**Title: The Scope and Limits of Simulation in Cognition and Automated Reasoning**

**Authors:**

Ernest Davis, Dept. of Computer Science, New York University, New York, NY, davise@cs.nyu.edu, http://www.cs.nyu.edu/faculty/davise

Gary Marcus, Dept. of Psychology, New York University, New York, NY, gary.marcus@nyu.edu, http://www.psych.nyu.edu/gary/

**Abstract:**

It has often been argued that simulation plays a central role in a wide range of cognitive tasks, including physical reasoning and natural language understanding, and correspondingly that simulation should be used as the basis for carrying out comparable tasks in artificial intelligence (AI) systems. Although computer simulation is undeniably a powerful tool, we argue here, focusing primarily on physical reasoning, that there are many cases in which simulation can play at most a limited role, both in cognition and in automated reasoning. In the analysis of the limits of simulation in automated physical reasoning, we argue that simulation is most effective when the task is prediction, when complete information is available, when a reasonably high quality theory is available, and when the range of scales involved, both temporal and spatial, is not extreme. When these conditions do not hold, simulation is less effective or entirely inappropriate. We discuss twelve features of physical reasoning problems that pose challenges for simulation-based reasoning, and show that simulation-based accounts of physical reasoning in human cognition are vulnerable to the same challenges. We also review evidence that suggests that in some instances humans do not use simulation, or do not use it very effectively, even when in principle simulation should be possible. We argue, finally, that simulation-based accounts of natural language understanding may suffer from similar limitations. Such considerations do not militate against simulation-based approaches altogether, but do suggest that other mechanisms, such as knowledge-based qualitative reasoning, may also play a critical role.

**1. Introduction**

Computer simulations – broadly speaking, computations in which the trajectory of a temporally evolving system is traced in detail – have become ubiquitous. Programmers have created extremely detailed simulations of the interactions of 200,000,000 deformable red blood cells in plasma (Rahimian, et al., 2010); the air flow around the blades of a helicopter (Murman, Chan, Aftosmis, & Meakin, 2003); the interaction of colliding galaxies (Benger, 2008); and the injuries caused by the explosion of an IED under a tank (Tabiei & Nilakantan, undated). Software, such as NVidia PhysX, that can simulate the interactions of a range of materials, including rigid solid objects, cloth, and liquids, in real time, is available for the use of game designers as off-the-shelf freeware (Kaufmann & Meyer, 2008). In artificial intelligence (AI) programs, simulation has been used for physical reasoning (Johnston & Williams, 2007), (Nyga & Beetz, 2012), robotics (Mombauri & Berns, 2013), motion tracking (Vondrak, Sigal, & Jenkins, 2008), and planning (Zicker & Veloso, 2009).

In cognitive psychology, likewise, simulation in a much broader sense has been proposed as the primary mechanism underlying physical reasoning, reasoning about other minds, language comprehension, and many other cognitive functions (Markman, Klein, & Suhr, 2009). For instance, Kaup, Lüdtke, and Maienborn (2010) suggest that “creating simulations is necessary for [the] comprehension” of language, and Battaglia et al (2013, p 18327) propose a model of physical reasoning

based on an “intuitive physics engine,” a cognitive mechanism similar to computer engines that simulate rich physics in video games and graphics, but that uses approximate, probabilistic simu- lations to make robust and fast inferences.

Similarly, Sanborn et al (2013) propose that “people’s judgments [about physical events such as colliding objects] are based on optimal statistical inference over a Newtonian physical model that incorporates sensory noise and intrinsic uncertainty about the physical properties of the objects being viewed.”

For certain reasoning tasks, simulation is unquestionably a powerful tool and an intuitively appealing one; however, it is important to recognize its limitations. Here, focusing on physical reasoning, we analyze the scope and limits of simulation as a theory of human reasoning and a technique of automated reasoning. We develop a categorization of tasks where simulation works well and is plausible as a cognitive mechanism; tasks where simulation does not work at all, and is therefore impossible as a cognitive mechanism; tasks where simulation could work, but other techniques are both more effective and more plausible as cognitive mechanisms; and tasks where simulation could work, but experimental evidence suggests that humans in fact use other mechanisms.

**2. Computer simulation of physical systems**

 In a typical computer simulation, the input is a detailed description of an initial scenario. The program then uses the dynamic laws of the domain to extrapolate an equally detailed description of the state of the scenario a short time later. The program continues to extrapolate each state to the next state until some stopping condition is met. The program returns the entire trajectory of states as its prediction of what will happen. Table 1 shows a description of this process in pseudocode.

*function Simulate(StartingState; StartingTime; BoundaryConditions, TerminationCondition)*

 *s ← StartingState;*

 *t ← StartingTime;*

 *Trajectory ← [〈t,s〉];*

 *repeat {*

 *Δ← chooseTimeStep(s);*

 *s ← projectForward(s,delta,BoundaryConditions);*

 *t ← t + Δ;*

 *add 〈t,s〉 to end of Trajectory;*

 *} until TerminationCondition;*

 *return Trajectory;*

*end*

Table 1: Algorithm 1: Physical simulation

For instance, consider a ball falling to the ground. In this simulation the initial state at a given time t is specified in terms of the height of the ball, x(t), and its velocity v(t), both measured upward. To extrapolate from one state at time t to the next state at time t+Δ, we calculate that the height decreases by Δ times the current downward velocity, and that the downward velocity increased by Δ times the acceleration of gravity, denoted g.

x(t+Δ) = x(t) + Δ\*v(t)

 v(t+Δ) = v(t) – Δ\*g

The simulation stops when x(t)≤0, since at that point the ball has hit the ground.

Some simulations, such as assessments of the aerodynamics of airplanes, aim for high precision; they are extremely specialized in terms of both the physical phenomena and the kinds of scenario under consideration; and they involve immense computational burdens. Others, such as simulations used for real-time animation, particularly in video games, aim at plausible images rather than physical precision, often in real time on a personal laptop, rather than off-line on supercomputers.

