The Collaborative Nature of Intelligence

James L. Crowley Author[0000−0001−7730−8968]

Abstract This chapter reviews scientific understandings of the collaborative nature of human intelligence and discusses implications for the construction of intelligent collaborative systems. We begin with a review of the role of collaboration in the evolution and dominance of social species and discusses how human intelligence is the result of collaboration shaped by reciprocal altruism. We contrast understandings of human intelligence from the perspectives of Educational Science, Psychology and Artificial Intelligence to argue for a unifying view of intelligence as an ability to adapt behaviors under limitations of time, resources and experience in order to survive and dominate in complex changing environments. From this we propose principles to guide the development of collaborative artificial intelligent systems, including requirements for emotional and ethical intelligence

Keywords: Collaborative Intelligence; Reciprocal Altruism; Artificial Intelligence; Human Intelligence; Prediction; Explanation; Adaptation; Ethical AI; Emotional Intelligence

1 Introduction

Humans are an intelligent species empowered by collaboration. Collaboration enables teams of humans to combine abilities, enabling collective behaviors well beyond the capabilities of any one individual. Collaboration has enabled mankind to collect and accumulate the knowledge and experience of billions of individuals, providing abilities to predict natural phenomena and master a large variety of ecosystems and climates. Collaboration has enabled the creation of increasingly powerful technologies as teams of individuals build on past experience to collectively create new artifacts. Collaboration has enabled mankind to create and master technologies

Institut Polytechnique de Grenoble
Univ. Grenoble Alpes, Grenoble France e-mail: James dot Crowley at grenoble-inp dot fr

for fire, agriculture, electricity and communications, empowering and transforming our species in ways that would be unrecognizable to our ancestors.

The collective efforts of humanity have recently resulted in the creation of a new technology that can imitate human social interaction, while surpassing human abilities to predict, explain, and control phenomena. Collaboration with this new technology has the potential to amplify human abilities on scales that can rival or exceed the mastery of fire or the invention of electricity. Fully harnessing this technology for the benefit of humanity will require understanding why humans are hardwired for collaboration, how collaboration enables human intelligence, and how collaboration with intelligent systems can empower individuals and society with previously unimagined new abilities.

2 What is Collaboration?

Collaboration is a process where two or more partners work together to achieve a shared goal [36]. Collaboration can arise either intentionally, through shared goals and communication, or structurally, as may be found among mutually dependent species that co-evolve in symbiosis without shared awareness. Collaboration empowers groups of individuals with abilities that go beyond those of any one member by integrating complementary abilities for perception, action and comprehension. Humans, in particular, are genetically programmed with a moral sense that favors collaborative behaviors under certain conditions. To understand how humanity can benefit from collaboration with artificial intelligence, it is useful to examine the natural origins of this moral instinct. A core concept is altruism.

2.1 Altruism

Altruism is behavior by an individual that benefits others while incurring a cost to the individual [9]. Many examples of altruism can be found in nature. In each example, individuals seem to act in a selfless manner, with actions that can be costly to self, while providing a much greater benefit to others. Altruism has played a central role in the evolution of species, giving rise to robust symbiotic systems and contributing fundamentally to the emergence of humanity as a dominant social species.

Altruistic behaviors would seem to directly contradict natural selection. Sacrificing oneself for the benefit of others should decrease one's own likelihood of reproduction, eventually leading to an extinction of traits that encourage altruistic behaviors. And yet, under certain conditions, Altruism can assure the reproductive success and dominance of species.

Darwin addressed this problem in Origin of Species [8] by discussing the existence of sterile worker ants. Darwin proposed that in the case of social insects, such as bees, ants and wasps that live in colonies, natural selection applies not only to the

individual, but also to the entire colony. In the case of social insects, this idea was highly prescient, as it is now understood that with most such species, members of a colony are descended from the same queen, and thus share the same or highly similar genetic codes. The apparent altruistic behaviors of individual workers are part of a collective effort to propagate the common genetic heritage of the colony.

The existence of altruism between genetically distinct individuals is harder to explain. A common case is the example of altruistic behaviors between genetically related individuals. Such behaviors are called kin selection and include altruistic behaviors that appear to contradict natural selection.

2.1.1 Kin Selection

In 1964, W. D. Hamilton developed a mathematical model that predicts the evolutionary success for species that exhibit altruistic behaviors towards kin [14]. This model is widely used to explain altruistic acts between genetically related individuals.

Hamilton's theory predicts the likelihood that an individual will perform an altruistic act as a function of three factors: genetic relatedness, benefit and cost. Specifically, the likelihood of an altruistic act, A, by an altruist towards a recipient is proportional to the product of the Benefit, B, to the recipient and the coefficient of relatedness, R, between the altruist and the recipient minus the Cost, C, to the altruist. Mathematically, this can be written as:

$$P(A) \propto R \cdot B - C \tag{1}$$

Where *P(A)* is the frequency of occurrence of the altruistic act, $\propto$ is the proportionality relation, R is a coefficient of relatedness between altruist and recipient, B is the reproductive benefit of the behavior to the recipient and C is the reproductive cost of behavior to the altruist. Cost and benefit refer to the probability of reproduction, with effects accumulating among populations of a large number of individuals over many generations.

Hamilton's theory relies on the insight that relatives share some of their genetic material because they have a common ancestor. If individuals behave altruistically to their relatives, then they have some chance of conferring benefits upon individuals who carry copies of their own genes. The altruist may suffer, but their genes prosper.

Several insights can be derived from Hamilton's rule. For example, the coefficient of relatedness, R, is the critical element for determining behaviors that are influenced by kin selection. Thus costly altruism can be expected to be limited to close kin, as the conditions for Hamilton's rule become progressively more difficult to satisfy as costs rise, and the higher the cost of the altruistic behavior, the more the behavior will be restricted to individuals with a close genetic relationship.

In order to meet the conditions of Hamilton's rule, individuals must be able to limit altruistic behavior toward kin. This requires some mechanism for discriminating kin from non-kin such as proximity of origin or the recognition of physical traits. For species in which the movement of individuals from their place of birth is

relatively slow, local interactions tend to be among relatives, with genetic similarity proportional to distance. In this case spatial location alone can provide sufficient information for kin discrimination.

For other species, however, the problem is more challenging. Hamilton [15] predicted that the ability to identify kin would be most fully developed in species that live in social groups where mechanisms for distinguishing kin from non-kin are not likely to be effective and there are multiple opportunities for passive, context-dependent behaviors. Such situations can give rise to complex collective behaviors. This is the case, for example, for pack behaviors observed in many mammalian species.

