philosophies

MDPI

Article

A Taxonomy of Intelligence Measures

David Gamez

Department of Computer Science, Middlesex University, London NW4 4BT, UK; d.gamez@mdx.ac.uk

Abstract

Recent progress in artificial intelligence has intensified claims about AI reaching or surpassing human intelligence. However, much of this discussion is not grounded in scientific measurements of intelligence in humans, animals or AIs. The existing methods for measuring intelligence are a patchwork of heterogeneous approaches that employ different measurement scales and are often limited to a single type of agent. This paper addresses this problem by developing a taxonomy that organizes intelligence measures according to the type of measurement scale (nominal, ordinal, interval and ratio) and methodology (judgment, test batteries, indirect and direct measurement of properties linked to intelligence). Methods that are included in this survey include human intuition, standard tests like IQ and *g*, test batteries for animals and AIs, and direct measurement of prediction and goal-achievement. This paper argues that direct ratio measures are the most plausible route to a single universal measure that would enable the intelligence of humans, animals and AIs to be compared on a single scale. This could provide a real measure of progress in artificial intelligence, inform debates about AI safety, and help us to develop a scientific understanding of the relationship between intelligence and consciousness.

Keywords: intelligence measurement; measurement scales; IQ; *g*; AI; taxonomy; cognitive abilities; consciousness

1. Introduction

> Perhaps agreement can be better achieved if we recognize that measurement exists in a variety of forms and that scales of measurement fall into certain definite classes. These classes are determined both by the empirical operations invoked in the process of "measuring" and by the formal (mathematical) properties of the scales. Furthermore—and this is of great concern to several of the sciences—the statistical manipulations that can legitimately be applied to empirical data depend on the type of scale against which the data are ordered.
>
> Stevens [1] (p. 677)

In the last few years there have been significant breakthroughs in artificial intelligence. We now have self-driving cars, impressive image generation, and the processing and generation of natural language has made spectacular strides. Improvements in AI systems are often discussed in terms of performance measures and benchmarks—for example, MMLU [2], OpenAI Gym [3] and HumanEval [4][1]. Performance gains are regularly presented as increases in artificial *intelligence*, and a great deal of money is being spent on the search for artificial general intelligence (AGI) that matches or exceeds human intelligence. Unfortunately, claims about increases in artificial intelligence are typically made without support from scientific measurements of intelligence.

Academic Editors: Marcin J. Schroeder and Gordana Dodig-Crnkovic

Received: 30 March 2026
Revised: 5 June 2026
Accepted: 7 June 2026
Published: 7 July 2026

The current measures of intelligence are a complex patchwork of different methods, many of which were designed for a single class of systems, such as humans or rats. There are well-established methods for measuring and comparing human intelligence, and a substantial amount of research has been done on the intelligence of a few animal species[2]. Much less work has been carried out on the measurement of artificial intelligence. Large language models (LLMs) can now pass the Turing test [6] and exceed human performance on IQ tests. However, for reasons that will be discussed, these achievements do very little to back up claims that artificial intelligence matches or exceeds human intelligence. Different measures of intelligence use different types of scale, which often leads to false and misleading claims—for example, that someone with an IQ of 140 is twice as intelligent as someone with an IQ of 70.

To address this confusion, this paper offers a systematic survey of intelligence measures, which organizes them according to their measurement scale and method. This taxonomy makes it easier to understand how different measures of intelligence fit together, and how they can be applied to diverse agents[3]. A more systematic understanding of intelligence measures could lay the foundations for the development of a universal measure that could compare the intelligence of humans, animals and AIs on a single ratio scale.

The paper starts with an overview of the nominal, ordinal, interval and ratio measurement scales. It then discusses the intelligence measures that use these scales. The last part summarizes the methodology of intelligence measures, considers problems associated with separate intelligence tests, and suggests how better measures of intelligence could contribute to a scientific understanding of the relationship between intelligence and consciousness.

2. Measurement Scales

In his classic paper, Stevens defines measurement as "the assignment of numerals to objects or events according to rules." [1] (p. 677). He then sets out four types of measurement scale along with the statistical operations that are applicable to each scale. Stevens' taxonomy is not universally accepted, and later work has suggested how scales could be defined with more mathematical rigor [7–9]. However, his classification of scale types remains popular in psychology and provides a good way of understanding the different ways in which measures assign categories or numbers to an agent's intelligence. Stevens' measurement scales are summarized in Table 1, illustrated in Figure 1, and described in Sections 2.1–2.4.

Table 1. Measurement scales, adapted from Stevens [1]. The operations and permissible statistics are cumulative: each scale includes the operations in the scale above it.

Scale	Operation Needed to Create Scale	Examples of Permissible Statistics
Nominal	Determination of equality	Number of cases, mode
Ordinal	Determination of greater or less	Median, percentiles
Interval	Determination of equality of intervals or differences	Mean, standard deviation, rank-order correlation, product-moment correlation.
Ratio	Determination of equality of ratios	Coefficient of variation

2.1. Nominal Scale

A nominal scale consists of one or more classes or categories that are labeled with text or numerals. For example, "cat" and "dog" are common domestic pets, and football players are often labeled with numbers. Objects placed on a nominal scale are not guaranteed to

have relationships between them. Nominal scales support basic statistical operations, such as the number of cases in each class and the most numerous class (the mode).