AI programs that deal with physical objects often use simulations, with good reason. Although numerous technical difficulties exist (many described below), simulation is conceptually and methodologically simple and comparatively straightforward to implement. Furthermore they can be used directly to produce a viewable animation, which is very helpful both for the end-user and for program development and debugging; moreover, physics engines of ever-increasing power, quality, and scope are publicly available for use. In some circumstances, they represent an ideal solution.

**3. Simulation: Challenges in Automated Systems**

It is easy, however, for the non-expert to overestimate the state of the art of physical simulation, and assume that there is a plug-and-play physics engine that works for pretty much any physical situation. Although physics engines are now commonplace in video games, when it comes to the real-world, their fidelity is often quite limited; plug-and-play engines capture only narrowly-defined environments; more sophisticated applications require hard work from experts. A few seconds of realistic CGI in a disaster film may well require several person-days of work; an accurate and complex scientific computation may require several person-months. Nils Thuerey (personal communication) writes,

There are ... inherent difficulties with these simulations: we are still very far from being able to accurately simulate the complexity of nature around us. Additionally, the numerical methods that are commonly used are notoriously difficult to fine-tune and control.

Plug-and-play physics engines are also subject to bugs and anomalies,[[1]](#footnote-1) and may require careful tuning to work correctly. In a systematic evaluation of seven physics engines, Boeing and Bräunl (2007) found that all seven gave significantly and obviously erroneous answers on certain simple problems involving solid objects.

In this section, we review twelve challenges that arise in the construction of simulations, some widely known, others less so; together, they help to articulate the scope and limits of when simulation can and cannot serve as an appropriate tool for physical reasoning – with important implications for cognition, as explicated in section 4.

**3.1 The challenge of finding an appropriate modeling approach**

The first hurdle in implementing a simulator is developing a domain model. In some cases, this is well understood. However choosing an appropriate model is often difficult, even for familiar objects, materials, and physical processes. The theory of physical systems that are changing temperature rapidly ("non-equilibrium thermodynamics") currently has very large gaps; the theory of liquids and gasses also has significant gaps. Even the theory of rigid solid objects, the simplest kinds of materials encountered in everyday activities, has some gaps.[[2]](#footnote-2).

Even in mundane situations, it may be challenging to find adequate models. Consider, for instance, cutting materials with tools. An ordinary household has perhaps a dozen kinds of tools for cutting: a few kinds of kitchen knives; more specialized kitchen equipment such as graters and peelers; a few kinds of scissors; a drill, a saw, a lawn mower, and so on. (A specialist, such as a carpenter or a surgeon, has many more.) Most people understand how they should be used and what would happen if you used the wrong tool for the material; if, for example, you tried to cut firewood with a scissors. But it would be hard to find good models for these in the physics or engineering literature.

**3.2 The challenge of discretizing time**

Most simulation algorithms employ a discrete model of time:[[3]](#footnote-3) The timeline consists of a sequence of separated instants. In some instances, converting continuous physical time[[4]](#footnote-4) into a discrete model is unproblematic, but in a surprisingly broad range of problems, difficulties arise from this conversion.

Consider, for example, the problem of choosing a proper time increment Δ in simulating rigid objects. If Δ is chosen too small, then many time steps must be calculated, increasing the computational burden. If Δ is too large, two objects may collide, interpenetrate, and pass through one another between one time point and the next. For instance, suppose that you are holding one compact disc (X) in the air, and you drop another (Y) directly on top of it from 1 meter above. By the time Y reaches X, it is travelling at a speed of about 4.5 m/sec. If the time increment is greater than a third of a millisecond, the collision will be missed (figure 1). As Boeing and Braunl (2007) demonstrated, current physics engines are not immune to this kind of error.



Figure 1: Object Y “passes through” X between successive instants

Alternatively, one can calculate exactly what is the time to the next collision or state change. But this kind of calculation can be extremely challenging.[[5]](#footnote-5)

Discretizing time can also lead to more subtle problems. For instance if the simulation of a rigid pendulum uses the simplest method for the updating of the physical state from one time point to the next, the simulation will incorrectly predict that the pendulum swings back and forth a few times, reaching a higher angle on each swing, until eventually it rotates in a full vertical circle in one direction. A more sophisticated updating procedure is required to avoid this. In domains with richer physics than a pendulum, such as fluid dynamics, these kinds of problems can arise in much more complex forms (Bridson, 2008) (Hairer, Lubich, & Wanner, 2006).

**3.3 The challenge of discontinuous dynamics**

In some problems, a small change to the starting situation leads to a correspondingly small change in the overall trajectory. For instance, if you change the angle of a cannon by a small amount, the path of the ball changes only slightly. In other problems, two nearly identical starting situations can lead to significantly different behaviors. In these, enormous precision is required in both measurement and simulation to ensure an accurate answer. Consider the problem of a rolling die, which is the archetype of a physical process whose outcome is hard to predict, and in which slight differences in initial conditions can lead to entirely different outcomes. Although it is relatively easy to carry out an approximate computer simulation of a die rolling and to render it in an animation, it is extremely difficult to accurately predict the outcome of an actual roll of dice, even if the starting conditions are specified precisely (Kapitaniak, Strzalko, Grabski, & Kapitaniak, 2012).

A related problem is that simulators can make predictions that are correct in a mathematical sense but impossible physically because of instability. For example, all seven simulators tested by Boeing and Bräunl (2007) predicted incorrectly that if three spheres were dropped one exactly on top of the next, they would stack.