A pack is a social group of animals that live and hunt together under the direction of a dominant leader. Pack behaviors have been documented in a large variety of mammals ranging from dolphins and orcas, to felines, canines and primates. Pack behaviors enable the group to collaboratively hunt larger prey that would not normally be accessible to individual members, as well as claim and defend territory, giving access to food and other resources. Pack social structures carry over from cooperative hunting into well defined rules for how the results of the hunt are shared, including rules for sharing benefits with members of the pack who were not engaged in the hunt.

2.1.2 Kin Selection Among Primates

Over the last 25 years, primatologists have collected abundant evidence for kin selection and pack behaviors in primates [28]. This evidence shows that kin selection plays a fundamental role in shaping primate social organizations, dispersal strategies, dominance hierarchies, and patterns of affiliative interactions.

Primates generally fit Hamilton's conditions for kin selection. Most primates live in large and relatively stable social groups (tribes), exhibit severe intragroup aggression and intense feeding competition, and live in groups that include both relatives and nonrelatives. Monogamous coupling of male and females is rare and male reproductive success is generally based on male dominance rank, with frequent incursions by non-resident males during mating season. As a result, association during early life provides a reasonably reliable basis for recognition of kin, as any member of a tribe of a similar age may likely be paternal kin, particularly for tribes where a single male monopolizes mating. There is evidence that some species, such a chimpanzees, may be able to recognize kin based on physical traits. None-the-less, according to Silk, tribal membership and age seems to be the primary selection mechanisms for altruism.

Dispersal patterns play an important role in the evolution of collaboration via kin selection. In most primate species, males disperse from their family groups while females remain with kin of varying degrees of relatedness and engage in a variety of altruistic behaviors including food sharing, grooming, alarm calling, and shared parenting. The result is emergence of long-lived maternal dominance hierarchies in which all members of the same maternal lineage occupy contiguous ranks.

Coalitions, in which one individual intervenes on behalf of another in an ongoing dispute, provide evidence for kinship selection. As juveniles mature, they lose support from their mothers and rely increasingly on coalitions for disputes. Support by a coalition can provide reproductive benefits, including rank acquisition and protection from escalation of attacks. Females, in particular, are more likely to support relatives than non-relatives in ongoing aggressive disputes, particularly against higher-ranking opponents. However, while Hamilton's model can be used to explain many behaviors and social structures in primates, some species exhibit social behaviors among unrelated individuals that are more complex and varied than predicted by kin selection. This is particularly the case for hominids, including gorillas, chimpanzees, orangutans, gibbons, bonobos, and humans. In these species, more complex models are required.

2.2 Reciprocal Altruism

Reciprocal Altruism is a behavioral strategy in which an individual engages in an altruistic act with the expectation that the altruistic act will be reciprocated in the future. The concept was proposed by Robert Trivers to explain the evolution of collaboration as instances of mutually altruistic acts [38]. Trivers illustrated the model with three examples observed in nature: (1) symbiotic cleaning behaviors of fish, (2) warning cries of birds, and (3) reciprocal altruistic behaviors of humans. The central idea is that altruists must be able to detect and terminate interaction with individuals who do not reciprocate. Such individuals are referred to as cheaters. Trivers' model shows how natural selection can operate against cheaters and favor reproduction for individuals who practice altruistic behaviors.

Reciprocal altruism applies to an important class of phenomena that cannot be explained by Hamilton's model for kinship selection. Reciprocal altruism explains the existence of stable, mutually beneficial altruistic relations between unrelated individuals, including members of different species, thereby providing a plausible explanation for the co-adaptation of species, for symbiotic phenomena such as cleaning fish and perhaps even human agriculture.

Behaviors for reciprocal altruism will come to dominate a population provided that altruistic individuals are able to limit altruistic acts to other individuals who reciprocate. This requires an ability to detect and discriminate against non-altruistic individuals. In other words, genes that favor a strong moral sense can come to dominate a population through the reproductive benefits of reciprocal altruism.

There is evidence that many aspects of human social behavior can be understood through the evolutionary logic of reciprocal altruism [37]. Appetites for friendship, gratitude, sympathy, trust, and cooperation—as well as aversions such as suspicion, indignation, and moralistic aggression—can be interpreted as regulatory emotional systems shaped by natural selection and mediated in part by hormonal responses. These reactive emotional mechanisms encourage mutually beneficial behaviors while enabling detection and exclusion of cheaters. Over many generations, this feedback

between cooperation and cheating has contributed to the emergence of a strong moral sense that underlies modern human intuitions about justice, fairness, and social obligation, and plays a significant role in the development of language, culture, and collective intelligence.

Unfortunately, evolution has also favored emergence of traits that empower individuals to gain benefits through cheating, including traits that enable cheaters to avoid detection. As a result, humans possess appetites for both altruistic behaviors and selfish behaviors in varying degrees. The expressions of these traits are sensitive to developmental processes imposed by family and culture through reward and punishment. Some cultures may be more successful than others in training susceptible individuals to suppress appetites for cheating and to enhance appetites for altruism. In the long term, this ability may even affect the survival and dominance of a culture. In any case, the constant struggle between abilities to cheat and abilities to detect cheating seems to have played an important role in the evolutionary development of human societies and the development of human emotional intelligence.

Interestingly, unlike Hamilton's model, Trivers' model is not limited to members of the same species. Many examples of reciprocal altruism can be found between populations of different species. This is commonly illustrated with the cleaning behaviors of certain species of fish, but can be used to explain many other altruistic behaviors including the emergence of symbiotic species and stable ecosystems.

Over many generations, reciprocal altruism has shaped the emotional mechanisms that regulate human social behavior. This evolutionary feedback between cooperation and cheating underlies modern intuitions about fairness and justice and has contributed to the emergence of language, culture, and collective intelligence.

2.2.1 The Human Appetite for Collaborative Behavior

To be effective, Collaborative Artificial Intelligence must comply with the appetites and aversions that underlie the innate human sense of morality. Humans have an innate appetite for collaboration and a strong intolerance of cheating or betrayal by collaborative partners. Benefits from reciprocal altruism over many generations have played an important role in the evolutionary development of human social behaviors. Human social behaviors are grounded in an innate sense of morality, expressed by emotional reactions driven by release of hormones in reaction to social situations. These reactive emotional behaviors have evolved through an escalating evolutionary feedback process in which enhanced abilities to cheat in reciprocal altruistic relations are countered by enhanced abilities to detect cheating. The evolutionary development of an emotional sense of justice is increasingly recognized as the foundation for modern human understanding of morality and ethics, as well as a driver for the development of human abilities for natural language and human intelligence.

Ethical behavior by AI systems is not merely desirable, it is imperative for effective collaboration. The survival benefits of reciprocal altruism have resulted in the evolution of a strong moral sense that drives humans to seek mutually altruistic partnerships, and to react with indignation and anger to detection or even suspicion

of cheating or betrayal. Suspicion of betrayal can easily poison collaboration and trigger an innate desire to punish the cheater. Systems must be designed to avoid triggering this strong moral sense by carefully avoiding actions that may result in suspicions of dishonesty, hypocrisy, or betrayal.

