Preface



When I began studying artificial intelligence, I started with what the experts were doing. I researched neural nets and read theses on the nature of consciousness. I participated in programming competitions for game AI and recommendation engines. But after years of work, my creations never seemed *intelligent*. I wanted to build something conscious; something smart in the same way that people are smart. But toward that goal, nobody appeared to be any more successful than myself.

I realized I was trying to create artificial intelligence before sufficiently understanding natural intelligence. I am not an outsider to this field. I earned academic degrees in biology and education, so it was humbling to admit that my knowledge of intelligent behavior was embarrassingly poor. I decided to stop researching and not restart before I had a concise, logical, and evidence-based definition.

Instead of a universally accepted definition of intelligence, I found research being conducted under many different and incompatible meanings. Intelligence studies were stubbornly trying to move forward by brute force instead of first attempting to answer the central question of "What is intelligence?"



I concluded that if I wanted clear, precise, and accurate definitions of key intelligence concepts, their emergent properties, and relationships to each other, I had no

alternative but to grudgingly make them myself. The result is this booklet, *The Foundations of Behavioral Logic*.

Imagining feelings predicted by a scenario.

Time Threshold

When a decision is made after a predetermined amount of time.

U

Urgency

An estimate of the amount of time before a **decision** is needed.

Understanding

The accuracy (its predictions compared to reality) and efficiency of beliefs.

WXYZ

Wild

Highly random behavior due to neglect.

CORE CONCEPTS

Definition Requirements

To define *intelligence* as best as I can, I'd better figure out the difference between a good and a poor definition. To make my definitions as clear and rigorous as possible, I did my best to follow to these rules:

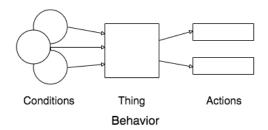
- 1. **No Jargon** Instead of inventing a new word, the most similar common word should be used. Webster's Unabridged Dictionary 1913 and 1828 editions were chosen as the standard reference.
- 2. **No Ambiguity** Each definition should only contain words that have commonly agreed upon meaning, or a definition of vague or controversial terms must be provided.
- 3. **No False Negatives** No example that would reasonably be considered to belong to a definition, but does not match the definition, should exist.
- 4. **No False Positives** No example that would reasonably **not** be considered to belong to a definition, but matches the definition, should exist.
- 5. **Observable Only** It cannot be inherently impossible to test a definition through the creation of a practical implementation.
- 6. **Parsimony** If part of a definition could be removed without decreasing accuracy or increasing ambiguity, then it should be removed.

I'll start with basic terms like *behavior*, *action*, *goal*, and *benefit* but please don't take them for granted. They lay the foundation for more advanced definitions. By the end, when

I define *consciousness*, *understanding*, *ideas*, and other higher level terms, if I have done my job well, they should seem as underwhelming and unimpressive as every other definition.

Behavior

Behavior is a quality of a thing where a set of conditions results in an action. It describes the cause and effect relation for a particular medium. Maybe a simple diagram is clearer.



A behavior may be intelligent or it may not be. If you push a rock, the effect might be for it to roll down a hill. This is a behavior of the rock, though it is **nonintelligent**. The rock has no other options to react to the pushing, nor does it have the ability to choose one.

Push a dog, and you can't be certain what reaction might occur. It is **intelligent** and has the power of choice.

Let's stop referring to things with intelligent behavior as *things*. The ability to make choices earns them special recognition. Let's use the term **creature** to refer to a thing that has intelligence.

S

Scenario

A combination of a situation and an idea.

Science

Increasing understanding.

Sensor

A mechanism for **measuring** the environment.

Simulation

Finding the consequent situation of a scenario using beliefs.

Simulational Learning

Acquiring **behaviors** using a model of the environment.

Situation

A group of time related information.

Slop

Randomness in behavior.

Spoiling

Random or excessive feelings of **pleasure** leading to **eccentric** behavior.

Step

One action in a **plan**.

Subjective Sensor

A sensor that evaluates behavior.

T

Training

The process of evaluating **sloppy** behaviors by trial and error, then **tuning** their blurry action parameters to increase benefit.

Tune

To decrease the randomness of **blurry action parameter** in behavior toward **beneficial** parameters.

Thinking

The process of creating thoughts to reach a decision.

