Chapter 1

COMMON SENSE

Abstract from a forthcoming book on logical AI

The main obstacle to getting computer programs with human level intelligence is that we don't understand yet how to give them common sense. Without common sense, no amount of computer power will give human level intelligence. Once programs have common sense, improvements in computer power and algorithm design will be directly applicable to making them more intelligent.

This chapter is an informal summary of various aspects of common sense. Formalisms will be given in later chapters.

1.1 What is common sense?

Common sense is a certain collection of reasoning abilities, perhaps other abilities, and knowledge.

In [**?**] I wrote that the computer programs that had been written up to 1958 lacked common sense. Common sense has proved to be a difficult phenomenon to understand, and most programs of 2004 also Lack common sense or have only a little. In the 1959 paper, I wrote "We shall therefore say that **a program has common sense if it automatically deduces for itself a sufficiently wide class of immediate consequences of anything it is told and what it already knows.**"[1]

Programs with common sense à la [**?**] are still lacking, and, moreover, the ideas of that paper are not enough. Logical deduction is insufficient, and nonmonotonic reasoning is required. Common sense knowledge is also required. Here's what I think is a more up-to-date formulation.

A program has common sense if it has sufficient common sense knowledge of the world and suitable inference methods to infer a sufficiently wide class of immediate consequences of anything it is told and what it already knows.

Requiring some intelligence as part of the idea of common sense gives another formulation.

[1] At least much that people consider obvious is not deduced.

A program has common sense if it can act effectively in the *common sense informatic situation*, using the available information to achieve its goals.

1.2 The common sense informatic situation

A program that decides what to do has certain information built in, gets other information from its inputs or observations; still other information is generated by reasoning. Thus it is in a certain *informatic situation*. If the information that has to be used has a common sense character, it will be in what we call the *common sense informatic situation*.

We need to contrast the *common sense informatic situation* with the less general *bounded informatic situations*.

Formal theories in the physical sciences deal with *bounded informatic situations*. A scientist decides informally in advance what phenomena to take into account. For example, much celestial mechanics is done within the Newtonian gravitational theory and does not take into account possible additional effects such as outgassing from a comet or electromagnetic forces exerted by the solar wind. If more phenomena are to be considered, scientists must make new theories—and of course they do.

Likewise present AI formalisms work only in a bounded informatic situations. What phenomena to take into account is decided by a person before the formal theory is constructed. With such restrictions, much of the reasoning can be monotonic, but such systems cannot reach human level ability. For that, the machine will have to decide for itself what information is relevant, and that reasoning will inevitably be partly nonmonotonic.

One example is the "blocks world" where the position of a block x is entirely characterized by a sentence $At(x, l)$ or $On(x, y)$, where l is a location or y is another block. The language does not permit saying that one block is partly on another.

Another example is the MYCIN [**?**] expert system in which the ontology (objects considered) includes diseases, symptoms, and drugs, but not patients (there is only one), doctors or events occurring in time. Thus MYCIN cannot be told that the previous patient with the same symptoms died. See [**?**] for more comment.

Systems in a bounded informatic situation are redesigned from the outside when the set of phenomena they take into account is inadequate. However, there is no-one to redesign a human from the outside, so a human has to be able to take new phenomena into account. A human-level AI system needs the same ability to take new phenomena into account.

In general a thinking human is in what we call the *common sense informatic situation*. The known facts are necessarily incomplete. [2] There is no *a priori*

[2]We live in a world of middle-sized objects which can only be partly observed. We only partly know how the objects that can be observed are built from elementary particles in general, and our information is even more incomplete about the structure of particular objects.

limitation on what facts are relevant. It may not even be clear in advance what phenomena should be taken into account. The consequences of actions cannot be fully determined.

An astrophysicist can learn that computing the orbit of a comet during its passage near the sun requires taking into account the forces resulting from gases boiled of the comet by the sun's heat. He can either deal with this informally, regarding the orbit as somewhat uncertain, or he can make a new mathematical theory taking outgassing into account. In either case, he remains outside the theory. Not only that, he can look at his own thinking and say, "How dumb I was not to have noticed ...."