At the same time, collaborative systems can benefit from abilities to inspire feelings of sympathy, friendship, respect, gratitude and trustworthiness that regulate altruistic behaviors. Humans have developed sophisticated relational skills to reassure partners of fidelity and inspire confidence. Awareness of such behaviors can enhance the effectiveness of artificial systems by empowering abilities to reciprocate and reassure partners of fidelity. Ethical challenges posed by collaboration with Artificial Intelligence are explored in the introductory chapter on Computation Human-AI Collaboration [6].

Exploiting human appetites for friendship and trust will require a well-developed ability to perceive and display emotions. Such abilities are a key component of human emotional intelligence and important for establishing and maintaining a collaborative relationship. Emotional intelligence may well be a key enabling technology for collaborative intelligence.

3 What is Intelligence?

A variety of definitions have been proposed for human intelligence as different individuals and cultures have attempted to describe and understand why some individuals thrive while others fail. The term has been extended to qualify behaviors of a variety of species and artificial systems. However, the question of whether animals or machines can be labeled as intelligent remains controversial.

In computer science, intelligence is often used to describe the actions and reactions of agents, where an agent is an autonomous entity that can perceive and act so as to change the state of an environment [27]. In natural science, intelligence is used to describe the actions and reactions of organisms [18]. In both areas, criteria used to assess intelligence depend on the observer and on the context in which the observer is making a judgment. Different observers have different standards for when or how an agent or its actions may be judged to be intelligent owing to different appreciations of what constitutes appropriate behavior. Actions and behaviors that may be considered as intelligent in some contexts may be seen as totally inappropriate in others.

Classic scientific views of intelligence have focused on powers of abstract reasoning based on abilities for formation and association of concepts [29]. Indeed, when combined with the use of language to communicate and learn from the comprehension of others, the ability to formulate abstractions is a powerful tool for predicting and dominating an environment. However, abstract reasoning is only a means to intelligence, not the end. The more important question is how abstractions can be used to empower survival and dominance.

3.1 Human Intelligence

The term intelligence emerged in the 15th century as a term for comprehension, based on the earlier Latin and French terms for understanding, knowledge and discernment. The concept of intelligence received increased attention in the early 20th century when intelligence was identified as quality for comparing the developmental potential of children. This original definition was subsequently supplanted by views of intelligence as an innate quality of an individual rather than an ability that could be developed and enhanced through training and collaboration.

3.1.1 The Intelligence Quotient (IQ)

In 1905, Alfred Binet and Theodore Simon developed a test to identify children who might need extra educational support [2]. They proposed that intelligence tests should be used to identify areas of weakness in a child's cognitive abilities for additional training. Their test consisted of a series of tasks of increasing difficulty, such as identifying objects, repeating sentences, and solving problems.

The Intelligent Quotient (IQ) was proposed by William Stern in 1912 [32] based on the test proposed by Binet and Simon. Binet and Simon had speculated that scores in different areas could be combined to estimate the mental age of a student, corresponding to the age where an average student would reach the same level of development. Stern proposed that the mental age could be divided by the chronological age to obtain a coefficient multiplied by 100 (the IQ) that could be used to compare a student's development to others of the same age.

In 1916, Lewis Terman adapted Binet and Simon's test for use in the United States and developed the Stanford-Binet Intelligence Scale [35]. This test included new tasks and norms specifically designed for American children. The Stanford-Binet test measured four categories of cognitive ability:

1. Verbal reasoning, with tests of language comprehension, vocabulary, and verbal reasoning;
2. Nonverbal reasoning, with tests for spatial perception and visual-spatial skills;
3. Quantitative reasoning, with tests of mathematical skills for arithmetic and problem-solving; and
4. Memory, with tests of short-term and long-term memory, as well as working memory.

Terman adopted Stern's Intelligent Quotient to provide a standardized measure for a student's intellectual potential.

The IQ is now widely used as a measure of innate intelligence. This was not the original intent. Binet and Simon believed that intelligence was not a fixed trait, but rather a set of cognitive abilities that could be developed through education and experience. Their test was designed to indicate areas where students might benefit from additional education. Terman's test was proposed as a tool to identify "gifted" students who merit additional opportunities, based on an implicit assumption that

precocious children will continue to develop intelligence at the same rate as they grow older. Over time the IQ has come to represent the degree to which an individual has been endowed with a capacity to acquire and master new skills. Mature adults now regularly brag of a high IQ in order to prove their general intelligence.

3.1.2 General Intelligence

In the mid 1920's, Charles Spearman proposed a two-factor theory of intelligence that considered intelligence is a general mental ability enabling specific cognitive tasks [30]. Spearman's two-factor theory presents intelligence as a composition of two components: an innate general intelligence, *G*, that underlies all abilities of the individual and specific abilities, *S*, that can be developed through training or experience. To validate this hypothesis, he tested how well individuals performed on various task including the abilities to discriminate pitch, perceive weights and colors, and understand directions and perform mathematics.

When analyzing the accumulated test scores, Spearman noted that those who did well in one area also tended to score higher in other areas and concluded that there must be some central factor that influences cognitive abilities. He termed this as general intelligence, or *G*. Spearman speculated that individuals have different levels of general intelligence enabling different abilities for acquisition of skills. To explain the differences in performance on different tasks, Spearman hypothesized that there must exist some specific factor, *S*, specific to a certain aspect of intelligence, resulting from training or experience.

Spearman's two-factor theory was highly influential among educators in the middle 20th century, leading to a number of efforts to develop tests to detect the *G* factor. The theory is now widely contested as having a number of limitations. For one, the theory is of limited scope, focusing solely on a few cognitive abilities while ignoring other abilities, such as emotional intelligence or social intelligence. The theory greatly oversimplified intelligence, claiming that intelligence could be reduced to a single, unified ability, rather than a collection of complementary abilities. Attempts to measure *G* tend to be sensitive to cultural and environmental factors. Most damning for a scientific theory is the lack of support by empirical evidence. There are no examples of well-designed experiments in which the *G* factor theory can be used to formulate verifiable predictions.

It is worth remembering, as researchers discuss the possibility of Artificial General Intelligence (AGI), that the concept of general intelligence is poorly defined and questionable as a benchmark for measuring the suitability of the actions and behaviors of agents. Claims of Artificial General Intelligence are on a poor foundation.

3.1.3 The Theory of Multiple Intelligences

In 1983 Howard Gardner challenged Spearman's assumption of a single unifying general ability by proposing a theory of multiple intelligences. The Theory of Mul-

tiple Intelligences [13] adopts a view that intelligence is not a single, unitary ability but rather a set of distinct abilities for linguistic, logical, spatial, kinesthetic, musical, interpersonal, and intrapersonal interactions.

