3) Reliable mathematical theory of consciousness (c-theory) converts the physical description into a description of the artificial system's consciousness.

4.4 Artificial Consciousness

A substantial amount of academic research has been carried out on artificial consciousness and working systems have been built to explore different aspects of this topic. Public awareness has been raised through films and TV series, such as Chappie, Ex Machina, Altered Carbon and Black Mirror [Gamez, 2020]. Artificial consciousness is a complicated field that can be broken down into four overlapping areas [Gamez, 2018]:

- **MC1**. *Machines with the same external behavior as conscious systems*. Humans behave in particular ways when they are conscious. For example, they are alert, they can respond to novel situations, they can inwardly execute sequences of problem-solving steps and they can learn. MC1 machine consciousness is the creation of AI systems that exhibit some or all of these external behaviors. Watson [Ferrucci, 2012] is an example of a MC1 system that mimics the external behavior of conscious humans when they are playing Jeopardy.
- **MC2**. *Models of the correlates of consciousness*. Theories about the neural and functional correlates of consciousness in humans can be modeled in a computer. For example, global workspace

implementations have been used to control a naval dispatching system [Franklin, 2003] and a video game avatar [Gamez et al., 2013].

- **MC3**. *Models of consciousness*. Phenomenal experiences have characteristic features that can be modeled in computers and used to control robots. One example of this type of system was developed by Chella et al. [2007], who used a virtual environment (analogous to the robot's consciousness) to control a museum guide robot. Gravato Marques and Holland [2009] built a system in which a robot used a simulation of itself to solve a motor control problem and executed the solution with its real body.
- **MC4**. *Machines that are phenomenally conscious*. When humans are conscious they are immersed in a bubble of experience that contains colors, smells, sounds, etc. (see Sec. 4.1). A machine that was immersed in a bubble of experience, which contained something similar to our colors, smells and sounds, would be MC4 conscious. MC4 consciousness will only be fully solved when we have discovered a mathematical theory of consciousness that can reliably map between physical and conscious states (see Sec. 4.3). We have no idea whether any of our current machines are MC4 conscious.

These categories are not exclusive: systems can implement several of them at the same time. For example, a robot based on the neural correlates of consciousness (MC2) could be phenomenally conscious (MC4) and exhibit conscious external behavior (MC1).

5 Relationships Between Intelligence and Consciousness

5.1 Natural Intelligence and Natural Consciousness

Intelligence is a *functional* property: the amount of intelligence in a system is independent of the way in which it is implemented. In Sec. 4.2 I outlined good reasons for thinking that consciousness is linked to spatiotemporal patterns in specific physical materials. Intelligence and consciousness can overlap when the implementation of the intelligence functions produces

spatiotemporal physical patterns (for example, neuron firing patterns) that are correlated with consciousness.

While there has been a substantial amount of work on the neuroscience of intelligence [Haier, 2017] and on the neural correlates of consciousness [Koch et al., 2016], we do not know enough about either to be able to say whether the brain's implementation of the functions linked to intelligence are the same as the neural correlates of consciousness. The best we can say is that some of the functions that have been proposed to be linked to consciousness in the brain are also likely to be linked to intelligence. For example, Aleksander and Dunmall [2003] claim that depiction, imagination, attention, planning and emotion are minimally necessary to support consciousness. These functional properties are clearly connected with intelligence – for example, we need imagination to do IQ tasks, such as Ravens' matrices, and planning is related to predictive intelligence and goal achievement. Other people have hypothesized that the brain's implementation of a global workspace is connected with its consciousness [Dehaene, 2014]. Global workspace theory has been shown to be good way to implement AI systems [Franklin, 2003; Gamez et al., 2013], so if global workspace theory is a correct theory of consciousness, then the brain's implementation of a global workspace is likely to be linked to its intelligence. While the exact relationship between prediction and consciousness is an open question, there is clearly a lot of non-conscious prediction going on in the brain, so there is unlikely to be exact alignment between the brain's predictive abilities and its consciousness. More abstract theories about consciousness, such as higher order thought [Rosenthal, 1986], recurrent processing [Maia & Cleeremans, 2005] and information integration theory [Tononi, 2008] point to brain mechanisms that might also be involved in intelligence. For example, a brain that can integrate more information (possibly using recurrent connections) and which contains meta-information about its internal states is likely to be more intelligent. Intelligence can be implemented in many different ways, so there is unlikely to be a strong relationship between the spatiotemporal patterns linked to consciousness and the intelligence functionality of the brain.