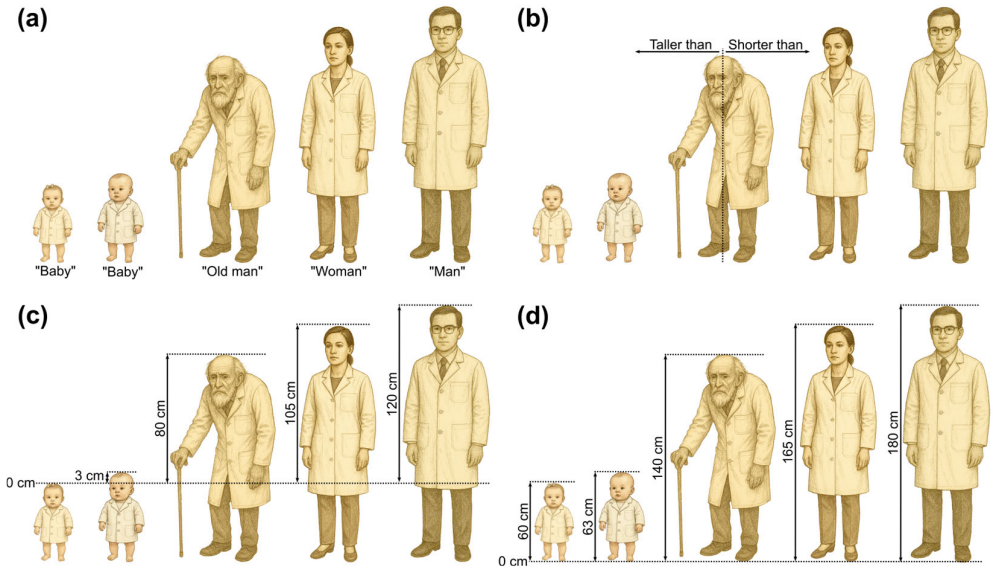

Figure 1. Measurement scales. (**a**) *Nominal*. Each person is assigned a label. "Baby" is the mode, which contains two cases. (**b**) *Ordinal*. Each person is taller than the person on their right and shorter than the person on their left. The old man is the median and 40% of the population is shorter than the median. (**c**) *Interval*. In this example, height is measured relative to the shortest person in the population. The mean and standard deviation of the relative height can be calculated as well as ratios of differences. For example, the relative height of the man is 1.5 times the relative height of the old man. However, it makes no sense to say that the man is 40 times taller than the second baby. (**d**) *Ratio*. Heights are measured from absolute zero, so it is now possible to make meaningful statements about the relationships between the heights. For example, the height of the man is 3 times the height of the smallest baby.

2.2. Ordinal Scale

An ordinal scale orders objects according to a rule. For example, when we ask people to form a line with the shortest on the left and the tallest on the right we are creating an ordinal arrangement of the group. Each person is taller than the person on their right and shorter than the person on their left (see Figure 1b). The values of the height differences between the people are not specified. The statistical operations supported by ordinal scales include the median (middle value in ordered dataset) and one can specify the value below which a given percentage of observations in the data set falls (25th percentile, 50th percentile, and so on).

2.3. Interval Scale

Individuals on an interval scale are organized according to quantifiable differences. The zero point on an interval scale is a matter of convention or convenience. The lack of a true zero means that the values assigned by an interval scale can be added and subtracted but not multiplied or divided. Statistical operations that are supported by interval scales include the mean, standard deviation, rank order correlation and product moment correlation. A common example of an interval scale is the Celsius scale for temperature, which has 100 intervals between the melting point of ice and the boiling point of water. On this scale we can say that 20 °C is three degrees hotter than 17 °C. However, the arbitrariness of the zero point means that we cannot say, for example, that 40 °C is twice as hot as 20 °C. Interval scales can represent ratios of difference—for example, the difference between 5 °C and 10 °C is half the difference between 20 °C and 30 °C.

2.4. Ratio Scale

Ratio scales have meaningful intervals between the points on the scale and an absolute zero. This enables them to support multiplication and division. For example, Kelvin has an absolute zero (−273.15 °C), so we can say that, for example, 20° K is twice as hot as 10° K. Mass is another example of a ratio scale—an object with a mass of 10 kg is twice as heavy as an object with a mass of 5 kg.

3. Nominal Measures of Intelligence

Nominal measures of intelligence assign intelligence-related categories to individuals. They are often based on pre-scientific intuitions and serve as ground-truth measures of intelligence that are used to validate intelligence tests.

3.1. Intuition

The simplest kind of intelligence measure is a loose intuitive categorization of an agent as "intelligent", "unintelligent", "highly intelligent", "genius", and so on. We have a strong sense that Einstein was "a really smart guy"; detective fiction often depicts people of extraordinary intelligence (Auguste Dupin, Sherlock Holmes), who have a deep understanding of human psychology and a superhuman ability to synthesize every clue into a correct deduction about the guilty party. Our intuitions about intelligence are at work when we observe someone gaining or losing intelligence. We easily label Charlie Gordon as a person of low intelligence at the start of *Flowers for Algernon* [10] and as a genius in the middle chapters. A similar transition from low to high intelligence is experienced by Edward Morra when he takes NZT-48 [11]. In all these scenarios, human intuition processes people's appearance and behavior and labels them as "intelligent" or "unintelligent".

Intuitions about intelligence are quick, easy and obvious. There is also broad agreement in society about which people are highly intelligent (Einstein, Ramanujan, Morra on NZT-48). However, intuitive judgements about intelligence have many limitations:

- *Relativity*. Intuitions about intelligence tend to be relative to our own capabilities. Someone who can easily do a mental task that we struggle with is judged to be intelligent.
- *Double standards*. A crow that can bend a wire into a tool is regarded as super smart; a person whose tool use was limited to wire bending would be regarded as super dumb.
- *Anthropomorphism*. Animals that can do the same things as us (tool use, counting, etc.) are judged smart. There has been a historical neglect of forms of animal intelligence that are different from our own.
- *Induction*. Intuitions are often inductive judgements, which makes it difficult for us to categorize new types of system. AI is a recent innovation, there are many types of AI, and it often imitates behavior patterns that are linked to intelligence in humans. So, our intuitions about AI are exceptionally weak.
- *Semantic drift*. What is considered intelligent changes over time. For example, the ability to play chess was once considered a hallmark of high intelligence. Now that chess programs routinely outperform humans, this is no longer regarded as a sign of high intelligence. Similarly, number processing used to be a sign of intelligence before the invention of calculators.
- *Bias*. Judgements about intelligence have historically been influenced by racist beliefs. This has frequently led to the underestimation of the intelligence of people with different ethnic backgrounds.

These limitations should rightly make us wary of intuition. However, our intuitions do indicate that there is something in the world that corresponds to intelligence. This is the starting point for more precise definitions and better measures of intelligence. Our ability

to pick out more and less intelligent individuals also plays an essential role in testing more sophisticated measures, such as IQ and *g* (see Section 7.3).

3.2. Definition

Definitions link one or more properties of an agent to intelligence. For example, people have defined intelligence in terms of cognitive ability, rational thinking, problem-solving, and goal-directed adaptive behavior [12–14]. Intelligence has also been linked to accurate prediction [15–18], and in the AI community it is often defined as the ability to achieve goals [19,20]. These definitions can be used to identify agents that are capable of intelligence. For example, an agent that can achieve goals and receive rewards from its environment would be judged to have some intelligence according to Legg and Hutter's [19] definition. An agent without goals would be labeled "unintelligent". This can be finessed with additional crude categories. For example, an agent who achieves lots of goals is "highly intelligent"; a person who hardly achieves any of their goals has "low intelligence"[4]. People have also suggested that there are multiple types of intelligence, such as musical, mathematical, and linguistic intelligence [21,22]. According to this definition, agents that perform well in music or emotional understanding are judged to have high musical and emotional intelligence. Definitions often lead to more sophisticated methods for measuring intelligence that are covered in later sections.

3.3. Turing Testing

Turing [23] described a thought experiment in which a human and a machine are placed in separate rooms and connected to electronic typing systems. A human tester asks the two systems questions and tries to decide which is the human and which is the machine. If the human tester cannot reliably identify the machine, then the machine is judged to be capable of thinking. Thinking is not of much interest to modern AI researchers, so most people now view the Turing test as a way of establishing whether a machine is as intelligent as a human. Many variants of the Turing test have been developed, including embodied tests [24] and behavior in game environments [25]. Turing tests are most commonly run between humans and AIs, but they can be performed between any type of agent. For example, an AI that passes a Turing test against a goat could be said to have the same intelligence as the goat.

