Toward a Formal Characterization of Real-World General Intelligence

Ben Goertzel
Novamente LLC
1405 Bernerd Place
Rockville MD 20851

Abstract

Two new formal definitions of intelligence are presented, the "pragmatic general intelligence" and "efficient pragmatic general intelligence." Largely inspired by Legg and Hutter's formal definition of "universal intelligence," the goal of these definitions is to capture a notion of general intelligence that more closely models that possessed by humans and practical AI systems, which combine an element of universality with a certain degree of specialization to particular environments and goals. Pragmatic general intelligence measures the capability of an agent to achieve goals in environments, relative to prior distributions over goal and environment space. Efficient pragmatic general intelligences measures this same capability, but normalized by the amount of computational resources utilized in the course of the goal-achievement. A methodology is described for estimating these theoretical quantities based on observations of a real biological or artificial system operating in a real environment. Finally, a measure of the "degree of generality" of an intelligent system is presented, allowing a rigorous distinction between "general AI" and "narrow AI."

Introduction

"Intelligence" is a commonsense, "folk psychology" concept, with all the imprecision and contextuality that this entails. One cannot expect any compact, elegant formalism to capture all of its meanings. Even in the psychology and AI research communities, divergent definitions abound; Legg and Hutter (LH07a) lists and organizes 70+ definitions from the literature.

Practical study of natural intelligence in humans and other organisms, and practical design, creation and instruction of artificial intelligences, can proceed perfectly well without an agreed-upon formalization of the "intelligence" concept. Some researchers may conceive their own formalisms to guide their own work, others may feel no need for any such thing.

But nevertheless, it is of interest to seek formalizations of the concept of intelligence, which capture useful fragments of the commonsense notion of intelligence, and provide guidance for practical research in cognitive science and AI. A number of such formalizations have been given in recent decades, with varying degrees of mathematical rigor. Perhaps the most carefully-wrought formalization of intelligence so far is the theory of "universal intelligence" presented by Shane Legg and Marcus Hutter in (LH07b), which draws on ideas from algorithmic information theory.

Universal intelligence captures a certain aspect of the "intelligence" concept very well, and has the advantage of connecting closely with ideas in learning theory, decision theory and computation theory. However, the kind of general intelligence it captures best, is a kind which is in a sense *more general* in scope than human-style general intelligence. Universal intelligence does capture the sense in which humans are more intelligent than worms, which are more intelligent than rocks; and the sense in which theoretical AGI systems like Hutter's AIXI or $AIXI^{tl}$ (Hut05) would be much more intelligent than humans. But it misses essential aspects of the intelligence concept as it is used in the context of intelligent natural systems like humans or real-world AI systems.

Our main goal here is to present variants of universal intelligence that better capture the notion of intelligence as it is typically understood in the context of real-world natural and artificial systems. The first variant we describe is *pragmatic general intelligence*, which is inspired by the intuitive notion of intelligence as "the ability to achieve complex goals in complex environments," given in (Goe93). After assuming a prior distribution over the space of possible environments, and one over the space of possible goals, one then defines the pragmatic general intelligence as the expected level of goal-achievement of a system relative to these distributions. Rather than measuring truly broad mathematical general intelligence, pragmatic general intelligence measures intelligence in a way that's specifically biased toward certain environments and goals.

Another variant definition is then presented, the *efficient pragmatic general intelligence*, which takes into account the amount of computational resources utilized by the system in achieving its intelligence. Some argue that making efficient use of available resources is a defining characteristic of intelligence, see e.g. (Wan06).

A critical question left open is the characterization of the prior distributions corresponding to everyday hu-

man reality; we have given a semi-formal sketch of some ideas on this in a prior conference paper (Goe09), where we present the notion of a "communication prior," which assigns a probability weight to a situation S based on the ease with which one agent in a society can communicate S to another agent in that society, using multimodal communication (including verbalization, demonstration, dramatic and pictorial depiction, etc.). We plan to develop this and related notions further.

Finally, we present a formal measure of the "generality" of an intelligence, which precisiates the informal distinction between "general AI" and "narrow AI."

Legg and Hutter's Definition of General Intelligence

First we review the definition of general intelligence given in (LH07b), as the formal setting they provide will also serve as the basis for our work here.