Thought

N

Non-intelligent Behavior

Behavior with no choice of actions.

Novelty

A propensity to choose previously unattempted actions.

Neglect

An inadequate amount of experience during childhood.

0

Opinion

A feeling associated with a situation from past experience.

P

Programmer

A source of **learning** external to a **creature**.

Pleasure

A **feeling** indicating **benefit**.

Pain

A feeling indicating harm.

Play

Participation in a game.

Plan

A sequence of actions produced from a chain of simulations.

Prediction

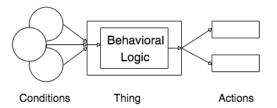
The consequent situation in a creature's **beliefs** that correspond to a particular **situation** and **action**.

Premature Conclusion

An error in **thinking** where the **decision** threshold is too low resulting in less beneficial actions than could easily be chosen.

Q

For a creature to display intelligent behavior, it must have some kind of mechanism that selects actions using information from its environment. Let's refer to that as *Behavioral Logic*.



A **brain** is a working implementation of behavioral logic. It could be built from any appropriate material: nerve cells, metal wires and switches, even mechanical gears and levers.

Behavior Tables

A simple brain is a **behavior table**; an ordered list of behaviors.

Let's imagine a creature with two types of sensors that can respond with either *yes* or *no*. The creature could also have three possible actions we will call *A*, *B* and *C*. Its behavior table might look like this:

Table 1

Situation Sensor A+B	Action
no · no	_
no · yes	Α
yes · no	В
yes · yes	С

A group of coinciding sense information I call a **situation**. A behavior table associates each unique situation with an **action**. Actions can be any sort of chemical, physical, or electrical change. An action could also affect behaviors. Different behavior tables could be chosen if the creature happens to be hungry, sleepy, afraid, or in any other variation of its environment. Let's say action *C* causes the creature to use table 2, and action *D* changes it back to the first table.

Table 2

Situation Sensor A+B	Action
no · no	D
no · yes	В
yes · no	В
yes · yes	В

A good name for these temporary alternate behavior tables is **Mood**.

A harmful habit formed by insufficient and/or anomalous experiences.

Imagination

A **situation** that does not represent the current environment.

Imitation

Generating an idea from an observation of another creature.

Indecision

A error in **thinking** where the threshold to make a **decision** can never be met.

Information

Measurements and **estimations** of a creature's environment converted to a format useable by its **brain**.

Instinct

An **intelligent behavior** acquired from an external programmer.

Intelligent Behavior

Behavior that is chosen from more than one option. (Antonym: non-intelligent behavior.)

Intelligence

Choosing actions to accomplish a goal.

J

K

L

Learn

To increase the **benefit** (or reduce **harm**) of behaviors.

M

Measure

To convert an environmental property into **information** using a strongly correlated phenomenon.

Mood

A persistent effect on a behavior table.

Experience

An instance of a behavior and its resulting feelings.

F

Feeling

Information from subjective sensors.

G

Game

A **situation** that approximates those expected in **adulthood**, but with no (or much lower) risk of **harm**.

Goal

The ultimate purpose of a **creature's** existence.

Webstser (1913) Goal 2. The final purpose or aim; the end to which a design tends...

Guess

An idea that is more than a slight variation on what is known to have worked in the past.

Gut Reaction

When a **simulational** learner chooses to not **think** and act according to **habit**.

H

Harm

Something that that decreases a **goal's** chance of success. (Antonym: **beneficial**.)

Habit

A behavior tuned from experience.

Idea

A proposed **action** to be used in a **simulation**.

Idiosyncrasy

Estimating Behavior Effectiveness

Just because a creature has a brain, there is no guarantee that it will make good choices. How can we assess if a behavior is beneficial or harmful?

Unlike physical properties such as temperature and light, there isn't a simple way to directly measure how "good" or "bad" a behavior is. Benefit is relative to what a particular creature is trying to accomplish. To evaluate the effects of a behavior, we can observe if it increases or decreases the odds that the creature's primary **goal** will succeed.

Webster's 1913 edition has an elegant definition of Goal.

Goal 2. The final purpose or aim; the end to which a

Estimating a behavior's effect on the odds of success of a goal is called **behavior evaluation**. If a behavior increases the chance of success, it is **beneficial**. If not, the behavior is **harmful**.