**3.4 The challenge of choosing an idealization**

Virtually all simulations represent idealizations; in some, friction is ignored, in others three dimensions are abstracted as two. In most situations, many different idealizations are possible; and the idealization should be chosen so that, on the one hand, the calculation is not unnecessarily difficult, and on the other, the important features of the situation are preserved. Consider, for instance, the simulation of a pendulum on a string. If you are using an off-the-shelf physics engine, then you would use the idealizations that the engine prescribes. For example, a typical engine might model the bob as an extended rigid shape and model the string as an abstract constraint requiring that the distance between the objects tied at opposite ends not exceed a fixed length. In that case, simulating the pendulum would require you to specify the structure of the pendulum, the mass and shape of the bob, the length of the string, and the initial position and velocity of the bob. A student in freshman physics, by contrast, will probably approximate the bob as a point mass which is constrained to move on a circle of fixed radius; the resulting simulation will certainly be easier to carry out and quite possibly more accurate. However, to set up the simulation in this way, the student must understand the essential form of the behavior in advance, viz. that the string remains fully extended while the bob swings.

Other scenarios are more complex. A bob may swing in a horizontal circle; spin on the axis of the string; rotate about its center of mass like a yo-yo, or fly through the air (figure 2). The string itself may be taut, loose, tangled, knotted, or twisted; it may get in the way of the bob; it may even unravel or snap. Although these behaviors are familiar to anyone who has spent time playing with objects on strings, few if any existing physics engines support any but the taut and loose conditions of the string and perhaps snapping.



Figure 2: A pendulum in various conditions

Efficient reasoning about these different possible behaviors of the string and the bob requires using a variety of different idealizations. Space can be two-dimensional or three-dimensional. A bob can be idealized as a point object, a rigid object, or an elastic object. A string of length L can be idealized as an abstract constraint restricting the motion of the bob; a one-dimensional curve of length L, with or without mass; or a three-dimensional flexible object, either uniform or with some internal structure (e.g. twisted out of threads or a chain of links.) Influence on the system can be limited to gravity, or can include friction and air resistance. In looking at any one simulation that has been well-worked out, it is easy to lose sight of how much careful work goes on in choosing the right idealization; as yet there is no algorithmic way to guarantee an efficient solution for arbitrary problems. Using the most realistic model possible is no panacea; highly realistic models both require more laborious calculation and more detailed information.

Moreover, different high-quality physics engines can give radically difficult predictions for a single problem, particularly when the problem involves a device with feedback. Figure 3, from (Boeing & Bräunl, 2012) shows the paths calculated by three different physics engines for a submarine robot trying to follow a wall. These are radically different, even though, as Boeing and Bräunl emphasize, all three results come from “valid and accepted fluid simulation methods”.



Figure 3: Alternative predictions by three different simulators. From (Boeing & Bräunl, 2012)

**3.5 The challenge of rapidly drawing “easy” inferences**

The output of a simulation is always precise, though not always accurate. Often, however, the reasoner has no need for the level of precision that a simulation provides. Consider, for example. a child who has built a tower of blocks and is now planning to knock it over by hitting it with another block. The child does not generally need to know details of the final disposition of each individual block nor a probability distribution over possible final dispositions; it may suffice to know that the tower will fall with a satisfying clatter. Likewise if you ride a bicycle on a bumpy road while carrying a half-full closed water canteen, all that matters is that the water stays inside the canteen, not the trajectory of the water splashing inside the canteen. In such cases, fully detailed simulations seem like inefficient and inappropriate tools.

There are many kinds of rules that allow quick inference or quick transference of results from a known situation to a new one. *Invariance under time and space*: If a mechanism worked in a particular way at home on Monday, it will work in the same way at work on Tuesday. *Invariance under irrelevant changes*: If a jar fits on a shelf, and you fill it with pennies, it will still fit on the shelf*. Invariance under changes of scale* (for certain physical theories, such as kinematics): A large pair of gears works in the same way as a much smaller scale model. *Approximation*: If a jug holds a gallon of water, then another jug of similar dimensions and shape will hold about a gallon. *Ordering on a relevant dimension*: If a toy fits in a box, then it will fit in a larger box, under a proper definition of “larger” (Davis, 2013). *Decomposition:* If a system consists of two uncoupled subsystems, then one can reason about each separately.

There can also be *rules of thumb*: useful generalizations for common cases. For instance, if you spill a cup of coffee in your office, it should not be necessary to resort to simulation to infer that the coffee will not end up in some other office. Rather, one can use a rule of thumb that a small amount of liquid dropped from a moderate "height onto a horizontal surface will end up not very far from the point where it was dropped, where "small amount", "moderate height", and "not very far" have some very approximate quantitative measurements.

Moreover, these alternative forms of inference may do well under circumstances where simulation scales badly. Consider the following scenario: you put a number of objects in a box, close the lid, and shake it up and down violently. We now wish to infer that the objects are still in the box. Simulating the motion of the objects will become rapidly more complicated as the number of objects increases, the shapes of the objects become more complex, and the shaking of the box becomes more complex and violent. By contrast a single simple rule, "An object in a closed container remains in the container" suffices to carry out the inference.

**3.6 The challenge of incorporating extra-physical information**

In some cases, reasoning about a physical system can be carried out more reliably and effectively using extra-physical information. Suppose that you see a baseball pitcher throw a ball. If you simulate the motion of the ball, using the best information you can gather about the angle, velocity, and so on, of the ball when it left his hand, and factor in the imprecision of this information, you will predict that it has a rather low probability of ending up anywhere close to the batter. You would obtain better results ― predicting that the ball will end up close to the strike zone, just inside it or just outside ― by relying on the known accuracy of the pitcher, plus quite specific information about the state of play and the pitcher's willingness to risk an out-of-strike zone pitch rather than a hit.

**3.7 The challenge of incomplete information**

Carrying out a physical simulation is generally only possible if the geometric and physical characteristics of the initial condition are known precisely. In many common real-world situations, reasoning must be carried out on the basis of partial, sometimes extremely limited, information. Perception may be imperfect or incomplete. For instance, an object may be partially occluded. (An opaque object always self-occludes its own far side from the viewer.) Knowledge of the physical situation may come from natural language text or sketches. Knowledge of aspects of the situation may come from inference; for example, if you see someone attempt unsuccessfully to pick up a suitcase, you can infer that the suitcase is unusually heavy; you can then use that inference for future prediction. Or the precise details may not have been determined yet. For instance, suppose that you are going to the furniture store to buy a dining room table. You can reason that you will not be able to carry it home on foot or riding a bicycle, even though you have not yet chosen a particular table.