Should it be needed, a human manipulating blocks can discover or be told that it is necessary to deal with configurations in which one block is supported by two others.

MYCIN's lack of common sense is yet more blatant, because its language cannot represent facts about events occurring in time.

The common sense informatic situation has the following features.

1. The theory used by the agent is open to new facts and new phenomena.

2. The objects and other entities under consideration are incompletely known and are not fully characterized by what is known about them.

3. Most of the entities considered are intrinsically not fully defined.

4. In general the informatic situation itself is an object about which facts are known. This human capability is not used in most human reasoning, and very likely animals don't have it.

The difficulties imposed by these requirements are the reason why the goal of Leibniz, Boole and Frege to use logical calculation as the main way of deciding questions in human affairs has not yet been realized. Realizing their goal will require extensions to logic beyond those required to reason in bounded informatic situations. Computer programs operating in the common sense informatic situation also need tools beyond those that have been used so far.

Here are some of the characteristics of such systems and some of the tools.

elaboration tolerance The languages used require *elaboration tolerance*. It must be possible to add facts without scrapping the theory and starting over. Elaboration tolerance seems to impose requirements on the languages used, i.e. on the set of predicate and function symbols.

These limitations apply to any buildable machines, so the problem is not just one of human limitations.

Science fiction and scientific and philosophical speculation have often indulged in the *Laplacian fantasy* of super-beings able to predict the future by knowing the positions and velocities of all the particles. That isn't the direction to go. Rather the super-beings would be better at using the information that is available to the senses—maybe having more and more sensitive senses.

nonmonotonic reasoning Elaboration tolerance imposes one requirement on the logic, and this is the ability to do *nonmonotonic reasoning*. The system must reach conclusions that further facts not contradicting the original facts are used to correct.

Taking into account only some of the phenomena is a nonmonotonic reasoning step. It doesn't matter whether phenomena not taken into account are intentionally left out or if they are unknown to the reasoner.

While nonmonotonic reasoning is essential for both man and machine, it can lead to error when an important fact is not taken into account. These are the errors most often noticed.

[3]

approximate concepts and approximate objects The *common sense informatic situation* necessitates the use of *approximate concepts* that cannot be fully defined and the use of *approximate theories* involving them.

reasoning in contexts and about contexts In bounded theories, the context is fixed at the time the theory is created. Therefore, the reasoner doesn't have to switch contexts. For example, the theorists undertakes to decide on a consistent notation. There are exceptions to this in computer programming wherein information is encapsulated in classes, but the external behavior of the classes is prescribed by the designer.

An agent in the common sense informatic situation is often confronted with new contexts.

approximate objects

composition of objects Consider an object composed of parts. It is convenient logically when what we knew about the parts and how they are put together enables us to determine the behavior of the compound object.

[3]Here's an extended example from the history of science.

Starting in the middle of the 19th century, Lord Kelvin (William Thomson) undertook to set limits on the age of the earth. He had measurements of the rate of increase of temperature with depth and of the thermal conductivity of rock. He started with the assumption that the earth was originally molten and computed how long it would have taken for the earth to cool to its present temperature. He also took into account gravitational contraction as a source of energy. He obtained numbers like 25 million years. This put him into conflict with geologists who already had greater estimates based on counting annual layers in sedimentary rock.

Kelvin's calculations were correct but gave the wrong answer, because no-one until Becquerel's discovery in 1896 knew about radioactive decay. Radioactive decay is the main source of energy that keeps the earth hot.

Kelvin's reasoning was nonmonotonic. Namely, he took into account all the sources of energy the science of the day knew about.

Nonmonotonic reasoning is necessary in science as in daily life. There can always be phenomena we don't know about. Indeed there might be another source of energy in the earth besides radioactivity.

Experience tells us that careful nonmonotonic reasoning, taking into account all the sources of information we can find and understand, usually gives good results, but we can never be as certain as we can be of purely mathematical results.