Gardner's theory posits that humans have multiple forms of intelligence that are relatively independent. In its original form, the theory identified eight intelligences: linguistic, logical, musical, spatial, kinesthetic, interpersonal, and intrapersonal. In later writings, Gardner added additional forms of intelligence including naturalistic, spiritual, pedagogical and digital intelligences.

Gardner derived his set of intelligences from studies of cognitive processing, brain damage, exceptional individuals, and cognition across cultures. To be included in his list, an intelligence was required to meet several criteria: It must occupy a place in evolutionary history, have a distinct developmental progression, have a potential for isolated demonstration following injury of other regions of the brain, be supported from experimental psychology and be supported by the existence of individuals who excel at that particular intelligence but not at others.

The theory of multiple intelligences is sometimes referred to as a cognitive-contextual theory of intelligence. These are theories of intelligence that concern how cognitive processes operate in various settings. Gardner's theories have been popular among educators who argue that a broader definition of intelligence can more accurately reflect the differing ways in which students learn and develop. However, his theory is criticized by established psychologists for its lack of empirical evidence, and dependence on subjective evaluation.

3.1.4 Adaptive Intelligence

Problems with definitions of intelligence based on abilities for abstract reasoning, planning, behavior, or social interaction have inspired a growing number of researchers to view intelligence as the capacity of an agent to adapt behavior to an environment while operating with insufficient knowledge and resources [41]. Wang placed the problem of limited resources at the heart of his definition of intelligence. This definition emphasizes the importance of adaptation of behavior as a mechanism for attaining situated behavior in a wide range of environments, and is in line with common views that human intelligence is a defining characteristic that has empowered humanity to dominate and populate the large range of natural environments found on Earth.

A similar view has been proposed by the Robert Sternberg [34] based on his earlier theory of triarchic intelligence [33]. Sternberg argues that survival and reproduction of the species is the fundamental requirement for life. Accordingly, he has proposed a framework that defines intelligence as a collection of cognitive processes that can be flexibly applied to different contexts and tasks in order to preserve and perpetuate the species.

In the following, we will define intelligence as an ability to dynamically adapt behaviors under strong constraints on time, resources, and experience so as to operate or dominate in complex and changing environments. Viewing intelligence as

adaptation of behaviors under constraints on time, resources and experience provides a unifying view for both human and artificial intelligence that converges nicely with common language uses for the term intelligence. This definition is in line with popular interpretations of the IQ as it includes acquisition of new skills and abilities for operating and solving problems in unusual situations with limited time and resources. It also includes learning from practical experience and developing new methods for resolving problems. This definition also avoids problems created by trivial solutions to the Turing test such as the use of a very large memory of predefined responses.

3.2 Artificial Intelligence

A variety of definitions for intelligence have been used by the scientific community of Artificial Intelligence, as enabling technologies have evolved over the years. A common starting point for many researchers has been the behavioralist definition used by Alan Turing of intelligence as human-level performance at interaction.

3.2.1 The Turing Test

Alan Turing was a British mathematician and philosopher, widely regarded as the father of Theoretical Computer Science. Turing established his reputation in 1936 by proposing a formal definition for computable functions based on finite state machines. In the late 1940s, Turing addressed the question of whether human intelligence is a computable function. As part of this effort, Turing proposed a test that he referred to as the imitation game, and is now known as the Turing Test [39]. The Turing Test is performed by a human evaluator who engages in a text-based conversation over a teletype without knowing whether they are interacting with a human or a machine. If the evaluator is unable to reliably distinguish the machine from a human, then the machine can be considered to have passed the test and to possess a form of intelligence.

The Turing Test provides an important reference when discussing Artificial Intelligence, as it is easily understood by non-technical audiences. The test evaluates the interaction of an agent (the machine) with its environment (the evaluator) and can be easily generalized to other domains, by focusing on abilities for human-level performance at tasks in other environments such as chess, mathematics, or problem solving.

However, the Turing test has many obvious limits. It does not make statements about how behavior is generated. As defined by Turing, the test is limited to text-based interaction and requires abilities for human-level performance at conversational and social interaction that are highly dependent on culture and education. In addition to the obvious social and linguistic biases, the Turing test has an important technical limitation. If the machine is allowed to operate with unbounded memory, then simplistic algorithms using table look-up can, in principal, pass the Turing Test.

3.2.2 Symbolic AI

In 1955, John McCarthy, then a young Assistant Professor of Mathematics at Dartmouth College, decided to organize a scientific meeting to clarify and develop ideas about thinking machines. McCarthy approached the Rockefeller Foundation to request funding for a summer seminar at Dartmouth for about 10 participants. In September 1955, McCarthy submitted a formal proposal for an 8-week symposium to discuss computers, natural language processing, neural networks, abstraction and creativity. The proposal was co-signed by Marvin Minsky, Nathaniel Rochester and Claude Shannon and is now widely credited with introducing the term "Artificial Intelligence". Invited participants included Allen Newell, Herb Simon, Arthur Samuel, Warren McCulloch and a variety of other individuals selected for their contributions in areas related to cybernetics, automata theory, and complex information processing. The Dartmouth conference became the founding event for a new scientific community devoted to computational approaches for intelligence, distinguishing the domain from cybernetics and control theory.

In proposing the symposium, McCarthy defined intelligence as the ability to achieve goals. He argued that intelligent agents, whether human or artificial, should be characterized as entities that can sense and reason about their environment, and use reasoning to take actions to achieve goals. McCarthy's definition of intelligence as a computational process that involves goal-directed behavior had a significant influence on the development of AI research and guided the development of methods for analysis and decision-making systems in a variety of domains. Following the Dartmouth Conference, Artificial Intelligence (AI) became recognized as a field of Computer Science that focused on programming computers to perform tasks that typically require human intelligence. Knowledge representation, inference, planning and problem solving were recognized as core problems and rule-based systems, symbolic reasoning, and search emerged as dominant paradigms.

Many of the paradigms that dominated early Artificial Intelligence were developed by Allen Newell and Herb Simon. In the 1950's while working at Rand Corporation, Newell proposed the General Problem Solver, or GPS, as a general algorithm that could be used to solve a wide range of problems including mathematical problems, logic puzzles, and geometric problems, by breaking them down, recursively, into ever smaller sub-problems. The GPS was one of the first AI algorithms to use heuristics, or "rules of thumb". Newell further developed these ideas in doctoral work at Carnegie-Mellon University under the direction of Herb Simon [22]. Working with the programmer J. C. Shaw, Newell and Simon developed the Logic Theorist, establishing theorem proving as a foundational technology for early research in AI [21].