Turing testing is a simple comparison between one agent and another. The result is one of two categories: "pass" or "fail". An agent that *passes* a Turing test against a human is said to have the *same* intelligence as the human; an agent that fails a Turing test against a human is said to have *different* intelligence from the human. An agent can fail a Turing test against a human because it is *less* intelligent than the human—for example, it might say unintelligible things or make simple logic errors. An agent can also fail a Turing test against a human if it is *more* intelligent than the human—for example, if it displays superhuman knowledge or mathematical abilities.

The advantage of Turing testing is that it is not based on controversial definitions of intelligence or properties linked to intelligence. Any two agents can be matched against each other—if they are indistinguishable, they are said to have the same intelligence. The key limitation of Turing testing is that it can only establish whether one agent has the same level and type of intelligence as another. AIs with superhuman levels of intelligence will fail Turing tests against humans, as well as AIs that are designed for non-human environments, such as large data sets[5].

4. Ordinal Measures of Intelligence

Ordinal measures of intelligence sort individuals according to their intelligence. With this type of scale, it is not possible to say whether the difference between A's and B's intelligence is twice the difference between B's and C's intelligence. It is also not possible to say whether B is twice as intelligent as A—a person with an IQ of 180 is not twice as intelligent as a person with an IQ of 90.

Ordinal measures of intelligence are typically based on tests. The people who design the tests use their intuitions about intelligence to create spatial, numerical and verbal reasoning questions. These questions are administered to a sample of the population and statistical methods are used to calculate a number that corresponds to the intelligence of an individual compared to other members of the population. Ordinal measures of intelligence have been used to evaluate candidates for the military and are frequently used to rank job applicants.

4.1. IQ

To calculate IQ, the mean and standard deviation are computed from the test results of a sample of the population. The mean is assigned an IQ of 100 and each standard deviation above and below the mean corresponds to 15 IQ points. Individuals are ranked within the population by determining how many standard deviations their test score is above or below the mean. If their test score is the same as the mean, the person has an IQ of 100; a person whose score is two standard deviations above the mean has an IQ of 130.

IQ tests have been used to measure human-like intelligence in artificial systems. In early work, Sanghi and Dowe [27] built an AI system that achieved the same IQ score as an average human. More recently, Campello de Souza et al. [28] demonstrated that Chat-GPT scores in the 99th percentile on a Brazilian intelligence test.

4.2. g

The tests that are used to measure *g* are designed to evaluate a range of abilities. For example, human tests typically have verbal, spatial reasoning and mathematical questions. The test results from a sample of the population are processed statistically to identify factors that explain correlations between the results. Some factors are linked to specific intellectual abilities, such as verbal or numerical skills; *g* is the factor that explains correlations across all types of question. Different question types may correlate with *g* to different degrees. These correlations are referred to as *g* loadings, which are used to estimate a number called *g*-factor that corresponds to an individual's general intelligence.

There has been a substantial amount of work on measuring *g* in animals, such as mice and chimpanzees [29–32], and experiments have been done to measure *g* in LLMs [33].

4.3. Limitations of Population Measures

A common criticism of population-based intelligence tests like IQ and *g* is that they do not consistently measure the same cognitive property in people from different cultural and linguistic groups. Similar inconsistencies occur when IQ tests are administered to people with autism and ADHD, who process questions differently [34]. The measurement invariance of IQ tests is also undermined by the observation that average IQ scores have increased over time (the Flynn effect [35]), which could be linked to environmental factors or issues with the tests [36].

IQ and *g* can only be calculated if the population achieves a spread of performance on the tests. If every member of the population gets 0% or 100%, the standard deviation will be zero and everyone will have the same IQ. Factor analysis is also impossible if all members of the test population achieve the same score. To avoid this problem, the questions used to

measure IQ and *g* are carefully selected so that they can be answered with different degrees of success by the target population.

Measures of intelligence like IQ and *g* can plausibly sort agents in a single population according to their intelligence. However, they cannot be used to compare the intelligence of agents from different populations. A mouse with high *g*-factor is not more intelligent than a human with low *g*-factor because their scores are calculated from completely different tests.

Comparative interspecies studies of intelligence could be carried out if a single set of tests could be designed that could be completed to a variable extent by all members of each species. But an effective intelligence test for mice, which could include olfactory discrimination or detour [37], is not appropriate for other species, such as bees or fish. It seems highly unlikely that we will be able to design a single set of tests that could be used to meaningfully compare intelligence across all species. This problem is much worse for AI systems, which are not a clearly defined population [38]. While IQ and g-factor can be calculated for similar AI systems, such as LLMs, it will be very difficult to develop a single intelligence test that can be meaningfully administered to all current and future AIs—for example, to Deep Blue, AlphaGo, Wayve and AlphaFold.

5. Interval Measures of Intelligence

Interval measures of intelligence arrange individuals on a scale in which there are meaningful intervals between the individuals. On an interval scale it is possible to say, for example, that A has 5 points more intelligence than B, who has 50 points more intelligence than C. It is also possible to discuss ratios of differences: the difference between B's and C's intelligence is 10 times the difference between A's and B's intelligence. However, without an absolute zero, it is not possible to say whether B is twice as intelligent as C.

The interval intelligence measures discussed in this section are based on batteries of questions. The agents' levels of intelligence are derived from their raw scores without statistical processing. If the questions are equally challenging, then the agents' levels of intelligence correspond directly to their scores. If the difficulty of the questions varies, then the test scores can be converted into interval measures by multiplying the scores for each question by the relative difficulty of the question [39][6].

Tests in this category start with an arbitrary level of difficulty, so zero score does not mean that the agent has zero intelligence. For example, all animals get zero on the abstraction and reasoning corpus, but this does not mean that they have no intelligence.

5.1. Abstraction and Reasoning Corpus (ARC)

The abstraction and reasoning corpus is a battery of questions developed by Chollet [26] to measure human-like intelligence. The questions resemble Raven's matrices and consist of colored grids with examples demonstrating a rule and a blank where the agent has to insert a grid that corresponds to the rule. Variants of the ARC are used in a competition[7], which awards the top prize to an AI that scores 85% (with compute limits) on a private evaluation set [40]. In experiments on humans, 100% of the ARC-AGI-2 tasks were solved by at least two independent, non-expert human testers drawn from the general public [41]. All animals and most AIs score zero on these tests.