Indeed this is often true in science and engineering and is often the goal of the search for a scientific theory. The common sense informatic situation is not so convenient logically. The properties of an object are often more readily available than the properties of the parts and their relations. Formalizations of facts about the relation between structured objects and their parts must not require that all facts about the objects be inferred from known facts about the parts.

A person knows that a certain object is composed of parts. He knows something about the structural relations and about the parts. Physically, the parts and their relations make up the object. If we knew all about these, we would know the object and its potential behavior. However, actual knowledge often runs the other way. We know more about the object than about its parts.

For example, a baseball has a visible and feelable surface, and we can see and feel the seams and can feel its compliance and its simplest heat transfer properties. We also know, from reading or from seeing a baseball disassembled, something about its innards. However, this knowledge of structure is less usable than the knowledge of the baseball as a whole.

It would be logically simpler if we knew about the structures of the objects in our environment and could establish the properties of the whole object from its structure. Thus it is quite pleasing that the properties of molecules follow from the properties of atoms and their interactions. Unfortunately, the common sense world is informatically more complex. We learn about complex objects and go from there to their structures and can only partly discover the structures.

This is not any kind of grand complementarity of the kind Bohr and his followers mistakenly tried to extend from quantum mechanics where it is one usable perspective. Moreover, this limitation on knowledge will apply to robots in a similar way as to humans.

This phenomenon, of often knowing more about the whole than about the parts, applies to more than physical objects. It can apply to processes. The phenomenon even existed in mathematics. Euclid's geometry was a powerful logical structure, but the basic concepts were fuzzy.

Formalization of facts about the relation between structured objects and their parts must not require that all facts about the objects be inferred from known facts about the parts.

knowledge of physical objects

knowledge of regions in space

knowledge of other actors

introspective knowledge

bounded informatic situations in contexts Bounded informatic situations have an important relation to the common sense informatic situation. For example, suppose there are some blocks on a table. They are not perfect cubes and they are not precisely aligned. Nevertheless, a simple blocks world theory may be useful for planning building a tower by moving and painting blocks. The bounded theory of the simple blocks world in which the blocks are related only by the $on(x, y, s)$ relation is related to the common sense informatic situation faced by the tower builder. This relation is conveniently expressed using the theory of contexts as objects discussed in Chapter . The blocks world theory holds in a subcontext *cblocks* of the common sense theory c, and sentences can be *lifted* in either direction between c and *cblocks*.

1.3 Localization

Maybe this section should be moved closer to discussions of nonmonotonic reasoning.

We do not expect events on the moon to influence the physical location of objects on the table. However, we can provide for the possibility that an astronomer looking through a telescope might be so startled by seeing a meteorite collide with the moon that he would fall off his chair and knock an object off the table. Distant causality is a special phenomenon. We take it into account only when we have a specific reason.

Closer to hand, we do not expect objects not touching or connected through intermediate objects to affect each other. Perhaps there is a lot of common sense knowledge of the physical motion of table scale ojects and how they affect each other that needs to be expressed as a logical theory.

1.3.1 The objects that are present

In Section **??** on circumscription, we discussed what objects can fly. We have

$$\begin{array}{l} \neg Ab\ Aspect1\ x \rightarrow \neg flies\ x, \\ Bird\ x \rightarrow Ab\ Aspect1\ x, \\ Bird\ x \wedge \neg Ab\ Aspect2\ x \rightarrow flies\ x, \\ Penguin\ x \rightarrow Bird\ x, \\ Penguin\ x \rightarrow Ab\ Aspect2\ x \\ , Penguin\ x \wedge \neg Ab\ Aspect3\ x \rightarrow \neg flies\ x, \text{etc.} \end{array} \tag{1.1}$$

When we do the circumscription we certainly want to vary $flies$. If we leave the predicates *bird* and *penguin* constant and vary only $flies$, we conclude that the objects that fly are precisely the birds that are not penguins.

If we are reasoning within a sufficiently limited context this is ok. However, we can't have it in a general purpose common sense knowledge base, because it excludes bats and airplanes from fliers and ostriches from the non-fliers.