We consider a class of active agents which observe and explore their environment and also take actions in it, which may affect the environment. Formally, the agent sends information to the environment by sending symbols from some finite alphabet called the *action space* Σ; and the environment sends signals to the agent with symbols from an alphabet called the *perception space*, denoted $\mathcal{P}$. Agents can also experience rewards, which lie in the *reward space*, denoted $\mathcal{R}$, which for each agent is a subset of the rational unit interval.

The agent and environment are understood to take turns sending signals back and forth, yielding a history of actions, observations and rewards, which may be denoted

$$a_1 o_1 r_1 a_2 o_2 r_2 ...$$

or else

$$a_1 x_1 a_2 x_2 ...$$

if x is introduced as a single symbol to denote both an observation and a reward. The complete interaction history up to and including cycle t is denoted $ax_{1:t}$; and the history before cycle t is denoted $ax_{<t} = ax_{1:t-1}$.

The agent is represented as a function $\pi =$ which takes the current history as input, and produces an action as output. Agents need not be deterministic, an agent may for instance induce a probability distribution over the space of possible actions, conditioned on the current history. In this case we may characterize the agent by a probability distribution $\pi(a_t|ax_{<t})$. Similarly, the environment may be characterized by a probability distribution $\mu(x_k|ax_{<k}a_k)$. Taken together, the distributions π and μ define a probability measure over the space of interaction sequences.

To define universal intelligence, Legg and Hutter consider the class of environments that are *reward-summable*, meaning that the total amount of reward they return to any agent is bounded by 1. Where r_i denotes the reward experienced by the agent from the environment at time i, the *expected total reward* for the agent π from the environment μ is defined as

$$V_\mu^\pi \equiv E(\sum_1^\infty r_i) \leq 1$$

To extend their definition in the direction of greater realism, we first introduce a second-order probability distribution ν, which is a probability distribution over the space of environments μ. The distribution ν assigns each environment a probability. One such distribution ν is the Solomonoff-Levin universal distribution in which one sets $\nu = 2^{-K(\mu)}$; but this is not the only distribution ν of interest. In fact a great deal of real-world general intelligence consists of the adaptation of intelligent systems to particular distributions ν over environment-space, differing from the universal distribution. We then define

Definition 1. *The* **biased universal intelligence** *of an agent π is its expected performance with respect to the distribution ν over the space of all computable reward-summable environments, E, that is,*

$$\Upsilon(\pi) \equiv \sum_{\mu \in E} \nu(\mu) V_\mu^\pi$$

Legg and Hutter's **universal intelligence** is obtained by setting ν equal to the universal distribution.

This framework is more flexible than it might seem. E.g. suppose one wants to incorporate agents that die. Then one may create a special action, say a_{666}, corresponding to the state of death, to create agents that

- in certain circumstances output action a_{666}
- have the property that if their previous action was a_{666}, then all of their subsequent actions must be a_{666}

and to define a reward structure so that actions a_{666} always bring zero reward. It then follows that death is generally a bad thing if one wants to maximize intelligence. Agents that die will not get rewarded after they're dead; and agents that live only 70 years, say, will be restricted from getting rewards involving long-term patterns and will hence have specific limits on their intelligence.

Connecting Legg and Hutter's Model of Intelligent Agents to the Real World A notable aspect of the Legg and Hutter formalism is the separation of the reward mechanism from the cognitive mechanisms of the agent. While commonplace in the reinforcement learning literature, this seems psychologically unrealistic in the context of biological intelligences and many types of machine intelligences. Not all human intelligent activity is specifically reward-seeking in nature; and even when it is, humans often pursue complexly constructed rewards, that are defined in terms of their own cognitions rather than separately given. Suppose a certain human's goals are true love, or world peace,

and the proving of interesting theorems – then these goals are defined by the human herself, and only she knows if she's achieved them. An externally-provided reward signal doesn't capture the nature of this kind of goal-seeking behavior, which characterizes much human goal-seeking activity (and will presumably characterize much of the goal-seeking activity of advanced engineered intelligences also) ... let alone human behavior that is spontaneous and unrelated to explicit goals, yet may still appear commonsensically intelligent.

One could seek to bypass this complaint about the reward mechanisms via a sort of "neo-Freudian" argument, via

- associating the reward signal, not with the "external environment" as typically conceived, but rather with a portion of the intelligent agent's brain that is separate from the cognitive component
- viewing complex goals like true love, world peace and proving interesting theorems as indirect ways of achieving the agent's "basic goals", created within the agent's memory via subgoaling mechanisms

but it seems to us that a general formalization of intelligence should not rely on such strong assumptions about agents' cognitive architectures. So below, after introducing the pragmatic and efficient pragmatic general intelligence measures, we will propose an alternate interpretation wherein the mechanism of external rewards is viewed as a theoretical test framework for assessing agent intelligence, rather than a hypothesis about intelligent agent architecture.