Newell and Simon subsequently proposed a formal framework for problem solving based on the notion of a problem space [23]. A problem space is a graph of states connected by actions, where a state is defined as a conjunction of predicates (truth functions). Actions transform states by negating some predicates and asserting others. A problem is defined in a problem space by specifying an initial state and criteria (predicates) for possible goal states. Planning algorithms were developed to search

for the least-cost path from the initial state to a goal state. Possible approaches include expanding the search forward from an initial state (forward chaining) or backward from a goal state (backward chaining), expanding a complete tree of possible actions with depth-first or breadth-first search, guiding the search with a cost function (heuristic search), and hierarchical search [24]. While trivial solutions can be found for finite state-spaces, unbounded spaces give rise to a variety of interesting challenges. Placing limits on available memory, computing resources, or time, and including the possibility of errors in perception or action, also raise a variety of challenging problems relevant to real-world applications that continue to occupy researchers in robotics and AI.

The problem space formalism presented intelligence as a form of rational behavior, in which the intelligence of an agent is judged by its ability to choose actions to accomplish goals. Inspired by economic theory, Newell formulated this as the "principle of rationality", claiming that to be considered as intelligent, an agent must choose actions to accomplish goals. Simon generalized this with the notion of "Bounded Rationality", recognizing that the cost of collecting information for optimal search [29] may exceed the benefit. Simon referred to bounded rational behavior as "satisficing", and was awarded a Nobel Prize in economics for showing that real economic agents could not behave rationally because of the cost of acquiring the information needed for optimal search. This state-space view of intelligence continues to dominate some areas of AI to this day, particularly in robotics and planning, and problems of how to achieve bounded rational behavior continue to motivate research.

In the 1970s, the use of predicates and logic for defining and searching problem spaces led a view that intelligence requires manipulation of symbols. In his 1980 Turing Award Lecture, Allen Newell formalized this as the "physical symbol system" hypothesis, stating that an intelligent system is essentially a system for manipulating symbols according to formal rules. With this view, even human intelligence is reduced to a form of symbol manipulation. The view of intelligence as symbol manipulation became a central dogma for AI in the 1980s, leading to dominance of approaches such as logic programming, and theorem proving.

3.2.3 Embodied Intelligence

In the 1980's Rodney Brooks proposed an alternative to symbolic computing for robotics that he referred to as Behavioral Robotics [4]. Brooks argued that intelligence arises from the interaction of a system with its environment. He defined intelligence as the ability of an autonomous system to achieve goals in a complex and changing environment, without relying on symbolic representations or explicit instructions. Brooks proposed an approach, referred to as the subsumption architecture, in which hierarchies of behaviors were defined as reactive perception-action cycles. With this approach, simple reactive behaviors can be used to construct robots that walk and navigate in complex environments with insect-like motions.

While the subsumption architecture was soon found to be of limited utility, behavioral robotics inspired a variety of successful reactive techniques for mobility

and manipulation. The approach has been found amenable to building systems that can acquire new abilities using reinforcement learning to learn from interaction with an environment.

Brook's ideas led to an approach known as "embodied intelligence" [31]. Embodied intelligence builds on the view that intelligence describes the interaction of an agent with its environment. The concept of embodiment emphasizes the importance of physical interaction and sensorimotor feedback as a fundamental property of intelligence. Intelligence arises from the dynamic interplay between the agent's body, its sensorimotor system, and the environment in which it operates.

Embodied intelligence states that to be considered as intelligent, an agent must be embodied, situated and autonomous [3]. Embodied means that the agent must possess a physical body that is able to act on the environment. Situated means that the actions and behaviors of the agent are shaped by and adapted to the physical constraints and affordances of the environment in which it operates. Autonomous means that the agent has the ability to act independently in order to make decisions and take actions based on its own goals and internal state, without external control or direction.

Programming an embodied intelligence to operate with a limited set of simple behaviors in a controlled environment is an interesting intellectual challenge. However, programming anything resembling natural intelligence for reliable operation in a high highly complex real-world environment has proven intractable. Early AI researchers, using symbolic programming techniques, were unable to develop a theory for AI technologies that operate reliably in natural environments. The reliability of such systems rapidly degrades as the complexity of the environment increases. Some form of adaptation is required. The inability of programmers using symbolic AI techniques to provide reliable behaviors for embodied intelligence in natural environments underscores the need for technologies for machine learning as a prerequisite for artificial intelligence. Such learning requires an ability to learn through interaction with an environment.

One of the more important scientific results of research on behavioral and embodied intelligence was the development of reinforcement learning. Reinforcement learning provides an alternative to supervised and unsupervised learning by enabling an agent to learn through interaction with an environment. The environment is typically seen as a form of a Markov Decision Process (MDP), and many reinforcement learning algorithms use deep learning techniques to learn reliable policies that maximize reward functions.

Classically, in reinforcement learning, an agent interacts with an environment in discrete time steps. At each time step the agent performs an action based on its current state and input from the environment and a behavioral policy that maps agent states and environmental states into actions. As a result of the action, the agent receives a reward that may be positive or negative. The agent then seeks to adapt the policy so as to maximize maximizes the expected cumulative reward. The classic problem, referred to as the credit assignment problem, is to know action has resulted in the reward. Attempts to learn behavioral policies through with symbolic systems

or classic dynamic programming methods proved intractable. However, Artificial Neural Networks are well suited for such learning.

3.2.4 Artificial Neural Networks

Artificial neural networks provide a distributed computing technology that can be trained to imitate any computable function. This technology has enabled substantial advances in areas such as computer vision, robotics, speech recognition and natural language processing, and appear to provide a technology for systems that can learn to operate reliably in extremely complex natural environments. In many areas, such systems can outperform humans in tasks requiring perception, action and cognition.

Artificial neural networks are constructed from large networks of perceptrons. The perceptron is an incremental learning algorithm that uses errors to learn a hyper-dimensional linear decision boundary. The perceptron was first proposed as a trainable classifier by Frank Rosenblatt [25] in 1958 based on a formal model for a Neuron proposed by McCullough and Pitts in 1943 [20]. After a brief period of popularity in the late 1950's, interest in perceptrons declined because of difficulties in training and lack of an enabling technology.

In the late 1970s, frustrations with attempts to use symbolic computing to provide solutions for simple perceptual tasks motivated researchers to re-examine perceptrons. An important advance occurred with the idea of replacing the non-linear threshold of the perceptron with a soft differentiable decision function. This allowed the perceptron to be trained by the classic gradient descent optimization algorithm with a distributed algorithm known as back-propagation.

The back-propagation algorithm implements training as a backwards flow of correction energy mirroring the forward flow of activation energy computed by a multi-layer network of perceptrons. A perceptron can be expressed as a vector of parameters that can be tuned from errors using training data for which the correct output is known. Driving the perceptron with a training sample results in an error, based on the difference between the observed and expected output. This error can be propagated back through the network enabling calculation of a correction term for each parameter, based on the derivative (or gradient) of the error term with respect to the parameter.