Weak inferences can also be made from phenomenological observations about consciousness to the potential intelligence of a system.

This connection is weak because most of the data and functions that produce intelligence are not consciously experienced. For example, when an idea spontaneously appears to me, I typically lack insight into the exact mechanisms by which it was arrived at, presumably because it was the result of unconscious processing. However, some of our reasoning is carried out consciously using imagination. With this type of reasoning, a consciousness with more contents could potentially solve more problems, achieve more goals and generate more predictions. So we might have weak grounds for believing that a system with more conscious contents has greater potential for intelligence. This is only a weak inference because there could be systems with rich states of consciousness that are not capable of intelligent behavior, and an impoverished binary consciousness, which could only contain 1 or 0, could potentially create every single document that has ever been written by humans. While the intensity of conscious contents plays a role in tagging states as online or offline, this does not appear to be strongly linked to intelligence.

5.2 Artificial Intelligence and Artificial Consciousness

The relationships between artificial intelligence and artificial consciousness vary with the type of artificial consciousness.

MC1 machines behave in a similar way to conscious humans. Many external behaviors linked to consciousness are also linked to intelligence, and most of the behaviors that we judge to be intelligent in humans can only be carried out consciously. So there is likely to be a close relationship between progress in MC1 machine consciousness and progress in artificial intelligence. As machines mimic more human behaviors, they will appear to be more conscious and more intelligent. However, there is also a potential dissociation between MC1 machine consciousness and AI. Machines could implement forms of intelligence that achieve low IQ or g-scores on human test batteries, but score highly on universal measures of intelligence. These highly intelligent machines might not exhibit any conscious human behaviors.

MC2 and MC3 machine consciousness research uses models of the correlates of consciousness and models of consciousness to produce more intelligent machines. This has already led to the development of systems

that exhibit human-like intelligence [Gamez et al., 2013] and intelligent navigation of a museum environment [Chella et al., 2007]. In the future, MC2 and MC3 research is likely to lead to more advanced forms of artificial intelligence. However, AI is a very diverse field and MC2 and MC3 are only two ways of building intelligent machines. A large number of other AI approaches, such as deep neural networks, can be used to develop intelligent systems, and these have few connections to research on consciousness.

We know almost nothing about the MC4 consciousness of artificial systems. It is possible that some of our current AI systems have conscious states that are as rich and vivid as our own. It is also possible that consciousness is only linked to systems that implement certain functions in something approximating biological hardware. Since consciousness is not a purely functional property and a given piece of intelligent behavior can be implemented in an infinite number of different ways [Putnam, 1988], there is not a necessary connection or nomological law linking intelligence and MC4 consciousness. The amount of overlap between MC4 machine consciousness and AI is an *empirical* question that can only be answered we have a reliable mathematical theory of consciousness and a practical universal measure of intelligence that does not depend on batteries of anthropocentric tests.

6 Conclusions

Many overlapping definitions of intelligence have been put forward, which are mostly based on human intelligence and generalize poorly to non-human animals and artificial systems. To address this issue, this paper has put forward four hypotheses that define intelligence in terms of prediction.