1.4 Remarks

Scientists and philosophers of science have often criticized *common sense* as inferior science. Indeed common sense notions of falling bodies not incorporating the discoveries of Galileo and his successors give wrong answers.[4]

Consider the fact that a glass dropped from a table will hit the floor, perhaps hard enough to shatter. Galileo tells us the falling will take a little less that 1/2 seconds, but we haven't much use for this fact in cautioning ourselves to avoid pushing the glass off the table. The common sense notions of physics are needed by a robot that can function in a home. Very likely if we could compute faster and estimate distances and times quantitatively, we could get substantial advantage from knowledge of elementary mechanics. Since we can't, humans have to make do with qualitative reasoning and practice skills that have some quantitative elements which are present but not explicit.

A person or program has common sense if it can act effectively in the *common sense informatic situation*, using the available information to achieve its goals.

Achieving human-level AI faces a problem that is generally ignored in building scientific theories and in philosophy. The theory builder or the philosopher of science discusses theories from the outside. However, for an AI system of human level, there is no outside. All its reasoning, including its metamathematical reasoning must be done in its own language, just as we humans discuss human reasoning in our own languages.

Common sense operates in a world with the following characteristics.

Humans and robots that humans build are middle-sized objects in a physical world that contains other purposeful entities. We have common sense information about this world and our possibilities for action and we need to give our robots this common sense information. This information is partial and will usually be partly wrong. We are, and our robots will be in the ¡em common sense informatic situation¡/em.

Distinguish between the facts about the world, which include atomic structure and how general relativity interacts with quantum mechanics and what facts are available for common sense use.

Basic facts about the Common sense informatic situation

1. The world in which common sense operates has the aspects described in the following items.

2. Situations are snapshots of part of the world.

3. Events occur in time creating new situations. Agents' actions are events.

4. Agents have purposes they attempt to realize.

5. Processes are structures of events and situations.

[4]One service mathematics has rendered to the human race. It has put common sense back where it belongs, on the topmost shelf next to the dusty canister labelled 'discarded nonsense'. —E. T. Bell. What did Bell mean by common sense?

6. 3-dimensional space and objects occupy regions. Embodied agents, e.g. people and physical robots are objects. Objects can move, have mass, can come apart or combine to make larger objects.

7. Knowledge of the above can only be approximate.

8. The csis includes mathematics and physics, i.e. abstract structures and their correspondence with structures in the real world.

9. Common sense can come to include facts discovered by science. Examples are conservation of mass and conservation of volume of a liquid.

10. Scientific information and theories are imbedded in common sense information, and common sense is needed to use science.

The common sense informatic situation includes at least the following.

1. The facts that may have to be used to decide what to do are not limited to an initially given set. New facts may have to be obtained.

2. The set of concepts required to act intelligently in a situation may have to be extended.

3. The ability to take into account new facts without complete reorganization of the knowledge base is called ¡EMelaboration tolerance.¡/EM The need for elaboration tolerance affects the logical formalisms.

4. Introspection about the methods one has been using may be required. Indeed the complete mental state up to the present moment sometimes has to be regarded as an object that can be reasoned about.

1.4.1 Refinement

Here are some truisms. They are all obvious, but formalisms must take them into account, and how to do this is often not obvious.

Consider a person's knowledge of his body.

He can see his arms, legs and other external parts. He cannot see his liver or his brain, but this is an accident of the senses humans happen to have. If we had ultrasonic detection abilities of bats, we might be able to see our internal organs. Indeed with artificial aids we can see them.

The organs are made up of tissues. Common sense knows little of them, but medical science knows a lot.

The tissues are made of cells. A lot is known about this.

Cells have a structure above the level of molecules, but not much is known about this.

Once we get down to the domain of molecules we are in chemistry which is based on atomic physics.

Below that is elementary particle physics. Whether that is the bottom is unknown.

Objects depend for their integrity and for their perceivable qualities on certain processes going on within them and between them and their neighbors. All material objects depend on the motion of their molecules for their temperatures, and their structural integrity often depends on temperature.

Human common sense can deal with all these phenomena. Some objects we can perceive well, others partly and others not at all. We are aware of objects as structured from parts.