The use of gradient descent by back-propagation was explored in a series of experiments conducted in a collaboration by Goeff Hinton at Carnegie Mellon University and Terry Sejnowski of UC San Diego in the late 1970s and early 1980s. A 1984 CMU technical report [17], published the following year in the journal Cognitive Science [1], reports their experiments with using a distributed form of gradient descent to train auto-encoders for reconstructing noisy data. The 1986 paper by Rumelhart, Hinton and Williams [26] is now widely used as the first publication of the distributed back-propagation algorithm for training multi-layer neural networks. This led to a second wave of popularity for neural networks in the late 1980s, as networks of perceptrons were shown to provide simple solutions to a number of classic hard AI problems such as recognizing spoken words and hand-written characters.

Back-propagation made it possible to train and operate arbitrarily large layered networks using SIMD (Single-Instruction-Multiple-Data) parallel computing. However, difficulties with reproducing experimental results, lack of a fundamental theory for how to design networks, and the high cost of network training, led machine learning researchers to prefer techniques with well understood mathematical foundations, such as Bayesian techniques, boosted learning, and support vector machines. The situation began to change in the early 2000's, driven by the widespread availability of SIMD Graphical Processing Units (GPUs) capable of computing neural networks on personal computers, as well as the availability of large data-sets of images and sound distributed with the World Wide Web.

3.2.5 Deep Learning

During the period 2005 to 2015, revolutionary progress in Machine Learning, Natural Language Processing, Computer Vision, and Spoken Language Understanding occurred, driven by three phenomena:

1. The continued growth of available computing power made possible with GPUs and ASICs (Application Specific Integrated Circuits) made widely available via cloud computing,
2. The availability of extremely large-scale data-sets of images, video, speech and natural language for training and testing available over the internet, and
3. A movement toward challenge-based research on machine learning.

Challenge based research, in which a community publishes a data-set and invites researchers to compete in writing code for tasks on this data, has been particularly important in stimulating the empirical scientific research driving much of this progress. In 2009, ImageNet (image examples for each of the words in WordNet) was combined with the European PASCAL Visual Object Class (POC) challenge team to create a joint research challenge known as the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [11]. In 2010 and 2011, a good score for the ILSVRC was a 30% to 40% error rate and the winning teams used standard computer vision techniques. In 2012, Alex Krizhevsky won the ILSVRC with a deep convolutional neural net called AlexNet [19]. AlexNet achieved an error rate of 16%, well below the second-best score of around 26%. This dramatic quantitative improvement marked the start of the rapid shift to techniques based on Deep Learning using Neural Networks by the computer vision community. By 2014, more than fifty institutions participated in the ILSVRC, almost exclusively with Neural Network Architectures. In 2017, 29 of 38 competing teams demonstrated error rates less than 5%.

Much of the progress during the period 2014 to 2020 was obtained by training networks at ever increasing depths, leveraging the growing availability of computing power provided by application specific integrated circuits and related technologies, and made possible by the availability of very large datasets of annotated data. When used with Reinforcement Learning, Deep Learning provided systems that have employed surprising strategies to greatly exceed human abilities at games such as Go

and Chess, and to solve practical problems such as protein folding. Deep learning was increasingly recognized by the popular media as a revolutionary new technology for AI, as convolutional networks were shown to provide effective solutions to long standing problems in computer vision and spoken language recognition, and recurrent networks provided solutions for problems in natural language processing.

3.2.6 Generative Systems and Large Language Models

Just as the general public was becoming used to the power of Deep Learning, a revolutionary new form of AI emerged in the form of Generative Systems [40]. Generative Artificial Intelligence (Generative AI) combines a discriminative component (an encoder) with a generative component (a decoder), communicating through a small number of hidden latent variables (a code) [12]. The latent variables provide a form of concept learning that distills compact representations from training data. Such representations resemble the abstractions used by humans to predict and explain phenomena.

As with human children, Generative Systems form abstractions by learning to reconstruct, recall and associate abstract representations from experience. By training on massive amounts of raw unstructured data, generative systems can learn to interpret and react to novel input, to create meaningful original output, and to exhibit human-level performance for a variety of tasks. Such systems can learn to interpret written or spoken natural language, images, sounds or any other sensory input by generating written or spoken natural language, realistic images, music, sounds, sensory-motor control, or other output.

The most powerful examples of such Generative Systems have been created with the Generative Pretrained Transformers (GPT) architecture created by OpenAI. In November 2022, OpenAI stunned the general public with the release of ChatGPT based on GPT-3. Suddenly a technology was available that allowed computing systems to carry on a creative conversation in natural language [5]. ChatGPT could be asked to search for information from the vast resources of the internet, and provide original, unscripted explanations of complex processes and concepts. This was rapidly followed by a system that could generate original realistic images from natural language prompts (DALL-E), followed by tools to generate video and music from text. Emergence of unexpected abilities, such as the capacity to generate computer code, and an ability to confidently fabulate plausible lies, made it obvious that Generative AI could be used to create revolutionary new tools for human empowerment. Technical details for Generative Artificial Intelligence, GPT and ChatGPT may be found in [7].

Generative systems are well adapted to cross-modal and multimodal interactions, interpreting input stimuli in one or more input modalities by generating reactions in one or more output modalities [42]. As a result, they can drive tangible embodiments and carry on an intelligent conversation. The most immediate impacts will be in the area of access to information that enhances human abilities to predict and explain. More profound impacts will likely emerge as Generative AI enables autonomous

systems with creative abilities and a capacity to adapt and self-repair. But are such systems really intelligent?

4 Intelligence as Adaptation of Behavior

Humanity has managed to thrive and dominate in an extremely large variety of environments and ecosystems. This evolutionary success is the result of three unique abilities:

1. The ability to predict,
2. The ability to explain, and
3. The ability to adapt.

The ability to predict enables individuals to anticipate danger and select courses of actions that favor survival and dominance. The ability to explain enables individuals to collaborate, using explanations to recruit and coordinate collaborators. The ability to adapt enables individuals to improve the efficiency and effectiveness of existing behaviors, and to develop new behaviors that favor survival in new or changing environments. Behaviors based on any of these are commonly recognized as examples of intelligence.

4.1 The Ability to Predict

A prediction is a statement about the unknown. In common usage, the term describes a forecast of an event or state that has not yet occurred, as well as statements about phenomena that may have already occurred but have not yet be perceived. Both usages are consistent with common notions of intelligence.

Predictions concern phenomena: anything that can be observed or imagined. Phenomena include real or hypothetical objects, events, states, and actions. Classically, Kant used the term phenomena to described perceptions of the real world. In the following, we extend the usage to include internal perceptions generated by imagination. If you can think of something, it is a phenomenon.