Progress has been made with the measurement of intelligence in natural systems and many scientists believe that g-score correlates with intelligence in humans and some animals. However, the test battery approach that is used to measure IQ and g-score in natural systems is unlikely to be generalizable to the wide variety of behaviors and intelligences of artificial systems. One solution to this problem is to design

tests that only measure human-like intelligence - in the AI context this is known as Turing testing. Another approach is to design universal intelligence measures that can be applied to any system at all, such as the prediction-based measure that was outlined in Sec. 3.4.

Most people agree that consciousness is the stream of colorful, noisy smelly experiences that start when we wake up in the morning and cease when we fall unconscious at night. In my own work I have described this as a bubble of experience. Many of the philosophical problems with consciousness can be neutralized with assumptions that provide a reasonable starting point for the scientific study of consciousness [Gamez, 2018]. These assumptions enable us to measure consciousness through first-person reports in normal adult humans. We can then carry out experiments that measure consciousness, measure the physical world and identify relationships between these two sets of measurements. Scientific research has already made considerable progress identifying some of the neural correlates of consciousness. In the future, we need to discover mathematical descriptions of the relationships between consciousness and the physical world. These can be used to make deductions about the consciousness of non-human animals and artificial systems.

Intelligence is a purely functional property; consciousness is not, so there cannot be a strong connection between consciousness and the many different ways in which intelligence can be implemented in artificial and natural systems. In natural systems, the spatiotemporal physical patterns linked to consciousness might overlap with the brain's implementation of intelligence. Weak inferences can also be made from the richness and structure of natural consciousness to the potential intelligence of a system. In artificial systems there is a reasonably close connection between MC1 machines and machines that exhibit human-like intelligent behavior. MC2 and MC3 technologies can be good ways of building more intelligent machines that think in a similar way to humans.

At the present time we do not have the theories or the data to make stronger conclusions about the relationships between intelligence and consciousness. We will be able to systematically study this relationship when we have a practical universal measure of intelligence, which can be applied to natural and artificial systems, and a reliable mathematical theory

of consciousness that can map between physical descriptions and descriptions of conscious states.

References

Aleksander, I. and Dunmall, B. [2003] Axioms and Tests for the Presence of Minimal Consciousness in Agents, *Journal of Consciousness Studies* **10**(4-5), 7-18.

Baker, M. [2016] 1,500 Scientists Lift the Lid on Reproducibility, *Nature* **533**(7604), 452-454.

Bartholomew, D. J. [2004] *Measuring Intelligence: Facts and Fallacies* (Cambridge University Press, Cambridge).

Bell, T. C., Cleary, J. G. and Witten, I. H. [1990] *Text Compression* (Prentice-Hall, Englewood Cliffs).

Bishop, J. M. [2009] A Cognitive Computation Fallacy? Cognition, Computations and Panpsychism, *Cognitive Computation* **1**, 221-233.

Bowers, J. S. and Davis, C. J. [2012] Bayesian Just-So Stories in Psychology and Neuroscience, *Psychological Bulletin* **138**(3), 389-414.

Boysen, S. T. and Capaldi, E. J. [1992] *The Development of Numerical Competence : Animal and Human Models* (L. Erlbaum Associates, Hillsdale, N.J.).

Burkart, J. M., Schubiger, M. N. and van Schaik, C. P. [2017] The Evolution of General Intelligence, *Behav Brain Sci* **40**, e195.

Cattell, R. B. [1971] *Abilities: Their Structure, Growth, and Action* (Houghton Mifflin, Boston).

Chella, A., Liotta, M. and Macaluso, I. [2007] Cicerobot: A Cognitive Robot for Interactive Museum Tours, *Industrial Robot: An International Journal* **34**(6), 503-511.

Chollet, F. [2019] On the Measure of Intelligence. *arXiv preprint* arXiv:1911.01547.

Clark, A. [2016] *Surfing Uncertainty: Prediction, Action, and the Embodied Mind* (Oxford University Press, Oxford).