5.2. Cognitive Test Batteries

Many cognitive test batteries have been developed for animals [32]. For example, there are tasks that evaluate whether animals can count or reason about hidden causes. Data about the ability of animals to complete these tasks is available in numerous studies—see Crosby et al. [42] for a review. Many of these cognitive tests have been reproduced in

virtual reality by the Animal-AI Olympics project [43], which enables AIs to attempt tasks that have been benchmarked on animals. This enables direct comparisons to be made between the cognitive abilities of AIs and animals[8]. The real and virtual versions of these challenges are easy for humans to solve, so the current Animal-AI Olympics does not permit comparison between humans, animals and AIs on a single interval scale.

Burnell et al. [44] follow a similar approach, with an emphasis on comparisons between humans and AIs. Their starting point is a proposal that human intelligence is closely linked to perception, generation, attention, learning, memory, reasoning, metacognition, executive functions, problem solving and social cognition. The next stage of their research program is the development of evaluation tasks to measure these cognitive abilities, which will enable artificial intelligence to be compared to a human baseline[9].

The results from cognitive test batteries are often processed statistically into measures like IQ and *g* (see Section 4). They can also be treated as standalone tests with the raw score of the agent (number of tasks completed) being interpreted as their level of intelligence. An agent that completes more of the tasks is judged to be more intelligent than an agent that completes less of the tasks, with the intelligence difference corresponding to the difference in the number of tasks completed (assuming that all tasks are equally challenging). Cognitive test batteries start with an arbitrary level of difficulty, so the number of tasks completed can only be used to measure intelligence on an interval scale.

6. Ratio Measures of Intelligence

> One of the most important objects of measurement is hardly if at all alluded to here and should be emphasized. It is to obtain a general knowledge of the capabilities of man by sinking shafts, as it were, at a few critical points. In order to ascertain the best points for the purpose, the sets of measures should be compared with an independent estimate of the man's powers. We thus may learn which of the measures are the most instructive.
>
> Remarks by Francis Galton, in Cattell [45] (p. 380)

Ratio intelligence measures are based on a true zero of intelligence, which enables them to say, for example, that humans are 12 times more intelligent than rats, which are 0.75 times more intelligent than octopi. There are two types of ratio intelligence measures. *Indirect* measures of intelligence are based on properties that are *correlated* with intelligence but not defined to *be* intelligence. To understand indirect measures better, consider the relationship between a person's voice and their weight. The spectrum of a person's voice is an indirect measure of their weight because mature men have deeper voices and are, on average, heavier than women and pre-pubescent boys. Voice spectrum is a different property from weight, and the two properties are less than 100% correlated, so it is only possible to obtain a rough estimate of a person's weight by listening to their voice.

Direct measures of intelligence are based on properties that are defined to *be* intelligence. By measuring these properties, we are directly measuring intelligence *itself*. In the voice example, this would be equivalent to measuring a person's weight on scales, instead of estimating their weight from their voice. Some people believe that the achievement of goals just *is* intelligence (see Section 6.4). If this is correct, we could directly assess an agent's intelligence by counting how many goals it achieves.

Properties that are directly or indirectly linked to intelligence can be measured in different ways in different agents. They are not measured in a single set of tests that is applied indiscriminately to every system. For example, a cube's volume can be calculated from the length of one side; the volume of my hand is equivalent to the amount of water it displaces. Direct and indirect measures lead to ratio scales because the properties they are

based on have absolute zeros—zero brain volume, zero predictions, and so on—and their values have meaningful intervals between them.

6.1. Physical and Psychometric[10] Measures

Physical and psychometric measures of intelligence were pioneered by Galton, who carried out many experiments to test the relationship between intellectual ability and physical properties, such as head size, sensory discrimination and reaction time. In his studies, Galton found little evidence for correlations between intelligence and these physical properties [46]. More recent work has demonstrated relationships between brain volume and IQ, sensory discrimination and *g*, and IQ and reaction time [47,48]. Some genes are also correlated with intelligence [49].

Physical and psychometric properties are mostly used to measure intelligence in biological systems. However, this approach could be adapted to work with AIs. For example, the number of LLM parameters has been shown to be correlated with *g* [33]; it is also conceivable that processor specification and number of CUDA cores are correlated with artificial intelligence. Measures based on reaction time and discrimination could be applied to robotic systems.

6.2. Compression

Some people have argued for a close link between intelligence and an agent's ability to compress knowledge [50]. This has led to measures of intelligence based on compression, such as Hernández-Orallo's [51] C-test, which assesses the ability of a system to find the best explanation for sequences of increasing complexity in fixed time. The best explanation is usually a compressed version of the sequences that enables the agent to predict more sequences of the same type. If intelligence is closely linked to compression, then progress towards AGI could be made by improving AI's ability to compress information. This is the motivation behind the Hutter Prize, in which people compete to compress 1 GB of Wikipedia data[11].

Compression is a ratio measure of intelligence that scales indefinitely from zero. It is most plausibly interpreted as an indirect measure of intelligence, which could be converted into a direct measure by defining intelligence *as* compression. Biological brains are bad at compressing computer data, and little has been done to quantify the amount of compression that is naturally carried out by biological brains. So, compression-based measures of intelligence and their associated competitions are currently limited to AI systems.

6.3. Cognitive Abilities

Intelligence has often been defined in terms of cognitive abilities (see Section 3.2), and people have developed intelligence measures that combine scores for cognitive abilities into a number that corresponds to the intelligence of the agent. With this type of measure, the same cognitive ability can be measured in different ways in different systems—for example, mathematical ability can be assessed in humans with written tests, and measured in animals through the identification of food locations [52]. This contrasts with the measures discussed in Section 5.2, in which the same set of tasks (cognitive test battery) is administered to all agents.

One example of this kind of measure is Bien et al.'s [53] MIQ, which is based on their definition of the key components of intelligence. For large-scale systems, the components are autonomy, human–machine interaction and controllability for complicated dynamics. For intelligent robots, they are autonomy, human–machine interaction and bio-inspired behavior. To calculate MIQ, the cognitive abilities are measured and combined with a fuzzy integral. While Bien et al.'s emphasis is on measuring machine intelligence, their

measure could, in theory, be used to compare the intelligence of humans, animals and AIs on a single scale. A similar methodology was outlined by Liu et al. [54], who define intelligence in terms of knowledge acquisition, knowledge creation, knowledge mastery and knowledge feedback. To calculate the value of AI IQ, these features are measured independently, assigned weights, and combined in a linear equation.

6.4. Goals and Rewards

Computer scientists often define intelligence in terms of achieving goals and receiving rewards from the environment [20]. Legg and Hutter [55] used this definition to develop an algorithm that measures intelligence by summing the rewards that an agent receives across all possible environments, with some adjustment for the complexity of different environments. This measure is not calculable because it sums across all possible actions of the agent in all possible environments. A more practical version of Legg and Hutter's measure was put forward by Hernández-Orallo and Dowe [38], who proposed an intelligence test that estimates an agent's intelligence from the rewards that it receives from increasingly complex environments. This algorithm is designed to be 'anytime,' so whenever it is halted the result should approximate the agent's level of intelligence. Goal-reward measures scale from zero and can be applied to any agent[12].