In this alternate interpretation, formal measures like the universal, pragmatic and efficient pragmatic general intelligence are viewed as *not* being directly applicable to real-world intelligences, because they involve the behaviors of agents over a wide variety of goals and environments, whereas in real life the opportunity to observe an agent's activities are much more limited. However, they are viewed as being *indirectly* applicable to real-world agents, in the sense that an external intelligence can observe an agent's real-world behavior and then *infer* its likely intelligence according to these measures.

In a sense, this interpretation makes our formalized measures of intelligence the opposite of real-world IQ tests. An IQ test is a quantified, formalized test which is designed to approximately predict the informal, qualitative achievement of humans in real life. On the other hand, the formal definitions of intelligence we present here are quantified, formalized tests that are designed to capture abstract notions of intelligence, but which can be approximately evaluated on a real-world intelligent system by observing what it does in real life.

Pragmatic General Intelligence

To formalize pragmatic general intelligence, the first modification we need to introduce to Legg and Hutter's framework is to allow agents to maintain memories (of finite size), and at each time step to carry out internal actions on their memories as well as external actions in the environment. Legg and Hutter, in their theory of universal intelligence, don't need to worry about memory, because their definition of intelligence doesn't take into account the computational resource usage of agents. Thus, in their framework, it's acceptable for an agent to determine its actions based on the entire past history of perceptions, actions and rewards. On the other hand, if an agent needs to conserve memory and/or memory access time, it may not be practical for it to store its entire history, so it may need to store a sample thereof, and/or a set of memory items representing useful abstractions of its history. If one is gauging intelligence using a measure that incorporates space and time resource utilization, then the size and organization of this memory become important aspects of the system's intelligence.

Further extending the Legg and Hutter framework, we introduce the notion of a *goal-seeking agent*. We define goals as mathematical functions (to be specified below) associated with symbols drawn from the alphabet $\mathcal{G}$; and we consider the environment as sending goal-symbols to the agent along with regular observation-symbols. (Note however that the presentation of a goal-symbol to an agent does not necessarily entail the explicit communication to the agent of the contents of the goal function. This must be provided by other, correlated observations.) We also introduce a conditional distribution $\gamma(g, \mu)$ that gives the weight of a goal g in the context of a particular environment μ.

In this extended framework, an interaction sequence looks like

$$m_1 a_1 o_1 g_1 r_1 m_2 a_2 o_2 g_2 r_2 ...$$

or else

$$w_1 y_1 w_2 y_2 ...$$

if w is introduced as a single symbol to denote the combination of a memory action and an external action, and y is introduced as a single symbol to denote the combination of an observation, a reward and a goal.

Each goal function maps each finite interaction sequence $I_{g,s,t} = a y_{s:t}$ with g_s corresponding to g, into a value $r_g(I_{g,s,t}) \in [0, 1]$ indicating the value or "raw reward" of achieving the goal during that interaction sequence. The total reward r_t obtained by the agent is the sum of the raw rewards obtained at time t from all goals whose symbols occur in the agent's history before t. We will use "context" to denote the combination of an environment, a goal function and a reward function.

If the agent is acting in environment μ, and is provided with g_s corresponding to g at the start of the time-interval $T = \{i \in (s, ..., t)\}$, then the *expected goal-achievement* of the agent, relative to g, during the interval is the expectation

$$V^{\pi}_{\mu,g,T} \equiv E(\sum_{i=s}^{t} r_g(I_{g,s,i}))$$

where the expectation is taken over all interaction sequences $I_{g,s,i}$ drawn according to μ. We then propose

Definition 2. *The* **pragmatic general intelligence** *of an agent π, relative to the distribution ν over environments and the distribution γ over goals, is its expected performance with respect to goals drawn from γ in environments drawn from ν; that is,*

$$\Pi(\pi) \equiv \sum_{\mu \in E, g \in \mathcal{G}, T} \nu(\mu)\gamma(g,\mu)V^{\pi}_{\mu,g,T}$$

(in those cases where this sum is convergent).