Cleeremans, A. [2005] Computational Correlates of Consciousness, *Progress in Brain Research* **150**, 81-98.

Collaboration, O. S. [2015] Estimating the Reproducibility of Psychological Science, *Science* **349**(6251), aac4716.

Crosby, M., Beyret, B. and Halina, M. [2019] The Animal-Ai Olympics, *Nature Machine Intelligence* **1**, 257.

Dehaene, S. [2014] *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts* (Penguin, New York).

Doya, K., Ishii, S., Pouget, A. and Rao, R. P. N., Eds. [2007]. Bayesian Brain: Probabilistic Approaches to Neural Coding. Cambridge, Mass., MIT.

Ferrucci, D. A. [2012] Introduction to "This Is Watson", *IBM Journal of Research and Development* **56**(3.4), 1:1-1:15.

Franklin, S. [2003] Ida - a Conscious Artifact?, *Journal of Consciousness Studies* **10**(4-5), 47-66.

Gamez, D. [2014a] Can We Prove That There Are Computational Correlates of Consciousness in the Brain?, *Journal of Cognitive Science* **15**(2), 149-186.

Gamez, D. [2014b] The Measurement of Consciousness: A Framework for the Scientific Study of Consciousness, *Frontiers in Psychology* **5**, 714.

Gamez, D. [2016] Are Information or Data Patterns Correlated with Consciousness?, *Topoi* **35**(1), 225-239.

Gamez, D. [2018] *Human and Machine Consciousness* (Open Book Publishers, Cambridge).

Gamez, D. [2019] The Intelligence of Sheep, *Animal Sentience* **25**(27).

Gamez, D. [2020] Consciousness Technology in Black Mirror: Do Cookies Feel Pain?, in *Black Mirror and Philosophy*, edited by D. K. Johnson (Wiley Blackwell, Hoboken), 273-281.

Gamez, D., Fountas, Z. and Fidjeland, A. K. [2013] A Neurally-Controlled Computer Game Avatar with Human-Like Behaviour, *IEEE Transactions on Computational Intelligence and AI In Games* **5**(1), 1-14.

Gardner, H. [2006] *Multiple Intelligences: New Horizons* (Basic Books, New York).

Gravato Marques, H. and Holland, O. [2009] Architectures for Functional Imagination, *Neurocomputing* **72**(4-6), 743-759.

Haier, R. J. [2017] *The Neuroscience of Intelligence* (Cambridge University Press, Cambridge).

Hameroff, S. and Penrose, R. [1996] Orchestrated Reduction of Quantum Coherence in Brain Microtubules: A Model for Consciousness? , *Mathematics and Computers in Simulation* **40**, 453-480.

Harnad, S. [1994] Levels of Functional Equivalence in Reverse Bioengineering: The Darwinian Turing Test for Artificial Life, *Artificial Life* **1**(3), 293-301.

Hernández-Orallo, J. [2000] Beyond the Turing Test, *Jounal of Logic, Language and Information* **9**(4), 447-466.

Hernández-Orallo, J. [2017] Evaluation in Artificial Intelligence: From Task-Oriented to Ability-Oriented Measuremen, *Artificial Intelligence Review* **48**(3), 397-447.

Hernández-Orallo, J. and Dowe, D. L. [2010] Measuring Universal Intelligence: Towards an Anytime Intelligence Test, *Artificial Intelligence* **174**, 1508-1539.

Hingston, P. [2009] A Turing Test for Computer Game Bots, *IEEE Transactions on Computational Intelligence and AI In Games* **1**(3), 169-186.

Humphreys, L. G. [1979] The Construct of General Intelligence, *Intelligence* **3**, 105-120.

Husserl, E. [1964] *The Phenomenology of Internal Time-Consciousness* Translated by J. S. Churchill. (Martinus Nijhoff, The Hague).

Hutter, M. [2021]. The Hutter Prize. Retrieved 12/11/21, 2021, from http://prize.hutter1.net/.