If a robot's reason for existence is to vacuum my basement, one measure of its effectiveness could be the amount of dust on the floor. A vacuum-bot with beneficial behaviors should leave less dust than one with harmful behaviors. A biological organism's primary goal is to reproduce. A living

creature's Darwinian fitness is typically measured by its number of offspring that also reproduce.

It may not be obvious if a behavior is beneficial or not if the creature's goal requires a long time to accomplish. The best assessment of behavior may require weighing many short and long term benefits and detriments.

Intelligence

Now that we have covered some basic terms related to behavior, we are ready for the *Behavioral Logic* definition of intelligence.

Intelligence - Choosing actions to accomplish a goal.

There are many definitions of intelligence, but this one is mine. It should describe anything with even the smallest amount of intelligence.

Learning

Where does intelligence come from? Some process must exist to create, evaluate, and modify behaviors. We will call this process **learning**.

Three possible ways for a creature to learn are:

- Being **programmed** by an external process or other intelligent creature.
- Training from evaluating first-hand experiences.

Benefit

Something that increases a goal's chance of success. (Antonym: harm.)

Benefit Threshold

Making a **decision** as soon as a predetermined benefit is evaluated.

Blurry Action Parameter

A **sloppy** aspect of a behavior meant to be **tuned** from **experience**.

Brain

A working implementation of behavioral logic.

C

Childhood

The initial time in a **beast's** existence before **experience** has **tuned** random behavior.

Consciousness

The ability for a **creature** to **evaluate imaginary situations**.

Creature

A thing that exhibits **intelligent behavior**.

D

Decision

Stopping the **thinking** process after a **time** or **benefit** threshold is reached to act on the best **idea**.

Depression

A overall decrease in the likelihood that a beast will choose any action.

E

Eccentric

Possessing idiosyncratic behavior.

Estimate

To convert an environmental property into **information** using one or more weakly correlated phenomena.

Evaluation

To estimate the effect on the odds of success of a goal.

Execute

Perform steps of a **plan** in order.

GLOSSARY

A

Abuse

Random or excessive pain.

Abort

Halting the execution of a **plan** before all steps are complete due to observed divergence from simulated results.

Action

A physical, chemical, electrical, or **mood** change.

Adulthood

The period of time in a **beast's** existence after sufficient behavior **tuning**.

Automaton

A **creature** that lacks the ability to **learn**.

B

Beast

A creature that can form habits.

Behavior

A quality of a thing where a set of conditions produces an **action**. A cause and effect relation for a particular medium.

Behavior: Conditions -> Thing -> Action

A collection of all of a creature's behaviors.

Behavior Damage

Harmful behaviors as a result of insufficient or anomalous experience.

Behavioral Logic

A method of choosing an action.

Behavior Table

A group of situations and associated actions.

Belief

An **prediction** of how a particular action will cause a **situation** to change.

• **Simulational** by imagining experience using a model of its environment.

The next three sections will address each of these learning processes including possible mechanisms, emergent effects, advantages, and disadvantages.

INTELLIGENT MACHINES

Learning Method I: Externally Programmed

Simple creatures lack the ability to evaluate and modify their own behaviors. They can only learn through the effects of other creatures and external processes such as natural selection (for living organisms) and human computer programmers. Whether intelligent or not, we will refer to all learning processes outside of to a creature as **programmers**.

Because they are executed without thought or opinion from the creature, programmed behaviors are **instincts**. From Webster's 1913:

Instinct

 $\ensuremath{\text{n.}}$ 2. the natural, unreasoning, impulse by which an

Programmed behaviors are not so much learned as taught.

Advantages of Programming

Learning from programmers is the simplest and most direct way for a brain to acquire beneficial behaviors.

9

Optimal

I have mentioned building artificially intelligent robots (and intend this document to be a high-level specification for such) but the greatest benefit of a thorough and accurate understanding of intelligence is probably not a robot butler (as cool as that is.) I believe the greatest benefit is a deeper understanding of ourselves. Powerful, simple and accurate models of our behavior would be a boon to the study and treatment of mental illness and for coaxing the most out of our greatest asset: the most powerful brains on the planet.