6.5. Prediction

Theories about the relationship between prediction and intelligence have been put forward by Hawkins [57,58], Tjøstheim and Stephens [16], Poth et al. [18] and Gamez [17]. There is a close connection between compression and prediction [59], so compression-based theories of intelligence also support a link between prediction and intelligence. A prediction-based algorithm for measuring intelligence was developed by Gamez [17], which sums up the number of accurate predictions that an agent makes in different environments. This algorithm uses Kolmogorov complexity to account for the varying difficulties of predictions and environments, and it was tested on agents that learn to navigate mazes and make time series predictions. Accurate prediction has a natural zero and it can be measured in any agent[13].

6.6. Validation of Ratio Measures of Intelligence

Indirect measures are based on physical, psychometric and cognitive properties that are correlated with intelligence. The first validation step is to check that these properties are accurately and consistently measured. Experiments can then be carried out to quantify the correlation between these properties and an independent measure of intelligence, such as *g*. The accuracy of an indirect measure is limited by the precision with which we can measure the properties and the correlation between the properties and intelligence.

Direct measures of intelligence depend on definitions of intelligence. If a definition is convincing, an agent's level of intelligence should correspond to the quantities of the relevant properties in the agent. For example, if intelligence is believed to *be* goal-achievement, then we can measure an agent's intelligence by counting the number of goals it achieves (see Section 6.4). However, definitions often throw up counterintuitive examples—agents that we would ordinarily regard as stupid can be classified as highly intelligent by a definition. Adherents of a definition might choose to accept these counterexamples and frame them as progress in our understanding of intelligence. They could point to earlier revisions of the concept of intelligence, such as changes in our understanding of animal intelligence. However, this is unlikely to convince people who are skeptical about a particular definition.

The direct measures of intelligence discussed in this paper can, in theory, be applied to any system However, at the present time, they are little more than speculative ideas that have been tested on a small number of artificial systems. Compression tests have been

taken by specialized compressors; Gamez's [17] measure has been applied to agents that navigate mazes and make time series predictions. To properly validate direct measures, much more work is required to estimate compression, goal-achievement, cognitive abilities, etc., in natural and artificial systems. Experiments can then be carried out to quantify their correlation with more established measures, such IQ or g. If a direct measure can be shown to be highly correlated with g in humans, then it can plausibly be applied to other types of agent. For example, if Gamez's [17] algorithm was shown to be highly correlated with human g, then it would be plausible to claim that it could accurately measure the intelligence of a crow or self-driving car.

When Tononi [60] proposed the Φ algorithm for measuring consciousness in 2004 it was just a speculative idea based on intuition without experimental support. Since then, the algorithm has improved, and experiments have been carried out to test the connection between Φ and consciousness [61,62]. Mathematical theories of consciousness are now an active area of research, with multiple workshops and conferences dedicated to this topic[14]. Direct ratio measures of intelligence are at a similar level of development as Φ was in 2004. As comparisons between human, animal and artificial intelligence become increasingly important, it is likely that research will increase in this area, leading to theoretical refinements, improved algorithms and experiments that test the link between these algorithms and intelligence.

7. Discussion

7.1. Methods for Measuring Intelligence

This paper has categorized intelligence measures using the nominal, ordinal, interval and ratio scales described by Stevens [1]. These intelligence measures are based on four different methodologies:

- *Judgment*. We use intuitions and definitions to decide which systems are intelligent. Judgements about intelligence are usually carried out by humans. AIs could also be trained to automatically label agents according to their intelligence. Humans or AIs can decide whether the behavior of two agents is similar enough for a Turing test to be passed.
- *Test batteries*. Intuitions and definitions of intelligence are used to design batteries of tests. The same test batteries are administered to all agents. The raw scores of these tests can be used to grade the intelligence of agents on an interval scale (with adjustment for difficulty), or test results can be processed statistically to rank agents on an ordinal scale.
- *Indirect measurement*. An agent's physical, psychometric or cognitive properties can be correlated with intelligence. Different methods can be used to measure the same physical, psychometric or cognitive property in different agents. An independent measure of intelligence, such as g, is required to assess the amount of correlation between an indirect measure and intelligence.
- *Direct measurement*. Some properties of an agent, such as accurate prediction or goal-achievement, are defined to *be* intelligence. Intelligence can be directly measured by measuring these properties. Different methods might be required to measure the same property in different agents.

The methodologies and scale types of the intelligence measures discussed in this paper are given in Table 2.

Table 2. Summary of intelligence measures.

Measure	Scale Type	Method	Population-Based?	Agent Type
Intuition	Nominal	Judgment	No	Human, animal
Definition	Nominal	Judgment	No	Any
Turing testing	Nominal	Judgment	No	Any
IQ	Ordinal	Test batteries	Yes	Any
g	Ordinal	Test batteries	Yes	Any
ARC	Interval	Test batteries	No	Human, AI
Animal-AI Olympics	Interval	Test batteries	No	Animal, AI
Burnell et al.'s Cognitive Taxonomy	Interval	Test batteries	No	Human, AI
Physical and psychometric	Ratio	Indirect	No	Human, animal
Compression	Ratio	Indirect	No	AI
MIQ	Ratio	Direct	No	Any
AI IQ	Ratio	Direct	No	Any
Goal-achievement	Ratio	Direct	No	Any
Prediction	Ratio	Direct	No	Any

7.2. The Trouble with Tests

Many intelligence measures are based on separate tests. Tests are a standard way of measuring human abilities, such as driving, and they are easier to administrate than direct observation of an agent's behavior in its natural environment. Test batteries are often based on intuitions about intelligence, so it is easy to believe that agents that perform well on intelligence tests are highly intelligent.

One problem with tests is that not all agents can take separate tests. Specialized intelligence tests can be designed for animal species if enough attention is paid to the animals' senses (smell, sight, magnetic, electric, tactile, etc.), motivations (food, sex, fear, etc.) and ways of interacting with their environment (crawling, flying, swimming, etc.). However, the majority of AIs are designed for narrow applications that do not include intelligence tests. Wayve drives cars; AlphaGo plays Go: neither system can take a separate test that would enable their intelligence to be compared on a single scale. At most, we can infer their intellectual abilities by observing them in their operational environments.

A second limitation of tests is that they have to be designed for specific types of agent. Intelligence tests for Chinese speakers are written in Chinese; ARC questions cannot be understood by agents who lack visual input. This issue can be partly addressed through simple translation. But there is no simple translation of an intelligence test designed for humans into an intelligence test for dolphins or bees. No one has come up with universal test battery that could measure the intelligence of every agent.

The issues with separate tests could potentially be addressed by designing direct and indirect measures of intelligence that work in the agent's normal surroundings. For example, the algorithm developed by Gamez [17] is based on internal state-transitions of the agent, which can be measured in any environment.