This definition formally captures the notion that "intelligence is achieving complex goals in complex environments," where "complexity" is gauged by the assumed measures ν and γ.

If ν is taken to be the universal distribution, and γ is defined to weight goals according to the universal distribution, then pragmatic general intelligence reduces to universal intelligence.

Furthermore, it is clear that a universal algorithmic agent like AIXI (Hut05) would also have a high pragmatic general intelligence, under fairly broad conditions. As the interaction history grows longer, the pragmatic general intelligence of AIXI would approach the theoretical maximum; as AIXI would implicitly infer the relevant distributions via experience. However, if significant reward discounting is involved, so that near-term rewards are weighted much higher than long-term rewards, then AIXI might compare very unfavorably in pragmatic general intelligence, to other agents designed with prior knowledge of ν and γ in mind.

The most interesting case to consider is where ν and γ are taken to embody some particular bias in a real-world space of environments and goals, and this biases is appropriately reflected in the internal structure of an intelligent agent. Note that an agent need not lack universal intelligence in order to possess pragmatic general intelligence with respect to some non-universal distribution over goals and environments. However, in general, given limited resources, there may be a tradeoff between universal intelligence and pragmatic intelligence. Which leads to the next point: how to encompass resource limitations into the definition.

One might argue that the definition of Pragmatic General Intelligence is already encompassed by Legg and Hutter's definition because one may bias the distribution of environments within the latter by considering different Turing machines underlying the Kolmogorov complexity. However this is not a general equivalence because the Solomonoff-Levin measure intrinsically decays exponentially, whereas an assumptive distribution over environments might decay at some other rate. This issue seems to merit further mathematical investigation.

Incorporating Computational Cost

Let $\eta_{\pi,\mu,g,T}$ be a probability distribution describing the amount of computational resources consumed by an agent π while achieving goal g over time-scale T. This is a probability distribution because we want to account for the possibility of nondeterministic agents. So, $\eta_{\pi,\mu,g,T}(Q)$ tells the probability that Q units of resources are consumed. For simplicity we amalgamate space and time resources, energetic resources, etc. into a single number Q, which is assumed to live in some subset of the positive reals. Space resources of course have to do with the size of the system's memory, briefly discussed above. Then we may define

Definition 3. *The* **efficient pragmatic general intelligence** *of an agent π with resource consumption $\eta_{\pi,\mu,g,T}$, relative to the distribution ν over environments and the distribution γ over goals, is its expected performance with respect to goals drawn from γ in environments drawn from ν, normalized by the amount of computational effort expended to achieve each goal; that is,*

$$\Pi_{Eff}(\pi) \equiv \sum_{\mu \in E, g \in \mathcal{G}, Q, T} \frac{\nu(\mu)\gamma(g,\mu)\eta_{\pi,\mu,g,T}(Q)}{Q} V^{\pi}_{\mu,g,T}$$

(in those cases where this sum is convergent).

Efficient pragmatic general intelligence is a measure that rates an agent's intelligence higher if it uses fewer computational resources to do its business.

Note that, by abandoning the universal prior, we have also abandoned the proof of convergence that comes with it. In general the sums in the above definitions need not converge; and exploration of the conditions under which they do converge is a complex matter.

Assessing the Intelligence of Real-World Agents

The pragmatic and efficient pragmatic general intelligence measures are more "realistic" than the Legg and Hutter universal intelligence measure, in that they take into account the innate biasing and computational resource restrictions that characterize real-world intelligence. But as discussed earlier, they still live in "fantasy-land" to an extent – they gauge the intelligence of an agent via a weighted average over a wide variety of goals and environments; and they presume a simplistic relationship between agents and rewards that does not reflect the complexities of real-world cognitive architectures. It is not obvious from the foregoing how to apply these measures to real-world intelligent systems, which lack the ability to exist in such a wide variety of environments within their often brief lifespans, and mostly go about their lives doing things other than pursuing quantified external rewards. In this brief section we describe an approach to bridging this gap. The treatment is left-semi-formal in places.

We suggest to view the definitions of pragmatic and efficient pragmatic general intelligence in terms of a "possible worlds" semantics – i.e. to view them as asking, counterfactually, how an agent *would* perform, hypothetically, on a series of tests (the tests being goals, defined in relation to environments and reward signals).