26

There is no known way to generate ideas that guarantee at least one is beneficial in finite time and certainly not in the brief time creatures have to accomplish their goal. As far as I know, producing ideas is just random guessing.

When to make a decision may be the most hazardous step of the thinking process. **Indecision** can lock thinking in a never ending loop. **Premature conclusion**s select the first barely acceptable action, even if one significantly better could be discovered with a little more thought.

No wonder people have so many psychological issues.

Now what?

I hope my attempt at describing my theoretical model of intelligence, *The Foundations of Behavioral Logic*, was clear and not entirely unpleasant to read. I am sure I could have written more clearly, but I have no desire to use rhetorical tricks to coerce anyone into accepting my beliefs if they are not worthy. If I have persuaded you, thank you for reading with an open mind. But even if I have not, the next step is the same: test these concepts with penetrating questions and measurable observations. If what I have described is valuable, then robots built using the concept of *behavioral logic* should perform better than those that aren't, and intelligent biological organisms should have analogous processes in the structure and function of their neuroanatomy. I am look forward to actively challenging these ideas, and hope you join me.

Efficient

Because programmed behavior does not require a creature to invest its own time and energy learning, an exclusively programmed creature can focus its efforts exclusively on the execution of behaviors.

Fast

Not only can behaviors without the burden of complex learning processes perform faster, programmed behavior does not require any first-hand experience. Programmed behaviors are immediately ready to use without any lessons or practice.

Disadvantages of External Programming

Inflexibility

An externally programmed creature does not have the ability to handle any situation it was not explicitly programmed for. If conditions change or it finds itself in a different habitat, it has no other option but to wait for its programmer to update its behavior. Behavioral logic that relies entirely on external programming is typically only effective in a single, stable environment.

Most chess programs can easily defeat me, but if I change the rules so pawns can teleport to any free space and the winning condition is moving the king to the middle of the board, a human could quickly adjust their strategy and easily beat a chess program. Stupid computer.

And this is the best case—when the programmer is doing a competent job. Poor programmers will create poor behaviors that the creature is powerless to improve.

External Programming - Summary

In addition to behavior tables, external programming can be considered to include all parts of a creature that it cannot modify itself, such as:

- Physical form
- Sensory apparatus and transmission
- Brain hardware
- Available actions

Because creatures that rely exclusively on programmed behaviors can only follow instructions and are incapable of learning anything on their own, we will refer to them as automatons. result of each action matches the predicted outcomes of each step of the plan.)

Simulation Advantages

Creatures with brains that are capable of simulation have beliefs about their world. They can imagine, think, decide, get ideas, plan, and imagine feelings of pleasure and pain. They are aware of the consequences of their actions. I believe this amount of self-awareness is enough to consider them **conscious**.

Conscious creatures can evaluate the benefit of behaviors faster and with less risk than beasts. They are better at handing novel situations and can explore a wider variation of behavior without risk of harm.

Simulation Disadvantages

Learning from simulations adds significant complexity to behavioral logic systems, increasing both the time and energy required for the execution of behaviors. Because of this increased complexity, more can go wrong. Here are a few possible problems.

A creature could have inaccurate beliefs. This can result in decisions that are imagined to be beneficial, but are disastrous in practice.

idea. So that's one idea, but what if that action results in simulated pain? Variations in proposed actions can be made in the same way as the blurriness of trained behaviors.

If minor variations all evaluate poorly, it may have to search for ideas that are significantly different than its current habits and instincts. Random actions and large variation of action variables are possible ways to **guess**. Imagining what would happen if nothing was done isn't a bad idea either.

If a creature has sufficient understanding of its environment, it could also get ideas by **imitating** others' behavior.

Planning Using Successive Simulations

A simulation does not have to end once an action is chosen. Instead of acting on the decision, a creature could instead start a new simulation using the prediction from a previously chosen action, and then think about the best action in the ensuing scenario. It could then create a new scenario from the results of that decision, and so on.

Plan - A sequence of actions (**steps**) produced from a chain of simulations.

For important decisions and the luxury of time to think, successive simulation can be used to generate plans. These plans can then be **executed** by performing the series of actions in immediate succession without the need to stop and think after each action. (As long as the intermediate

HABITS OF THE BEASTS

Learning Method II: Trained

Creatures that must rely on an external programmer will struggle in environments where conditions change unpredictably and frequently. In chaotic environments, creatures need the flexibility to alter their behavior in response to current conditions.