Predictions empower individuals to anticipate phenomena before they are perceived, enabling defense against threats and exploitation of opportunities. Most species of animals learn to make predictions through passive conditional association, or operant conditioning. Human abilities to predict go beyond simple association to include complex phenomena and future events.

The classic example of passive conditional association is the widely repeated story of Pavlov's experiments with dogs. Dogs naturally salivate at the sight of meat. Pavlov observed that in the laboratory, dogs would salivate whenever they saw the white coat of the lab assistant who delivered the meat. To study this phenomenon, Pavlov introduced the sound of a tone whenever the animals were fed. Eventually,

an association was formed, and the animals were observed to salivate whenever they heard the sound, even if no food was present. The dogs learned to predict meat whenever they heard the tone.

Operant conditioning is a type of associative learning that focuses on consequences that follow an action or behavior. Operant behaviors are typically acquired (or conditioned) through a learning process where voluntary behaviors are modified by the addition or removal of reward or punishment. The subject learns to predict that certain actions or behaviors will result in a phenomena.

Human abilities to predict go beyond conditional associations. Humans are able to formulate complex models that include the notion of causality. Causality assumes that there is a normal temporal sequence of events that may be altered by deliberate intervention by an agent. The agent may be human, natural, super-natural or hypothetical. Causality provides a tool for reasoning about possible sequences of events that may be generated by an individual, or by the actions of others. Causal models enable humans to design processes that can bring about desired consequences. They also enable formulation of rules that regulate collaborative behaviors that are fundamental to human society.

4.2 The Ability to Explain

An explanation is a statement that may be used make predictions. Humans use explanations to share predictions, to share models that provide the ability to make predictions, and formulate instructions for collaborative behaviors.

As humanity has developed abilities to communicate with spoken language, the human brain has evolved a corresponding ability to communicate and reason with stories [16]. Stories describe the natural flow of events and give examples of how a natural flow of events can be altered by the deliberate intervention of agents. Stories enable groups of humans to develop and share subjective realities that regulate collective behaviors and facilitate collaboration and cooperative competitions that make society possible.

Stories, in the form of religious texts, encode human moral impulses for reciprocal altruism with examples that show the consequences good and evil. Such texts communicate the shared subjective realities that empower individuals to collaborate without the need for negotiation. Rules of the road, moral principles, religious doctrine, and civil laws encode the lessons from stories with prescriptions of appropriate and inappropriate behaviors, enforced by examples of the consequences of failure to comply.

Explanation can be defined as a statement that makes something clear or justifies an action. Explanations allow humans to share not only predictions but also the ability to predict. Explanations enable humans to teach new skills, share awareness, coordinate behaviors, share experience, and collectively develop solutions to problems. For reactive skills, explanation focus attention on phenomena that allow the skill to be controlled and assimilated. For situational awareness, explanations draw

attention to the underlying entities and relations as well as associated actions and intentions. For operational collaboration, an explanation can provide a description of the sequence of intended actions that can take a situation to a desired state, as well as a description of the sequence of intermediate situations that can be used to ensure the proper execution of actions and operations. Explanations can be used to share knowledge about how to obtain information and coordinate actions based on habits and social norms. Explanations can also facilitate agreement on protocols for interaction and collaboration, facilitating coordinated action and recognition of intention. Explanations can be used to describe hypotheses or approaches for creative collaboration. An ability to generate and interpret explanations is key for all levels of collaboration with intelligent systems.

Stories are rendered as narratives. Explanations can be structured as narratives that provide answers to the classic 5WH questions of Who, What, Why, When, Where, and How. Specifying these elements is key to establishing a shared situation model and a shared agreement on operational plans and authority, whether describing past, present or future situations. ”What” and ”where” describe the entities and relations that compose the explanation. ”Why” describes the desired or intended goal. ”How” concerns the sequence of actions or operations that can be used to reach the goal. ”When” describes the conditions under which the actions and operations can be performed. ”Who” assigns the operational authority to perform the actions, or establishes a protocol for determining authority during the operation.

A narrative structure based on the 5WH questions can substitute for a lack of experience by providing grounding for interpreting instructions. Narratives can also be used to diagnose and learn from the results of operations ”after the fact” when operations fail to provide a desired outcome. A narrative explanation can help identify whether the failure was due to incomplete or inaccurate model of the situation or the result of erroneous assumptions or some other cause. A narrative that explains the understanding of the situation and the reasons for selecting actions can be used to learn from the failure and improve operations for the future.

4.3 The Ability to Adapt

Adaptation can be defined as the acquisition of new abilities through explanation or experience. An ability to adapt is even more fundamental to survival and dominance than comprehension or explanation. The enormous complexity of the real world imposes a requirement to continue to acquire and adapt abilities. Indeed, most definitions of intelligence include an ability to learn as a core ability.

Adaptation is grounded in reactive abilities. Reasoning and comprehension require energy and time, reducing the effectiveness of an action or behavior. They also consume attention, limiting cognitive abilities to detect and monitor other phenomena that can threaten or favor survival. Assimilation transforms a behavior from slow reasoned processes that is expensive in time and energy into an immediate reaction that can be efficiently engaged with minimal effort. Reducing the cognitive

cost, in particular, can allow the individual to choose from a much larger palette of behaviors, empowering effective operation in a larger range of circumstances and environments.

Four functions characterize human abilities to learn from interaction with an environment [10]. These four functions are:

1. Attention: an ability to select and amplify some information, while diminishing or eliminating others.
2. Engagement: a process driven by curiosity that encourages the brain to formulate and test new hypotheses.
3. Feedback: a process that compares predictions with observations and corrects models of the environment.
4. Consolidation: a process that that renders learned abilities automatic or reactive.

Attention, Engagement, Feedback and Consolidation are the secret ingredients of human learning, not only enabling collaboration, but informing a deeper understanding of why and how to collaborate as individuals, groups and societies. Explanation can accelerate each of these functions. Explanations can enhance assimilation of new behaviors by focusing attention on external and internal phenomena associated with successful performance of a behavior. Explanation enhances engagement by challenging the individual to improve performance. Explanations provide awareness of errors allowing the individual to improve models of the environment and enhance attention to critical phenomena. Explanations can reward correctly assimilate behaviors, further encouraging practice and consolidation.

Collaboration is key for intelligence and dominance. The evolution of life from self-reproducing molecules, to complex ecosystems composed of multiple interacting species is the result of the enabling power of collaboration to enable adaptation, survival and reproduction in increasingly complex unpredictable environments.

Generative AI provides an enabling technology for collaborative systems that can vastly empower human abilities to predict, explain and adapt. The question is how to employ this new technology to assure benefit and enhance the quality of life for humanity.