Of course, no representation is truly complete or entirely precise; in any representation, some aspects are omitted, some are simplified, and some are approximated. However, the simulation algorithm requires that the initial conditions of the scenario be *fully specified relative to a given level of description*. That is, the representational framework specifies some number of critical relations between entities and properties of entities. A complete representation of a situation relative to that framework enumerates all the entities that are relevant to the situation, and specifies all the relations in the framework that hold between those entities. The description must be detailed and precise enough that the situation at the next time step is likewise fully specified, in the same sense.

**3.7.1 Limits to Monte Carlo simulation as an approach to partial knowledge**

One common method for using simulations in situations with partial knowledge is to use Monte Carlo methods, or probabilistic simulation. Given a probability distribution over a space of physical situations, a program generates some large collection of specific situations following the distribution; it carries out a simulation based on each of these; and then it considers these behaviors as a random sample of the possible behaviors. Then for any particular proposition Φ of interest, the probability that Φ will be true in the actual situation can be estimated as the fraction of simulations in which Φ is true. Conditional probabilities can be computed in a similar way: to calculate the conditional probability P(Φ|λ), generate random scenarios satisfying λ and find the fraction of these that satisfy Φ. This technique can be adapted to situations in which one starts with a set of constraints rather than a distribution by using a distribution that is, in some sense, uniform over the space of situations satisfying the constraints.

For example, Battaglia, Hamrick, and Tenenbaum (2013) consider a situation where an observer is looking at a tower of blocks, but has imperfect information; specifically, the perceived position of the block varies from the true position by an error that follows a Gaussian distribution. Their program uses this distribution to generate a collection of different towers, and simulates the behavior of each of these hypothetical towers. The probability that the true tower is stable is estimated as the fraction of simulations in which the tower remains stable. Conditional probabilities such as "the probability that the blue block and red block end up less than an inch apart, given that the tower falls over" can be estimated by generating random towers, selecting those in which the tower falls over, and then, within that restricted set, considering the fraction in which the blue block and red block end up less than an inch apart.

Probabilistic simulation can often be effective, but it has its own problems. First, the concept of a random shape is problematic. If you know that a block is a rectangular solid, then you can reasonably specify the distributions of its length, width, and height as distributions of some standard form (e.g. Gaussians), or collectively as a joint distribution. However, if you have imperfect knowledge of the *shape* of the blocks, then it is much less clear how to proceed. There does not seem to be any natural probability distribution over the class of all shapes.

Second, Monte Carlo simulation runs the risk of missing scenarios that are important and possible, though improbable, particular if, as suggested in (Battaglia, Hamrick, & Tenenbaum, 2013), the number of simulations run is actually small (3-7). Suppose, for example, that you are standing underneath a scaffold next to a building, and you observe that a tool on the scaffold is precariously balanced. You wish to know whether you are safe, so you carry out a half-dozen simulations of the tool falling off the scaffold. If the tool does not hit you in any of them, you conclude, erroneously, that there is no need to worry.

Third, the calculation of the conditional probability P(Φ|λ) relies on generating random elements satisfying λ. However, if λ is only rarely satisfied, it may be difficult to generate these. For instance, suppose that you see a screw in a bolt, and you have reliable information that the screw fits properly in the bolt (that is, after all, a safe default assumption when a screw is on the bolt). You cannot, however, see the threads of either. You wish to infer something about the number of turns needed to remove the screw from the bolt; almost certainly greater than half a turn and almost certainly less than 50 turns. That is, you want to estimate P(Φk|λ) where Φk is the proposition, "It will require exactly k turns to remove the screw from the bolt" and λ is the proposition, "The screw fits on the bolt".

If you sample random shapes for the screw and the bolt separately, then you will never come across a pair in which the screw fits on the bolt; and if you have no samples for λ, you certainly cannot estimate P(Φk|λ). Rather, you must generate the shapes for the screw and bolt concurrently. But at this point you need a quite specialized Monte Carlo shape generator that can generate fitting screws and bolts simultaneously. If the generator is to be created by the automated reasoner, and not supplied by a human programmer, then that in itself would require powerful spatial and physical reasoning capacities that can hardly be the product of unaided simulation.

**3.7.2 Reasoning with incomplete information about containers**

One important domain of physical reasoning where reasoning with incomplete information is both common and effective is in reasoning about containers ― boxes, bottles, cups, bags, and so on ― and their contents. Containers are ubiquitous in everyday life, and children learn to use containers at a very early age.

A container can be made of a wide range of materials: rigid materials, paper, cloth, animal body parts, or combinations of these. The only requirement is that the material should maintain its shape to a sufficient degree that holes do not open up through which the contents can escape. Under some circumstances, one can even have a contained space whose bottom boundary is a liquid; for instance, an insect can be trapped in a region formed by the water in a basin and an upside-down cup. It can also have a wide range of shapes. The only constraint on the shape of a closed container is that there exists an internal cavity. The only constraint on an open container is that the material of the container plus a horizontal surface at its top delimit an internal cavity. Either a closed or an open container may additionally have small holes that penetrate from inside to outside e.g. a cage or a sieve.

The material of the contents of a container is even less constrained. In the case of a closed container, the only constraint is that the material of the contents cannot penetrate or be absorbed into the material of the container (e.g. you cannot carry water in a paper bag or carry light in a cardboard box); and that the contents cannot destroy the material of the container (you cannot keep a gorilla in a balsa wood cage). Using an open container requires additionally that the contents cannot fly out the top (Davis, 2011). Using a container with holes requires that the contents cannot fit or squeeze through the holes.