A simple behavior **training** mechanism has these components.

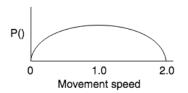
- 1. Initial **blurriness** (random variation) in behavior.
- 2. Specialized **subjective senses** that don't simply report the facts, but indicate benefit or harm.
- 3. The ability to **tune** (reduce variation toward optimal parameters) blurry parameters toward increased benefit.

Blurriness

Unless there is an initial randomness in behavior, there is no opportunity for a creature to explore outside of its original programming.

One way to implement blurriness is to randomize **action parameters**variations of actions that accept a range of values. For example, an action of moving an appendage

could have a variable speed of movement. It could be initially programmed with a wide range of random values to unpredictably move slow, medium, or fast.



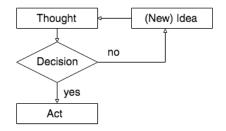
Some behaviors could be blurrier than others. The most important and predictable behaviors should have very little variation. No creature should have to learn to breathe through trial and error.

Subjective Senses

To judge if the result of an action was beneficial or harmful, sensors don't transmit an objective measurement, but a subjective judgement. Unlike most visual and audial information, subjective senses indicate if something feels good or bad, so we call information received from subjective sensors **feelings**.

In the previous example, stress sensors could cause the creature to feel **pain** to warn of imminent damage if the appendage moves too fast. Chemical receptors could transmit a feeling of **pleasure** if the action resulted in touching a desirable target.

Gaining Experience



After a simulation, a **decision** must be made to either stop thinking and act, or consider a new idea. One way might be to choose the first action that exceeds a predetermined **benefit threshold** when we want to select the first idea that is "good enough."

With the luxury of time, a creature could continue to think about situations even after a beneficial action is found in order to search for an innovative solution of greater benefit.

In **urgent** situations with limited time, the most beneficial (or least harmful) idea could be chosen after a short **time threshold**.

In dire emergencies, a **gut reaction** could skip the entire thinking process and act out of habit.

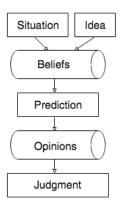
Getting Ideas

All that thinking won't do any good unless the creature can propose an idea that is beneficial. Where do new ideas come from?

If a creature is already trained by experience, it already knows of a trained action that could be used for the first

22

action and the current situation can be used to locate the corresponding belief in the beliefs table. From that, we can **predict** a resulting situation.



Once we have a predicted situation, we need to judge if that situation puts the creature in a beneficial or harmful position. Fortunately, it should already have an opinions table that correlates situations with a subjective sensory measurement. It can repurpose this table to judge how good or bad the predicted results of their actions should be.

Making a Decision

But what should be done if the simulated feelings were bad? We need a way to find a beneficial action.

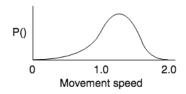
One incidence of a situation, the chosen action (and variable values) and resulting subjective senses is called an **experience**. Positive experiences should increase the probability of the same action variables being selected in future situations. Painful experiences should decrease the chance of using those variable values again.

Here are 4 hypothetical experiences of the appendage movement example.

Experiences

Speed	Stress	Taste
1.8	yes	yes
1.1	no	yes
0.5	no	no
1.3	no	yes

If we use these experiences to tune the original chances (decreasing the likelihood of dangerous movement speeds and increasing those that tasted food) our new distribution of movement speeds should look more like this.



Now this situation is neither reacted to out of instinct, nor indiscriminately random. After tuning through experience, it

Childhood

Childhood is the period of time at start of a creature's life characterized by maximum randomness of behavior. Children are at a significant disadvantage when competing against either programmed automatons or adults with behaviors tuned from experience. One way to protect children during this time is to use games—easier and safer versions of their environment. Puppies play by wrestling with each other long before needing to fight for territory. Cats bring injured prey to their kittens for easy hunting practice.

Another way to decrease the risks of childhood is simply to make as many children as possible and hope some will survive to adulthood.

As a creature's experience increases, blurriness decreases and subjective senses become less important. Adults with adequate experience should behave more like automatons.

Opinions

Tuning behavior through experience works, but only for behaviors that result in immediate benefit or harm. Is there a way to give our beasts the wisdom to look ahead to longer term benefit?