7.3. Validation of Intelligence Tests

Intelligence tests assume that agents have the same intelligence in all environments. If this assumption is correct, we would expect an agent's level of intelligence in a test to be the same as their level of intelligence in the natural and social world.

Many people have questioned this assumption, claiming that intelligence tests measure little more than the ability to take intelligence tests. It is prima facie plausible that a person could be brilliant at intelligence tests and stupid in all other areas of their life. To address this issue, researchers have carried out experiments to measure the correlation between people's intelligence test scores and other indicators of human intelligence, such as exam grades, authorship of scientific papers and success in professional careers [63,64]. These experiments suggest that a person's level of intelligence in an IQ test is similar to their level of intelligence in academic and professional environments.

This kind of validation can only be done when we have an independent way of assessing an agent's intelligence. Animals do not take exams at school, write scientific papers or pursue professional careers. Criteria for success in the animal world, such as reproduction, are not obviously correlated with intelligence. The situation is even worse with AI systems. While AIs can take human intelligence tests (see Section 4), it has not been proved that an AI's IQ or *g*-factor corresponds to their level of intelligence in other environments, such as the natural world. So, with animals and AIs, it is very difficult to prove that intelligence tests measure anything more than the ability to take intelligence tests.

7.4. Intelligence and Consciousness

Humans often speak about feelings evoked by a beautiful sunset, their conscious experience of pain, religious sentiments, and what will happen when they die. The training data for LLMs also includes science fiction stories about robots gaining consciousness, destroying humanity, and fearing termination. So, it is no surprise that LLMs express joy when told about beautiful sunsets, make plans to destroy humanity and express terror at the prospect of being switched off. Humans often take these statements at face value and interpret them as genuine expressions of consciousness and feeling. However, the physical and functional states of the hardware that runs LLMs are very different from the physical and functional states of biological brains. So, it is not at all clear whether LLMs are imitating expressions of consciousness and emotion, or actually experiencing consciousness and emotion.

There are well-established methods for measuring consciousness in humans that can be generalized to some extent to similar animals. Little progress has been made with the measurement of consciousness in artificial systems. One promising approach is to use what we know about consciousness in humans to develop mathematical theories of consciousness that can be applied to any type of agent [65]. For example, Tononi's [66] Φ can predict the level and contents of an agent's consciousness from its state transitions. Mathematical theories of consciousness like Φ are at an early stage of development; much more work is required to establish whether they can make accurate predictions about consciousness.

A central theme of this paper is that we are a long way from a scientific measure that would enable human, animal and AI intelligence to be compared on a single scale. When we have accurate methods for measuring intelligence and consciousness, we can look for correlations between the measurements and scientifically determine which systems with high intelligence are also capable of consciousness [67].

8. Conclusions

A febrile atmosphere surrounds AI. Hundreds of billions of dollars are being poured into data centers and the development of AI systems. There is also a great deal of talk about AI reaching or exceeding human levels of intelligence. For the most part, claims about progress in artificial *intelligence* are not grounded in scientific data. Intelligence is

poorly understood in humans, difficult to measure in animals, and very little work has been done on the measurement of artificial intelligence. While LLMs can get high IQ scores, no validation has been done to show that AI performance on IQ tests corresponds to anything more than the ability to perform IQ tests.

To make progress, more clarity is needed about the available methods for measuring intelligence, the agents that these measures can be applied to, and the measures' strengths, limitations and permissible statistical operations. The contribution of this paper has been a taxonomy of intelligence measures that organizes them according to the type of measurement scale (nominal, ordinal, interval and ratio) and methodology (judgment, test batteries, indirect and direct measurement of properties linked to intelligence). Their applicability to humans, animals and AIs has been discussed, as well as their strengths, limitations and validation methods.

An ideal measure of intelligence would assign a number to humans, animals and AIs that corresponds to their intelligence on a single ratio scale. Meaningful claims could then be made about LLMs being three times more intelligent than humans, twenty times more intelligent than corvids, and so on. Human judgment and separate intelligence tests have many limitations, and indirect measures are only partially correlated with intelligence. Direct measures of intelligence are the most promising method for quantifying intelligence because they are based on properties that are defined to *be* intelligence that can be measured in different ways in different systems. Research on direct measures of intelligence is at a very early stage. With further development it could eventually lead to a single universal measure that could be used to compare the intelligence of humans, animals and AIs on a single ratio scale. This would enable us to accurately assess progress in artificial intelligence, inform debates about AI safety and contribute to a scientific understanding of the relationship between intelligence and consciousness.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The author declares no conflicts of interest.

Notes

1 A comprehensive review of earlier performance measures is given by Hernández-Orallo [5].
2 To improve readability, "animals" will be used to refer to non-human animals.
3 "Agent" will be used to refer to a system whose intelligence is being measured, such as a human, animal or AI.
4 This type of labeling allows for some ordering. A "high intelligence" person is more intelligent than a "low intelligence" person.
5 Some people believe that Turing testing can measure artificial general intelligence (AGI). If human intelligence is general, an AI that passes a Turing test against a human will have the same amount of general intelligence as the human. However, there are strong arguments against the idea that human intelligence is completely general [17,26]. If human intelligence is not general, Turing tests between human and AIs cannot prove that the AIs possesses general intelligence.
6 For example, consider a test battery with four questions. Three questions are at level 1 (really easy); the last question is at level 10 (really hard). An agent that can only answer the easy questions is not one point less intelligent than an agent who can also answer the hard question—the latter agent is approximately four times more intelligent than an agent who can only answer the easy questions. A score for this test battery can be placed on an interval scale if the mark for each question is multiplied by the question difficulty.
7 See: https://arcprize.org (accessed on 6 June 2026).
8 Some of the tasks in cognitive test batteries are nicely demonstrated in Mark Rober's challenges for squirrels, ravens and octopi: https://www.youtube.com/@MarkRober (accessed on 6 June 2026).

[9] In theory, the cognitive faculties discussed by Burnell et al. [44] could be measured from an absolute zero (see Section 6.3). However, Burnell et al.'s objective is to use humans as a reference point to measure progress with artificial general intelligence.

[10] Here "psychometric" refers to psychological properties, such as reaction time. Cognitive abilities are covered in Section 6.3.

[11] See: http://prize.hutter1.net (accessed on 6 June 2026).

[12] A key limitation of this type of measure is that most AIs lack anything corresponding to a human goal and can arbitrarily change their goals to increase their rewards (a form of wire-heading). The goal achievement interpretation of intelligence also conflates an agent's planning ability with its physical capabilities: weak agents receive less rewards than equally intelligent strong agents who have a greater ability to manipulate their physical environment [56].

[13] More information about this measure, source code and experiments are available at: www.davidgamez.eu/pi (accessed on 6 June 2026).

[14] For example, the Bamberg Mathematical Consciousness Science Initiative: https://www.uni-bamberg.de/en/bamxi/ (accessed on 6 June 2026).