Those are all the constraints. In the case of a closed container, the material of the contents can be practically anything.

If the only goal is to infer that the object remains inside the container, then simulation is at best extremely inefficient and at worst essentially impossible. Suppose, for example, that you have an eel inside a closed fish tank and you wish to infer that it remains in the tank. If we are to solve this problem by probabilistic simulation, we would need, first to understand how an eel swims, and second to simulate all kinds of possible motions of the eel and confirm that they all end with the eel inside the tank. If we do not know the mechanisms of swimming in eels, then the only way to use probabilistic simulation is to generate some random distribution of mechanisms that eels might use. Clearly, this is not plausible.

**3.7.3 The physical dynamics of unknown entities**

Suppose that you are walking along the beach and you see an oddly shaped green object. The object has a cavity at the top, and inside the cavity is some white stuff. You want to know what will happen if you try to lift the green thing, putting one hand on either side.

Since you have no idea what the green thing or the white stuff is, you obviously cannot predict very much about this. Still you are not entirely at a loss for conclusions to draw. It is very unlikely that picking up the green thing will cause the white stuff to fly up twenty feet into the air or will cause a large sinkhole to appear under your feet. It is impossible that the green thing will turn out to be a mountain or that the white stuff will turn into a canary.

With some more observation, further conclusions can be drawn. If the green thing has sharp edges, then it is probably not soft or malleable, and therefore is likely to preserve its shape when you pick it up, unless it consists of multiple pieces, in which case you might be able to see cracks between the pieces. If the white stuff has a horizontal top surface and fills the cavity below the top surface, then it may well be a liquid. Simulation-based reasoners cannot readily draw these sorts of inferences.

**3.8 The challenge of irrelevant information**

In physical reasoning, as in most forms of reasoning, it is important to focus on the relevant aspects of the situation and ignore the irrelevant.

In modeling and simulation, irrelevancies are generally excluded manually. The human being who sets up the simulation formulates the problem in terms of the relevant features and ignores the irrelevant ones. However, this expedient will not do for a general AI program or a cognitive theory. There must be a well-defined process to exclude irrelevant information in physical reasoning, and it is hard to see how this exclusion process itself can rely on simulation.

Suppose for example that you are eating a picnic in the middle of a country fair. You are pouring lemonade from a thermos, while all around you the fair is bustling; the Ferris wheel is turning, carnival touts are barking, hot dog vendors are grilling hot dogs and so on. In most cases, in reasoning about pouring the lemonade, you can safely ignore all this other activity.

As a second example, let us return to the problem of automatically generating samples of screws and bolts discussed in section 3.7. Suppose that you have some irrelevant information about the shape of the bolt: it is a regular heptagon and the manufacturer's monogram is engraved on one side. If this is included as part of the given information λ, then the problem of generating random samples satisfying λ becomes that much harder.

**3.9 The challenge of range of scales**

Consider the following scenario. At night, you look up at the Little Dipper, and see Polaris, the North Star. However, if you close your eyes, then you don’t see Polaris.

How do we know this? Although in principle it would be possible to compute this using a physics engine, in practical terms, such an approach is wildly unwieldy. The natural way to use simulation to understand this scenario is to simulate a photon being emitted in a random direction from Polaris, and then keep simulating until you have a respectable collection of photons that meet the pupil of the observer. (A minimum of three photons are needed just to get the correct color.) The trouble derives from the scales involved; to see Polaris at its observed brightness and color because every 100 ms, about 500,000 photons from Polaris reach your pupil. Polaris is 433 light-years away, so these were emitted in 1580 AD. The probability that a particular photon emitted from Polaris will meet one particular 1 cm2 pupil is about 3\*10-42; even with high-speed computers, it is not practical to compute this by tracking individual photons over such vast scales .[[6]](#footnote-6)

Less straightforward solutions that still depend primarily on simulation raise their own complications. Suppose that we can avoid that problem of probabilistic simulation and directly generate a simulation of a photon that starts at Polaris and travels in a straight line to the eye. This involves the problem of interpolating a simulation with constraints at multiple time points, discussed in the next section. Even so, the problem of discretizing time remains. If we discretize time into equal time-steps, then to detect the interaction of the photon with the 1 mm thick eyelid would require a time step of 10-11 seconds, and therefore about 1020 time steps would be needed to span the 400 years. In all but one of these time steps, the photon moves forward 1 millimeter through empty space or the atmosphere. Clearly this is not feasible. Rather the simulation must represent only the time points when something interesting happens. One can preserve a small role for simulation, but most of the work then goes into decide what to include in the simulation; executing the simulation is providing at most a very small final piece in the puzzle.

Reasoning that integrates wide ranges of scales is common in science. Stellar evolution is related to nuclear reactions; this involves a mass ratio of about 1057. Species evolution is related to genes; this is a factor of about 1034. Everyday life also involves scale ratios that are large, though not astronomical. You are setting out for a trip to California of 3000 miles, and you turn the keys in the ignition through a distance of 1 cm; this is a ratio of 5\*108. A soldier is hit by a bullet in an event taking 10-3 seconds and carries it for the rest of his 70 year lifetime; this is a factor of 2\*1012.

**3.10 The challenge of tasks other than prediction**

Simulations are most readily used for the task of generating predictions, such as where a pendulum will be at a given instant. There are, however, many other important kinds of physical reasoning tasks, to which simulation-based techniques are in general less well suited. Some examples:

**Interpolation between two states at two different times:** Looking at Polaris is a simple example. The photon left Polaris in 1580 AD; it arrives at the eye now; one wishes to interpolate between the two.