Knill, D. C. and Pouget, A. [2004] The Bayesian Brain: The Role of Uncertainty in Neural Coding and Computation, *Trends Neurosci* **27**(12), 712-719.

Koch, C., Massimini, M., Boly, M. and Tononi, G. [2016] Neural Correlates of Consciousness: Progress and Problems, *Nat Rev Neurosci* **17**(5), 307-321.

Legg, S. and Hutter, M. [2007a]. A Collection of Definitions of Intelligence. *Proceedings of Advances in Artificial General Intelligence Concepts, Architectures and Algorithms: Proceedings of the AGI Workshop 2006*, edited by B. Goertzel and P. Wang, IOS Press, pp. 17-24.

Legg, S. and Hutter, M. [2007b] Tests of Machine Intelligence, in *50 Years of Artificial Intelligence*, edited by M. Lungarella, F. Iida, J. Bongard and R. Pfeifer (Springer, Berlin, Heidelberg), 232-242.

Legg, S. and Hutter, M. [2007c] Universal Intelligence: A Definition of Machine Intelligence, *Minds and Machines* **17**, 391-444.

Maia, T. V. and Cleeremans, A. [2005] Consciousness: Converging Insights from Connectionist Modeling and Neuroscience, *Trends in Cognitive Sciences* **9**(8), 397-404.

Marino, L. and Merskin, D. [2019] Intelligence, Complexity, and Individuality in Sheep, *Animal Sentience* **25**(1), 1-26.

Neff, G. and Nagy, P. [2016] Talking to Bots: Symbiotic Agency and the Case of Tay, *International Journal of Communication* **10**, 4915-4931.

Neisser, U., Boodoo, G., T. J. Bouchard, J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg,

R. J. and Urbina., S. [1996] Intelligence: Knowns and Unknowns, *American Psychologist* **51**(2), 77-101.

Pockett, S. [2000] *The Nature of Consciousness: A Hypothesis* (Writers Club Press, San Jose, California).

Portugues, R., Severi, K. E., Wyart, C. and Ahrens, M. B. [2013] Optogenetics in a Transparent Animal: Circuit Function in the Larval Zebrafish, *Curr Opin Neurobiol* **23**(1), 119-126.

Putnam, H. [1988] *Representation and Reality* (MIT Press, Cambridge, Massachusetts; London).

Robertson, K. F., Smeets, S., Lubinski, D. and Benbow, C. P. [2010] Beyond the Threshold Hypothesis: Even among the Gifted and Top Math/Science Graduate Students, Cognitive Abilities, Vocational Interests, and Lifestyle Preferences Matter for Career Choice, Performance, and Persistence, *Current Directions in Psychological Science* **19**(6), 346-351.

Rosenthal, D. M. [1986] Two Concepts of Consciousness, *Philosophical Studies* **49**(3), 329-359.

Sanghi, P. and Dowe, D. L. [2003] A Computer Program Capable of Passing Iq Tests. 4th Intl. Conf. on Cognitive Science (ICCS'03). Sydney**:** 570-575.

Shaw, R. C. and Schmelz, M. [2017] Cognitive Test Batteries in Animal Cognition Research: Evaluating the Past, Present and Future of Comparative Psychometrics, *Animal Cognition* **20**, 1003-1018.

Sternberg, R. J. [1985] *Beyond Iq: A Triarchic Theory of Human Intelligence* (Cambridge University Press, Cambridge).

Thurstone, L. L. [1938] *Primary Mental Abilities* (University of Chicago Press, Chicago).

Tononi, G. [2008] Consciousness as Integrated Information: A Provisional Manifesto, *Biological Bulletin* **215**(3), 216-242.

Turing, A. [1950] Computing Machinery and Intelligence, *Mind* **59**, 433-460.

Warwick, K. [2000] *Qi: The Quest for Intelligence* (Piatkus, London).