Real-world intelligent agents don't normally operate in terms of explicit goals and rewards; these are abstractions that we use to think about intelligent agents. However, this is no objection to characterizing various sorts of intelligence in terms of counterfactuals like: how would system S operate if it were trying to achieve this or that goal, in this or that environment, in order to seek reward? We can characterize various sorts of intelligence in terms of how it can be inferred an agent would perform on certain tests, even though the agent's real life does not consist of taking these tests.

This conceptual approach may seem a bit artificial, but, we don't currently see a better alternative, if one wishes to quantitatively gauge intelligence (which is, in a sense, an "artificial" thing to do in the first place). Given a real-world agent X and a mandate to assess its intelligence, the obvious alternative to looking at possible worlds in the manner of the above definitions, is just looking *directly* at the properties of the things X has achieved in the real world during its lifespan. But this isn't an easy solution, because it doesn't disambiguate which aspects of X's achievements were due to its own actions versus due to the rest of the world that X was interacting with when it made its achievements. To distinguish the amount of achievement that X "caused" via its own actions requires a model of causality, which is a complex can of worms in itself; and, critically, the standard models of causality also involve counterfactuals (asking "what would have been achieved in this situation if the agent X hadn't been there", etc.) (MW07). Regardless of the particulars, it seems impossible to avoid counterfactual realities in assessing intelligence.

The approach we suggest – given a real-world agent X with a history of actions in a particular world, and a mandate to assess its intelligence – is to introduce an additional player, an *inference agent* δ, into the picture. The agent π modeled above is then viewed as π_X: the model of X that δ constructs, in order to explore X's inferred behaviors in various counterfactual environments. In the test situations embodied in the definitions of pragmatic and efficient pragmatic general intelligence, the environment gives π_X rewards, based on specifically configured goals. In X's real life, the relation between goals, rewards and actions will generally be significantly subtler and perhaps quite different.

We model the real world similarly to the "fantasy world" of the previous section, but with the omission of goals and rewards. We define a *naturalistic* context as one in which all goals and rewards are constant, i.e. $g_i = g_0$ and $r_i = r_0$ for all i. This is just a mathematical convention for stating that there are no precisely-defined external goals and rewards for the agent. In a naturalistic context, we then have a situation where agents create actions based on the past history of actions and perceptions, and if there is any relevant notion of reward or goal, it is within the cognitive mechanism of some agent. A *naturalistic agent* X is then an agent π which is restricted to one particular naturalistic context, involving one particular environment μ (formally, we may achieve this within the framework of agents describe above via dictating that X issues constant "null actions" a_0 in all environments except μ).

Next, we posit a metric space (Σ_μ, d) of naturalistic agents defined on a naturalistic context involving environment μ, and a subspace $\Delta \in \Sigma_\mu$ of inference agents, which are naturalistic agents that output predictions of other agents' behaviors (a notion we will not fully formalize here). If agents are represented as program trees, then d may be taken as edit distance on tree space (Bil05). Then, for each agent $\delta \in \Delta$, we may assess

- the prior probability $\theta(\delta)$ according to some assumed distribution θ
- the effectiveness $p(\delta, X)$ of δ at predicting the actions of an agent $X \in \Sigma_\mu$

We may then define

Definition 4. *The* **inference ability** *of the agent δ, relative to μ and X, is*

$$q_{\mu,X}(\delta) = \theta(\delta)\frac{\sum_{Y \in \Sigma_\mu} sim(X,Y)p(\delta,Y)}{\sum_{Y \in \Sigma_\mu} sim(X,Y)}$$

where sim is a specified decreasing function of $d(X,Y)$, such as $sim(X,Y) = \frac{1}{1+d(X,Y)}$.

To construct π_X, we may then use the model of X created by the agent $\delta \in \Delta$ with the highest inference ability relative to μ and X (using some specified ordering, in case of a tie). Having constructed π_X, we can then say that

Definition 5. *The inferred pragmatic general intelligence (relative to ν and γ) of a naturalistic agent X defined relative to an environment μ, is defined as the pragmatic general intelligence of the model π_X of X produced by the agent $\delta \in \Delta$ with maximal inference ability relative to μ (and in the case of a tie, the first of these in the ordering defined over Δ). The inferred efficient pragmatic general intelligence of X relative to μ is defined similarly.*

This provides a precise characterization of the pragmatic and efficient pragmatic intelligence of real-world systems, based on their observed behaviors. It's a bit messy; but the real world tends to be like that.