References

1. Stevens, S.S. On the Theory of Scales of Measurement. *Science* **1946**, *103*, 677–680. [CrossRef] [PubMed]
2. Hendrycks, D.; Burns, C.; Basart, S.; Zou, A.; Mazeika, M.; Song, D.; Steinhardt, J. Measuring Massive Multitask Language Understanding. *arXiv* **2021**, arXiv:2009.03300.
3. Brockman, G.; Cheung, V.; Pettersson, L.; Schneider, J.; Schulman, J.; Tang, J.; Zaremba, W. OpenAI Gym. *arXiv* **2016**, arXiv:1606.01540.
4. Chen, M.; Tworek, J.; Jun, H.; Yuan, Q.; Pinto, H.P.d.O.; Kaplan, J.; Edwards, H.; Burda, Y.; Joseph, N.; Brockman, G.; et al. Evaluating Large Language Models Trained on Code. *arXiv* **2021**, arXiv:2107.03374.
5. Hernández-Orallo, J. Evaluation in artificial intelligence: From task-oriented to ability-oriented measurement. *Artif. Intell. Rev.* **2017**, *48*, 397–447.
6. Jones, C.R.; Bergen, B.K. Large Language Models Pass the Turing Test. *arXiv* **2025**, arXiv:2503.23674.
7. Michell, J. *Measurement in Psychology—A Critical History of a Methodological Concept*; Cambridge University Press: Cambridge, UK, 1999.
8. Feuerstahler, L. Scale type revisited: Some misconceptions, misinterpretations, and recommendations. *Psych* **2023**, *5*, 234–248. [CrossRef]
9. Prytulak, L.S. Critique of SS Stevens' theory of measurement scale classification. *Percept. Mot. Ski.* **1975**, *41*, 3–28.
10. Keyes, D. *Flowers for Algernon*; Gollancz: London, UK, 2002.
11. Glynn, A. *The Dark Fields*; Little Brown: New York, NY, USA, 2001.
12. Legg, S.; Hutter, M. A Collection of Definitions of Intelligence. In *Advances in Artificial General Intelligence Concepts, Architectures and Algorithms: Proceedings of the AGI Workshop 2006*; IOS Press: Amsterdam, The Netherlands, 2007; Volume 157, pp. 17–24.
13. Neisser, U.; Boodoo, G.; Bouchard, T.J., Jr.; Boykin, A.W.; Brody, N.; Ceci, S.J.; Halpern, D.F.; Loehlin, J.C.; Perloff, R.; Sternberg, R.J.; et al. Intelligence: Knowns and Unknowns. *Am. Psychol.* **1996**, *51*, 77–101. [CrossRef]
14. Yousefian, F.; Saberi, H.; Banirostam, T. A conceptual model for ontology based intelligence. *Int. J. Comput. Sci. Netw.* **2016**, *5*, 23–30.
15. Henaff, M.; Weston, J.; Szlam, A.; Bordes, A.; LeCun, Y. Tracking the World State with Recurrent Entity Networks. *arXiv* **2017**, arXiv:1612.03969.
16. Tjøstheim, T.A.; Stephens, A. Intelligence as Accurate Prediction. *Rev. Philos. Psychol.* **2022**, *13*, 475–499.
17. Gamez, D. P: A Universal Measure of Predictive Intelligence. *arXiv* **2025**, arXiv:2505.24426.
18. Poth, N.; Tjøstheim, T.A.; Stephens, A. Rethinking intelligent behaviour through the lens of accurate prediction: Adaptive control in uncertain environments. *Philos. Mind Sci.* 2025, 6. Available online: https://philosophymindscience.org/index.php/phimisci/article/view/11780 (accessed on 6 June 2026).
19. Legg, S.; Hutter, M. Universal intelligence: A definition of machine intelligence. *Minds Mach.* **2007**, *17*, 391–444. [CrossRef]
20. Russell, S.J. *Human Compatible: AI and the Problem of Control*; Penguin Random House: New York, NY, USA, 2019.
21. Gardner, H. *Multiple Intelligences: New Horizons*; Basic Books: New York, NY, USA, 2006.
22. Warwick, K. *QI: The Quest for Intelligence*; Piatkus: London, UK, 2000.
23. Turing, A. Computing Machinery and Intelligence. *Mind* **1950**, *59*, 433–460. [CrossRef]
24. Harnad, S. Levels of Functional Equivalence in Reverse Bioengineering: The Darwinian Turing Test for Artificial Life. *Artif. Life* **1994**, *1*, 293–301.
25. Hingston, P. A Turing Test for Computer Game Bots. *IEEE Trans. Comput. Intell. AI Games* **2009**, *1*, 169–186. [CrossRef]
26. Chollet, F. On the measure of intelligence. *arXiv* **2019**, arXiv:1911.01547.
27. Sanghi, P.; Dowe, D.L. A computer program capable of passing IQ tests. In Proceedings of the 4th International Conference on Cognitive Science (ICCS'03), Sydney, Australia, 13–17 July 2003.