5 Conclusions

Collaborative Intelligence is the collective ability of a group of agents to adapt behaviors in complex and changing environments by generating innovative solutions to novel challenges under constraints of resource, time and experience. Collaborative intelligence is made possible by integration of complementary abilities for prediction, explanation and adaptation that enables emergence of collective responses to challenges. This requires a form of altruistic interaction in which each member of the group acts within their domain of abilities to bring about the shared objectives. It also requires an ability to predict and explain courses of action so as to reach shared

agreement on the goals of the collaboration, and on the domain of authority within which member can act to attain those goals.

Generative AI is increasingly recognized as a rupture technology with potential for profound economic and sociological impact. The challenge is how to harness this potential to enhance the quality of life for humanity. Scientific foundations from evolutionary biology, anthropology, psychology, and computer science suggest a number of principles to guide the development of a technology for collaborative intelligent systems. These principles provide recommendations for the development of enabling technology, as well as ethical and moral guidelines for deployment, to assure that collaborative intelligence improves quality of life by empowering individuals and society.

Spoken Language has empowered human intelligence by enabling individuals to work together, and to learn from the experience of others. An ability to effectively communicate ideas and emotions with language is fundamental to sharing comprehension through explanation. Explanation and discussion can also be extremely effective for coordinating actions and inspiring confidence.

Emotional intelligence will likely prove to be a key enabling technology for collaborative artificial intelligence. Artificial systems can benefit from an awareness of human emotional intelligence, and use this awareness to inspire affection and trust. The ability to sense and understand the moods and emotions of human partners can enable intelligent systems to encourage collaboration by inspiring human appetites for altruistic behaviors, leading to more effective collaborative groups.

As discussed in section 2, Ethical behavior by AI systems is not merely desirable, it is imperative for effective collaboration. Humanity has evolved a strong moral sense that drives humans to seek mutually altruistic partnerships, and to react with indignation and anger to detection or even suspicion of cheating or betrayal. Humans have a persistent appetite for collaboration and a deep intolerance of betrayal by collaborative partners. An ability for machines to sense and accommodate human appetites for friendship and trust are likely to prove fundamental to collaborative artificial intelligence.

References

1. D. H. Ackley, G. E. Hinton, and T. J. Sejnowski. A learning algorithm for boltzmann machines. *Cognitive Science*, 9(1):147–168, 1985.
2. A. Binet and T. Simon. *The Development of Intelligence in Children*. Williams and Wilkins, Baltimore, MD, 1916.
3. C. Breazeal. *Designing Sociable Robots*. MIT Press, 2002.
4. R. A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 2(1):14–23, 1986.
5. T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, and D. Amodei. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901, 2020.
6. M. Chetouani. *Handbook of Human-AI Collaboration*, chapter Introduction to Computational Human-AI Collaboration. Springer, 2026.

7. J. L. Crowley. *Handbook of Human-AI Collaboration*, chapter Generative Artificial Intelligence. Springer, 2026.
8. C. Darwin. *The Origin of Species*. 1859.
9. R. Dawkins. *The Selfish Gene*. Oxford University Press, Oxford, 1976.
10. S. Dehaene. *How We Learn*. Penguin Books, 2021.
11. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.
12. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
13. H. Gardner. *Frames of Mind*. Basic Books, New York, 1983.
14. W. D. Hamilton. The genetical evolution of social behaviour i and ii. *Journal of Theoretical Biology*, 7:1–52, 1964.
15. W. D. Hamilton. Kinship, recognition, disease, and intelligence: constraints of social evolution. *Animal Societies: Theories and Facts*, pages 81–100, 1967.
16. Y. N. Harari. *Sapiens: A Brief History of HumanKind*. Harper Collins, New York, 2015.
17. G. E. Hinton, T. J. Sejnowski, and D. H. Ackley. Boltzmann machines: Constraint satisfaction networks that learn. Technical Report CMU-CS-84-119, Carnegie Mellon University, Pittsburgh, PA, May 1984.
18. F. Keijzer. Agents and organisms: Why the difference is important for the representation discussion (and cognitive science in general). In *Proceedings of the AISB 2015 Symposium on AI and Games*, 2015.
19. A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, volume 25, 2012.
20. W. S. McCullough and W. H. Pitts. A logical calculus of ideas immanent in neural activity. *Bulletin of Mathematical Biophysics*, 5:115–133, 1943.
21. A. Newell, J. C. Shaw, and H. A. Simon. Elements of a theory of human problem solving. *Psychological Review*, 65(3):151–166, 1958.
22. A. Newell and H. A. Simon. Gps, a program that simulates human thought. In E. A. Feigenbaum and J. Feldman, editors, *Computers and Thought*, pages 279–293. McGraw-Hill, New York, 1963.
23. A. Newell and H. A. Simon. Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, 19(3):113–126, 1976.
24. N. J. Nilsson. *Principles of Artificial Intelligence*. Morgan Kaufmann, 1980.
25. F. Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.
26. D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by backpropagating errors. *Nature*, 323:533–536, Oct. 1986.
27. S. J. Russell and P. Norvig. *Artificial intelligence: a modern approach*. Pearson, 1995.
28. J. B. Silk. The evolution of cooperation in primate groups. In H. Gintis, S. Bowles, R. Boyd, and E. Fehr, editors, *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*. MIT Press, Cambridge, MA, Feb. 2005.
29. H. A. Simon. *The Sciences of the Artificial*. MIT press, 1957.
30. C. E. Spearman. *The Abilities of Man: Their Nature and Measurement*. Macmillan, New York, 1927.
31. L. Steels and R. Brooks, editors. *The "Artificial Life" Route to "Artificial Intelligence": Building Situated Embodied Agents*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1995.
32. W. Stern. *Psychologische Methoden der Intelligenzpruefung*. Barth, Leipzig, 1912.
33. R. J. Sternberg. *Beyond IQ: A Triarchic Theory of Human Intelligence*. Cambridge University Press, 1985.
34. R. J. Sternberg. Adaptive intelligence: Its nature and implications for education. *Education Sciences*, 11:823, 2021.
35. L. M. Terman. *The Measurement of Intelligence*. Houghton Mifflin, Boston, 1916.
36. L. G. Terveen. An overview of human-computer collaboration. *Knowledge-Based Systems*, 1995.

37. R. Trivers. *Reciprocal altruism: 30 years later*, pages 67–83. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.
38. R. L. Trivers. The evolution of reciprocal altruism. *The Quarterly Review of Biology*, 46(1):35–57, Mar. 1971.
39. A. M. Turing. Computing machinery and intelligence. *Mind*, 59(236):433–460, 1950.
40. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017.
41. P. Wang. On defining artificial intelligence. *Journal of Artificial General Intelligence*, 10(2):1–37, 2019.
42. L. Ye, M. Rochan, Z. Liu, and Y. Wang. Cross-modal self-attention network for referring image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10502–10511, 2019.