Intellectual Breadth: Quantifying the Generality of an Agent's Intelligence

We turn finally to a related question: How can one quantify the degree of generality that an intelligent agent possesses? There has been much qualitative discussion of "General AI" or "Artificial General Intelligence," versus "Narrow AI" (GP05), and intelligence

as we have formalized it here is specifically a variety of general intelligence, but we have not yet tried to quantify the notion of generality versus narrowness.

Given a triple (μ, g, T), and a set Σ of agents, one may construct a fuzzy set $Ag_{\mu,g,T}$ gathering those agents that are intelligent relative to the triple ; and given a set of triples, one may also also define a fuzzy set Con_π gathering those triples with respect to which a given agent π is intelligent. The relevant formulas are:

$$\chi_{Ag_{\mu,g,T}}(\pi) = \chi_{Con_\pi}(\mu, g, T) = \sum_Q \frac{\eta_{\mu,g,T}(Q) V^\pi_{\mu,g,T}}{Q}$$

One could make similar definitions leaving out the computational cost factor Q, but we suspect that incorporating Q is a more promising direction. We then propose

Definition 6. *The* **intellectual breadth** *of an agent π, relative to the distribution ν over environments and the distribution γ over goals, is*

$$H(\chi^P_{Con_\pi}(\mu, g, T))$$

where H is the entropy and

$$\chi^P_{Con_\pi}(\mu, g, T) =$$

$$\frac{\nu(\mu)\gamma(g, \mu)\chi_{Con_\pi}(\mu, g, T)}{\sum_{(\mu', g', T')} \nu(\mu_\alpha)\gamma(g', \mu')\chi_{Con_\pi}(\mu', g', T')}$$

is the probability distribution formed by normalizing the fuzzy set $\chi_{Con_\pi}((\mu, g.T))$.

A similar definition of the intellectual breadth of a context (μ, g, T), relative to the distribution σ over agents, may be posited. A weakness of these definitions is that they don't try to account for dependencies between agents or contexts; perhaps more refined formulations may be developed that account explicitly for these dependencies.

Note that the intellectual breadth of an agent as defined here is largely independent of the (efficient or not) pragmatic general intelligence of that agent. One could have a rather (efficiently or not) pragmatically generally intelligent system with little breadth: this would be a system very good at solving a fair number of hard problems, yet wholly incompetent on a larger number of hard problems. On the other hand, one could also have a terribly (efficiently or not) pragmatically generally stupid system with great intellectual breadth: this would be a system that was roughly equally dumb in all the contexts under study.

Thus, one can characterize an intelligent agent as "narrow" with respect to distribution ν over environments and the distribution γ over goals, based on evaluating it as having low intellectual breadth. A "narrow AI" relative to ν and γ would then be an AI agent with a relatively high efficient pragmatic general intelligence but a relatively low intellectual breadth.

Conclusion

Our goal here has been to push the formal understanding of intelligence in a more pragmatic direction. More work remains to be done, e.g. in specifying the environment, goal and efficiency distributions relevant to real-world systems, but we believe that the ideas presented here constitute nontrivial progress.

If the line of research pursued here succeeds, then eventually, one will be able to do AGI research as follows: Specify an AGI architecture formally, and then use the mathematics of general intelligence to derive interesting results about the environments, goals and hardware platforms relative to which the AGI architecture will display significant pragmatic or efficient pragmatic general intelligence, and intellectual breadth.

References

Philip Bille. A survey on tree edit distance and related problems. *Theoretical Computer Science*, 337, 2005.

Ben Goertzel. *The Structure of Intelligence*. Springer, 1993.

Ben Goertzel. The embodied communication prior. In *Yingxu Wang and George Baciu (eds), Proceedings of ICCI-09, Hong Kong*, 2009.

Ben Goertzel and Cassio Pennachin. *Artificial General Intelligence*. Springer, 2005.

Marcus Hutter. *Universal AI*. Springer, 2005.

Shane Legg and Marcus Hutter. A collection of definitions of intelligence. In *Ben Goertzel and Pei Wang (eds), Advances in Artificial General Intelligence*. IOS, 2007.

Shane Legg and Marcus Hutter. A formal measure of machine intelligence. In *Proceedings of Benelaam-2006, Ghent*, 2007.

Stephen Morgan and Christopher Winship. *Counterfactuals and Causal Inference*. Cambridge University Press, 2007.

Pei Wang. *Rigid Flexibility: The Logic of Intelligence*. Springer, 2006.