**Planning:** Given a starting situation and a goal, find a plan that, when executed in the starting situation, will bring about the goal. For example, given a pitcher full of water and an empty glass in the starting situation and goal of having water in the glass, plan to pour the water from the pitcher into the glass. Most simulations are ill-equipped to develop plans of this sort, and Monte Carlo sampling from a vast space of possible plans seems like an unrealistic solution.[[7]](#footnote-7)

**Inferring the shape of an object from its physical behavior:** For instance, a human who observes that a bucket of water is slowly dripping from the bottom can readily infer that there is a small hole. A simulation-based reasoner can only derive such an inference by using a supervisory system that samples from a large number of possible buckets with arbitrary alterations.

**Inferring the physical properties of an object from its physical behavior:** For example, if an object is blown away by a gentle wind, a human reasoner can readily infer that it is a light object. Few simulation systems are set up to derive this sort of inference.

**Design:** Construct an object to achieve a specified functionality. Suppose, for instance, you want to design a terrarium for your pet toad that will sit on the window sill. One can immediately infer a number of constraints; it can't be completely closed, or the toad will suffocate; it can't be open on top, or the toad will jump out; it must be possible to get food and water to the toad and clean the terrarium; it needs to have a flat bottom; it needs to fit on the shelf; and so on. These constraints must be derived *before* the design and dimensions of the terrarium can be chosen or simulated.

**Comparative analysis:**  How does a modification to a problem affect the solution? For instance, you can reason that if you make an object heavier, then it will become more difficult to lift. This inference can be done without knowing the actual weight of the object (de Kleer & Brown, 1985).

**3.11 The challenge of the frame problem**

Simulation theories, might at first blush, seem to be immune to the so-called frame problem – of efficiently reasoning about the aspects of a situation that do not change over time-- that has often bedeviled knowledge-based theories (McCarthy & Hayes, 1969); in reality, the frame problem arises in simulation in very much the same ways and to the same extent as in knowledge-based theories, avoidable In simple cases (Shanahan, 1994; Reiter, 2001) daunting in many more complex cases.

An example of a case where simulation runs into the frame problem is as follows. Consider a situation where there are two stable towers of blocks on the ground with a substantial space between them, tower A and tower B. You now drop a block onto tower A. It is immediately obvious to a human onlooker that whatever happens in tower A, tower B will be unaffected. To determine this, a physics engine, such as we discuss in sections 2 and 3, would need compute all the forces between blocks in tower B, and calculate that the net force and net torque on each block is zero. Moreover, in order to accurately predict how tower A will collapse, the engine would need to use a very fine discretization of time with many time steps; and it would need repeat this identical calculation over tower B at each time step. The key hallmark of the frame problem here is the needless repetition of the calculation that nothing in tower B changes ; in particular categories of cases, it may be possible to recognize that the situation is repeated, and to retrieve the previous results rather than redo the calculation; but as the situation becomes more complicated, this rapidly becomes increasingly difficult to do.)

**3.12 The challenge of using common sense to check simulations**

As we have seen, technical problems in simulations can give results that are not merely wrong in detail, but physically nonsensical. If the wrong time increment is used in simulating solid objects, the simulation can predict that one object passes through another. If a simulation of a pendulum uses the wrong updating rule, it will predict that the pendulum swings higher and higher (section 3.2).

If there is a human in the loop, they can examine the output of the simulator and at least see that this cannot be right, particularly if the simulator outputs an animation. If there is no human in the loop, then it would be very desirable for the physical reasoning system to be able to detect these kinds of problems by itself. If the simulation is the only reasoning mechanism, this cannot be done.

**3.13 Summary**

Simulation is effective for automated physical reasoning when the task is prediction, when complete information is available, when a reasonably high quality theory is available, and when the range of scales involved, both temporal and spatial, is not extreme. If these conditions do not hold, then simulation becomes to some extent problematic. We have identified many categories of problems where these conditions do not hold, and demonstrated by examples that these categories include many natural and simple examples.

Furthermore, even when these conditions do hold, there may be simpler approaches either based on other kinds of physical knowledge or based on non-physical knowledge.

Probabilistic simulation can be effective when a plausible probability distribution can be posited, and when it is easy to generate a random sample of instances satisfying the known constraints. In many cases these conditions are not met. If the given information is too weak, as in the cases of objects of unknown shape or material, then there may be no natural probability distribution. If a complex constraint is known, as in the case of the screw and bolt that fit together, then it may be difficult to generate a random sample satisfying the constraint.

Table 2 enumerates the categories of challenges we have discussed, and reviews the examples.

|  |  |
| --- | --- |
| Challenges | Examples |
| Finding an appropriate model | Tools for cutting. |
| Discretization of time | Rigid objects passing through one another; pendulum swinging higher and higher |
| Discontinuity and instability | Dice, Stacking spheres |
| Choosing an idealization | Pendulum |
| Easy inference  | Knocking over a tower of blocks, containers, spilling a cup of coffee |
| Non-physical information | Baseball pitch |
| Incomplete information | Bolt on screw, containers, unknown entities |
| Irrelevant information | Picnic at a fair, irrelevant shape details |
| Range of scale | Astronomical examples |
| Tasks other than prediction | Planning, inferring shape and properties, design |
| Frame problem | Separated towers of blocks  |
| Checking that simulations make sense | Rigid object, pendulum |

Table 2: Challenges of simulation, with examples drawn from this paper

Some of these problems are well-discussed in the literature of modeling and simulation; others not. The problems of finding a dynamic model, of discretizing time, and of discontinuity are discussed very extensively; these are central issues in the theory of modeling and simulation, and there is a large technical literature that addresses them. The problems of choosing an idealization and of checking that simulations make sense are discussed, but comparatively rarely in technical terms. The problems of incomplete information, irrelevant information, range of scale, non-predictive tasks, and the frame problem have been examined only to a limited degree, and the technical literature within the theory of modeling and simulation is small.[[8]](#footnote-8) The issues of easy inference or of inference from non-physical information are hardly discussed. For simulation to serve as a *general* solution to the problem of physical reasoning, in either human or automated form, *all* would need to be solved. At present, no actually implemented automated system has come close.