Beliefs

Initial Situation	Action	Consequent Situation
no · no	move randomly	no · yes
no · no	no action	no · no
yes · no	move randomly	yes · no
yes · no	no action	yes · no
yes · no	move toward smell	yes · yes

Each row of this table represents a **belief** that can predict the effects of an action. A creature's **understanding** of its environment is a measure of how complete, accurate, and efficient its beliefs are.

This is probably a good place to reiterate that my examples are meant to illustrate the simplest working implementation. Their purpose is to show that a practical implementation is possible, and to help explain the concept. I assume that real working brains will be much more sophisticated, but advanced versions should still follow the same basic principles and, more importantly, perform the same function, only better.

Simulating Behavior

A **thought** is the transformation of a situation and a proposed action into predicted subjective sense data. It starts with an **idea** for a possible action. This proposed

THE THINKERS

Learning Method III: Simulational

The major drawback of programmed behaviors is inflexibility. The time required for external programmers to make changes can be too great for creatures living in frequently changing or a variety of environments.

Training overcomes this limitation, but at great cost of time, energy, and risk to acquire at best sub-optimal behaviors. What habitual learners need is a way to evaluate the benefit of new behaviors quickly and safely. They need the ability to imagine a situation and predict the benefit of an action.

Beliefs

A simple way to predict the effects of actions would be to remember how situations changed after an action was performed. This is **science** in its most basic form.

One way would be, instead of relying only on subjective senses, to recall feelings associated with a past situation. That information could be used to create subjective **opinions** of objective situations.

Opinions		
Situation	Subjective Sense	
no · no	pain (hunger)	
no · yes	-	
yes · no	-	
yes · yes	pleasure (good taste of food)	

A vertebrate's optic nerve cannot transmit pain, but if it relays a scene that resembles a previous pleasant experience, a creature could feel the same pleasure. A chess program could examine board state and remember if a similar configuration in the past was associated with a win or loss.

Habit Advantages

Creatures that form behaviors through experience can adapt to a changing environment without having to wait on an outside programmer.

This also allows trainable creatures to thrive in different environments with the same initial programming. Creatures that can form habits are generalists.

Habit Disadvantages

There are some advantages to learning through experience, but the drawbacks are numerous.

A significant amount of time at the beginning of trainable creatures' existence is spent with haphazard, possibly harmful, behaviors; wasting time and resources on actions that do not directly further its primary goal. During this time, mistakes with devastating consequences can too easily be made.

Habit-forming creatures are not only threatened by physical damage, but risk behavioral damage—harmful behaviors acquired from poor training. Anomalous experiences can create idiosyncrasies—behaviors tuned to harmful instead of beneficial effect. Insufficient training can leave highly random, wild behavior into adulthood. Bad behaviors can render a creature as incapable of accomplishing its goal as one physically broken.

Even though learning from experience is considered a *higher* form of intelligence than programmed (it requires more complex behavioral logic than simple behavior tables) it often results in behaviors that are far from maximally beneficial. The more behaviors a creature has, the more likely that some will not be adequately trained before adulthood. It should be no surprise that habit forming creatures frequently display harmful behaviors. It is hoped that the benefit of many good behaviors outweigh the effects of a small number of bad ones.

The amount of initial random variation of a habit forming creature's behaviors and actions should be set to an optimal amount. Too little, and a behavior is indistinguishable from programmed. Too much, and the search space is too large to be tuned in time for adulthood.

Subjective Senses are only a rough estimator of benefit or harm. Due to inherent inaccuracy in behavior evaluation, tuning to subjective sensory information may not actually be increasing the odds of success of the primary goal.

The problems associated with learning through trial-anderror are numerous and devastating. Why would any creature want to accept the risk, energy and time investment for mediocre behaviors? There is only one explanation: The benefit of habit forming is greater than the harm, at least in certain environments.

Habitual Learning - Summary

Habitual learners are much more than automatons. They feel pleasure and pain. They remember their experiences. They can form long term strategies by associating objective patterns with failure or success. They deserve a better name.

Beast - A creature that can form habits.

It's not flattering, but better than *automaton*. From Webster's 1913 edition:

Beast 3. As opposed to man: Any irrational animal.