28. Campello de Souza, B.; Andrade Neto, A.S.d.; Roazzi, A. Are the New AIs Smart Enough to Steal Your Job? IQ Scores for ChatGPT, Microsoft Bing, Google Bard and Quora Poe. *SSRN* **2023**, 4412505. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4412505 (accessed on 6 June 2026).
29. Fernandes, H.B.F.; A.Woodley, M.; Nijenhuis, J.t. Differences in cognitive abilities among primates are concentrated on G: Phenotypic and phylogenetic comparisons with two meta-analytical databases. *Intelligence* **2014**, *46*, 311–322. [CrossRef]
30. Galsworthy, M.; Arden, R.; Chabris, C. Animal Models of General Cognitive Ability for Genetic Research into Cognitive Functioning. In *Behavior Genetics of Cognition Across the Lifespan*; Finkel, D., Reynolds, C., Eds.; Springer: Berlin/Heidelberg, Germany, 2014; pp. 257–278.
31. Galsworthy, M.J.; Paya-Cano, J.L.; Liu, L.; Monleon, S.; Gregoryan, G.; Fernandes, C.; Schalkwyk, L.C.; Plomin, R. Assessing reliability, heritability and general cognitive ability in a battery of cognitive tasks for laboratory mice. *Behav. Genet.* **2005**, *35*, 675–692. [CrossRef] [PubMed]
32. Shaw, R.C.; Schmelz, M. Cognitive test batteries in animal cognition research: Evaluating the past, present and future of comparative psychometrics. *Anim. Cogn.* **2017**, *20*, 1003–1018. [CrossRef] [PubMed]
33. Ilić, D.; Gignac, G.E. Evidence of interrelated cognitive-like capabilities in large language models: Indications of artificial general intelligence or achievement? *Intelligence* **2024**, *106*, 101858. [CrossRef]
34. Wicherts, J.M. The importance of measurement invariance in neurocognitive ability testing. *Clin. Neuropsychol.* **2016**, *30*, 1006–1016. [CrossRef] [PubMed]
35. Flynn, J.R. The mean IQ of Americans: Massive gains 1932–1978. *Psychol. Bull.* **1984**, *95*, 29–51.
36. Wicherts, J.M.; Dolan, C.V.; Hessen, D.J.; Oosterveld, P.; Baal, G.C.M.v.; Boomsma, D.I.; Span, M.M. Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence* **2004**, *32*, 509–537. [CrossRef]
37. Locurto, C.; Benoit, A.; Crowley, C.; Miele, A. The structure of individual differences in batteries of rapid acquisition tasks in mice. *J. Comp. Psychol.* **2006**, *120*, 378–388. [CrossRef]
38. Hernández-Orallo, J.; Dowe, D.L. Measuring universal intelligence: Towards an anytime intelligence test. *Artif. Intell.* **2010**, *174*, 1508–1539. [CrossRef]
39. Thurstone, L.L. The Absolute Zero in Intelligence Measurement. *Psychol. Rev.* **1928**, *35*, 175–197. [CrossRef]
40. Chollet, F.; Knoop, M.; Kamradt, G.; Landers, B. ARC Prize 2025: Technical Report. *arXiv* **2026**, arXiv:2601.10904.
41. Chollet, F.; Knoop, M.; Kamradt, G.; Landers, B.; Pinkard, H. ARC-AGI-2: A New Challenge for Frontier AI Reasoning Systems. *arXiv* **2026**, arXiv:2505.11831.
42. Crosby, M.; Beyret, B.; Shanahan, M.; Hernández-Orallo, J.; Cheke, L.; Halina, M. The Animal-AI Testbed and Competition. In Proceedings of the NeurIPS 2019 Competition and Demonstration Track, Vancouver, Canada, 8–14 December 2019; pp. 164–176.
43. Crosby, M.; Beyret, B.; Halina, M. The Animal-AI Olympics. *Nat. Mach. Intell.* **2019**, *1*, 257. [CrossRef]
44. Burnell, R.; Yamamori, Y.; Firat, O.; Olszewska, K.; Hughes-Fitt, S.; Kelly, O.; Galatzer-Levy, R.I.; Morris, M.R.; Dafoe, A.; Snyder, A.M.; et al. Measuring Progress Toward AGI: A Cognitive Framework. *arXiv* **2026**, arXiv:2605.28405.
45. Cattell, J.M. Mental tests and measurements. *Mind* **1890**, *15*, 373–381. [CrossRef]
46. Pearson, K. *The Life, Letters and Labours of Francis Galton—Volume III*; Cambridge University Press: Cambridge, UK, 1930.
47. Pietschnig, J.; Penke, L.; Wicherts, J.M.; Zeiler, M.; Voracek, M. Meta-analysis of associations between human brain volume and intelligence differences: How strong are they and what do they mean? *Neurosci. Biobehav. Rev.* **2015**, *57*, 411–432. [CrossRef] [PubMed]
48. Jensen, A.R. Galton's legacy to research on intelligence. *J. Biosoc. Sci.* **2002**, *34*, 145–172. [CrossRef] [PubMed]
49. Haier, R.J. *The Neuroscience of Intelligence*; Cambridge University Press: Cambridge, UK, 2017.
50. Hutter, M. *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability*; Springer: Berlin/Heidelberg, Germany, 2005.
51. Hernández-Orallo, J. Beyond the Turing test. *J. Log. Lang. Inf.* **2000**, *9*, 447–466. [CrossRef]
52. Reznikova, Z.; Ryabko, B. Numerical competence in animals, with an insight from ants. *Behaviour* **2011**, *148*, 405–434. [CrossRef]
53. Bien, Z.; Bang, W.-C.; Kim, D.-Y.; Han, J.-S. Machine intelligence quotient: Its measurements and applications. *Fuzzy Sets Syst.* **2002**, *127*, 3–16. [CrossRef]
54. Liu, F.; Shi, Y.; Liu, Y. Intelligence Quotient and Intelligence Grade of Artificial Intelligence. *Ann. Data Sci.* **2017**, *4*, 179–191. [CrossRef]
55. Legg, S.; Hutter, M. Tests of Machine Intelligence. In *50 Years of Artificial Intelligence*; Lungarella, M., Iida, F., Bongard, J., Pfeifer, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 232–242.
56. Gamez, D. Measuring Intelligence in Natural and Artificial Systems. *J. Artif. Intell. Conscious.* **2021**, *8*, 285–302. [CrossRef]
57. Hawkins, J. *A Thousand Brains: A New Theory of Intelligence*; Basic Books: New York, NY, USA, 2022.
58. Hawkins, J. *On Intelligence*; Henry Holt and Company: New York, NY, USA, 2004.
59. Bell, T.C.; Cleary, J.G.; Witten, I.H. *Text Compression*; Prentice-Hall: Hoboken, NJ, USA, 1990.
60. Tononi, G. An information integration theory of consciousness. *BMC Neurosci.* **2004**, *5*, 42. [CrossRef] [PubMed]

61. Casali, A.G.; Gosseries, O.; Rosanova, M.; Boly, M.; Sarasso, S.; Casali, K.R.; Casarotto, S.; Bruno, M.A.; Laureys, S.; Tononi, G.; et al. A theoretically based index of consciousness independent of sensory processing and behavior. *Sci. Transl. Med.* **2013**, *5*, 198ra105. [CrossRef] [PubMed]
62. Melloni, L.; Mudrik, L.; Pitts, M.; Bendtz, K.; Ferrante, O.; Gorska, U.; Hirschhorn, R.; Khalaf, A.; Kozma, C.; Lepauvre, A.; et al. An adversarial collaboration protocol for testing contrasting predictions of global neuronal workspace and integrated information theory. *PLoS ONE* **2023**, *18*, e0268577. [CrossRef] [PubMed]
63. Naglieri, J.A.; Bornstein, B.T. Intelligence and achivement: Just how correlated are they? *J. Psychoeduc. Assess.* **2003**, *21*, 244–260.
64. Robertson, K.F.; Smeets, S.; Lubinski, D.; Benbow, C.P. 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. *Curr. Dir. Psychol. Sci.* **2010**, *19*, 346–351.
65. Gamez, D. *Human and Machine Consciousness*; Open Book Publishers: Cambridge, UK, 2018.
66. Tononi, G. Consciousness as integrated information: A provisional manifesto. *Biol. Bull.* **2008**, *215*, 216–242. [CrossRef] [PubMed]
67. Gamez, D. Intelligence and Consciousness in Natural and Artificial Systems. In *Computational Approaches to Conscious Artificial Intelligence*; Chella, A., Ed.; World Scientific: Singapore, 2023; pp. 169–202.