**4 Simulation in cognitive models**

In recent years, the notion that a significant segment of human reasoning might be carried out by simulation[[9]](#footnote-9) has become increasingly popular, although advocates for the simulation theory of cognition vary widely in the claims that they make for the scope of the theory and in the breadth of their concept of simulation. For example, Battaglia, Hamrick, and Tenenbaum (2013) conjecture that most intuitive physical reasoning is carried out using probabilistic simulation:

Probabilistic approximate simulation thus offers a powerful quantitative model of how people understand the everyday physical world. This proposal is broadly consistent with other recent proposals that intuitive physical judgments can be viewed as a form of probabilistic inference over the principles of Newtonian mechanics.

Likewise, Sanborn, Mansinghka, and Griffiths (2013) write

Combining Newtonian physics with Bayesian inference, explaining apparent deviations from precise physical law by the uncertainty in inherently ambiguous sensory data, thus seems a particularly apt way to explore the foundations of people’s physical intuitions.

Simulation has also been seen as a central tool for the understanding of language, For instance, Zwaan (2004) writes (similarly to Kaup et al. (2010) whom we cited in the introduction) that

Comprehension is the vicarious experience of the described events through the integration and sequencing of traces from actual experience cued by the linguistic input.

Along similar lines, Bergen (2012) suggests (p 13, though see qualifications in his chapter 10) that perhaps "we understand language by simulating in our minds what it would be like to experience the things that language describes." (Bergen also suggests "outside of the study of language, people use simulations when they perform lots of different tasks, from remembering facts to listing properties of objects to choreographing objects”.)

Other authors have presented views in which simulation is a key component, but view simulation as one tool among many. Hegarty (2004) for example writes, "Mental simulations ... can be used in conjunction with non-imagery processes such as task decomposition and rule-based reasoning Along somewhat similar lines, Smith et al. (2013) argue that physical reasoning is generally carried out using simulation, but admit the possibility of exceptions:

In some specific scenarios participants’ behavior is not fit well by the simulation based model in a manner suggesting that in certain cases people may be using qualitative, rather than simulation-based, physical reasoning.

Importantly, what is considered a simulation also varies widely from one author to the next. In most of the literature, embodied simulations are taken, implicitly or explicitly, to be either visualizable images, static or dynamic, or activation of motor control; and most of the experimental results that are cited in support of simulation-based theories involved one of these. (As (Dennett & Viger, 1999).) note, Barsalou (1999) defines the related concept of “perceptual symbols” so broadly it is hard to see what would be excluded.)

**4.1 “Physics engine in the head”**

Advocates of a simulation model of physical reasoning such as (Battaglia, Hamrick, & Tenenbaum, 2013) generally place themselves in opposition to “intuitive physics” or “naïve physics” theories such as (Carey, 1992), (Keil, 2003), or (McCloskey, 1983) in two respects: They posit, first, that human reasoners use a physics engine as the basis for reasoning; second, that humans use scientifically correct theories such as Newtonian physics, rather than incorrect theories. It is important not to conflate these two issues; there could be a physics engine based on approximate or incorrect laws of physics, and alternative methods of reasoning could use correct physics. The distinction is particularly significant as the literature on intuitive physics is rarely very specific about the reasoning mechanism proposed. We begin, in this section, by discussing cognitive models based on the use of a physics engine; we will return to the question of intuitive physics vs. correct physics in section 6.1.

How plausible is it that human physical relies heavily on a “physics engine in the head”, in the phrase of Peter Battaglia (pers. comm.)? One line of evidence for this view comes from, Battaglia et al (2013), who carried out an experiment in which participants were asked to predict whether a tower was stable and, if not, in which direction it would fall. They found that participants' responses were consistent with a cognitive model in which a subject applies the Newtonian theory of rigid solid objects, and carries out probabilistic simulation, where the probabilistic element corresponds to uncertainty in the positions of the blocks.

In the same vein, Smith, Battaglia, and Vul (2013) carried out an experiment in which participants catching a bouncing ball in a computer animation. They similarly conclude that the subjects are using probabilistic simulations. However they add the refinement that the dynamic laws, as well as the estimate of the initial situation, are also probabilistic and that the participants' judgments are biased by prior expectations on object destinations. (Subsequent studies by the same authors, however, make significantly more limited claims, as discussed below.)

Sanborn, Mansingkha and Griffiths (2013) similarly claim that human errors in reasoning about events with colliding balls, often attributed to the use of invalid heuristics, can be accounted for using a “noisy Newton theory” in which people are in fact using the accurate Newtonian theory of collisions to reason, but get the wrong answer because of perceptual inaccuracies

Hegarty (2004) discusses a number of experiments that suggest that participants used simulation to solve a number of physical reasoning problems involving gears, pulleys, and liquids (figure 4).



Figure 4: Reproduced from (Hegarty, 2004)

In our opinion, though simulation may be a tool in human physical reasoning, there are necessarily many aspects of human reasoning for which it is insufficient..

First, as demonstrated above in Section 3, many different kinds of physical reasoning cannot or can hardly be carried out using simulation. In *all* such cases, simulation is implausible as a cognitive theory; if a problem cannot feasibly be solved using simulation, then *a fortiori* the mind cannot be using simulation to solve it. For instance, we have argued above that an automated reasoner cannot use simulation in reasoning about shaking dice inside a box or seeing Polaris; the same limitations necessarily apply to cognitive models. Likewise, one can construct problems similar to Hegarty's in which it is much less plausible that subjects will use simulation. For instance, one can ask how the behavior of a system of gears will change under *irrelevant* changes such as carving out holes from the inside of the gear; or ask whether more or less force will be needed in the pulley system if the weight is made heavier (a *comparative analysis* task with *incomplete information*), or ask them to *design* a glass that will hold less water than the one pictured. Presumably people can answer these questions easily, and as we have argued in section 3, simulation is an inappropriate tool for these.

Second, as argued in sections 3.7 and 3.8 above, there are inferences that, though they in principle could be made using simulation, can be made much more easily and robustly using either non-simulative physical principles or using non-physical information; and in many such cases it seems that the mind is in fact using these other principles. Consider for example, the experiment in Smith et al. (2013) testing whether people can predict the motion of a bouncing ball on a table with obstacles (in a computer animation). A green and a red target are marked on the table, and subjects are asked which of the two targets the ball will reach first. In some cases a simulation-based model accounts well for the experimental data. However, in cases in which the obstacles are placed so that the green target is impossible to reach, subjects will answer that the ball will reach the red target first much more quickly than the simulation model would predict (figure 5). (Note that this is an instance of reasoning about containers, discussed in section 3.8.2.)



Figure 5: Simulating a bouncing ball, from (Smith, Dechter, Tenenbaum, & Vul, 2013)

Third, there is experimental evidence that people make mistakes that are difficult to explain in a simulation-based theory.For instance, if subjects are indeed using simulation to predict collapsing towers of blocks, then they should certainly be able to use simulation for problems involving balance beams constructed of solid objects; the physical principles are the same as those involved in a collapsing tower (in fact, a subset, since neither friction nor collisions are involved) and the analysis is simpler. However, several studies e.g. (Siegel, 1976) have shown that people make systematic errors in predicting the behavior of balance beams, and we have demonstrated that such errors do not correspond to any plausible model of perceptual uncertainty (Marcus & Davis, to appear)

Fourth, there is experimental evidence that, in many situations, people cannot carry out mental simulations of moving objects accurately. Levillain and Bonatti (2011) carried out an experiment in which subject are shown a video of two balls moving toward a collision with one another, but passing behind an occluding barrier before the collision takes place. Participants are quite inaccurate in predicting when or where the collision would take place, or how the balls would move after the collision.

Smith and Vul (2013) carried out experiments in which participants are required to predict the behavior of a pendulum bob if the string is cut in mid-swing, They report that that when subjects are asked either to predict where the bob will land, or asked to choose a point to cut the string, their answers are accurate, and correspond to a simulation-based model. However, when they are asked to draw a picture of the trajectory, their answers are quite inaccurate and idiosyncratic.

Finally, accounts that propose *probabilistic* simulation – in which multiple simulations would need to occur – conspicuously lack one of the most appealing features of simple simulation as a cognitive theory; namely, its correspondence to subjective experience. People report carrying out mental simulation in solving problems, but they rarely if ever report carrying out multiple mental simulations. [[10]](#footnote-10)

It is certainly quite possible that in many of these situations, people construct *visualizations*; however, for the “physics engine in the head” theory to be correct, the inference must be the *source* of the visualization rather than *vice versa.* For example, in our example above of increasing the weight on the pulleys above, people may well visualize themselves having to pull harder to lift the weight; but they can hardly be using simulation to infer that this is the correct visualization rather than one in which they can lift the weight with less force. Similar considerations apply to non-physical visualizations as well. For instance, Moulton and Kosslyn (2009) discuss a hypothetical case of a groom who is considering telling a risqué wedding toast, visualizes the look of horror on the bride's face, and changes his mind. But how does the groom come to *visualize* his bride with a look of horror? The real work is beforehand; the cognitive process that generates the visualization must surely be drawing on a previous inference that the bride will be horrified based on knowledge of speech acts, ceremonial occasions, emotions, and so on. It is not plausible that there is a process that goes directly from the joke to the *image* of horror and then interprets the image to infer the emotion of the bride.

**4.1.1 Memory-based physical reasoning**

An alternative theory of constructing simulations is that simulations, in the sense of “movies in the head”, are derived by retrieving relevant memories and modifying them to fit the current situation. For instance if you see an open thermos of hot chicken soup knocked off a table edge, you recall other cases where you have seen containers of liquid spill and predict that the results here will be similar.

Along these lines, Sanborn, Masingkha, and Griffiths (2013) propose that their “noisy Newton” model of collisions could be implemented using stored memories of observed collisions. When a new collision must be reasoned about, the reasoner retrieves the stored memories and evaluates the new situation by a process of interpolation among the stored memories, taking into account uncertainty in the perception of the new situation.

The theory initially seems plausible, and we do not doubt that in some cases retrieval from memory plays a significant part of the reasoning process.[[11]](#footnote-11) But in many situations, the machinery of memory per se is too impoverished to yield clear answers, without significant support from other mechanisms Consider, for example, Hegarty’s (2004) gear problem. Since the problem involves this particular pair of gears, with a dent and with a bump, which the subjects have presumably never seen before, they certainly cannot retrieve a memory that gives the answer to this precise question. It even seems unlikely that subjects in general have a memory of seeing the interaction of one gear with a bump and another with a corresponding dent. Subjects presumably have seen gears interact (otherwise, they could hardly answer the question, and could retrieve memories of seeing those. But in order to apply that memory to the current problem, they must be generalize the experience: either to the narrow rule, “Two interacting gears of the same radius will rotate with the same angular velocity” or to the broader rule, “Two interacting gears rotate in such a way that the contact point moves at equal speeds along the two boundaries,” or something similar. And since they will need to carry out this generalization in any case, there is not much reason, either experimental or theoretical, to presume that the generalization is performed at query time; it would seem much more effective to carry it out when the subject learns about gears, and then have it available. Memory per se contributes little

Second, finding the correct measure of similarity between a new situation and previous memories and finding the proper way to modify an earlier memory to fit a new situation often require substantial background knowledge and reasoning. Consider again the example of knocking over a thermos of soup, and imagine a subject who has often seen a thermos of soup but, as it happens, has never seen either a thermos or soup spill, though he has seen cups of coffee, glasses of wine, glasses of juice, pots of water and so on spill. The subject must succeed in retrieving the experience of seeing other open containers of liquid spill as the relevant memory rather than the experience of seeing thermoses of soup not spill (because they remain upright). Then the subject must determine how the features of the current situation combine with those of the memories; e.g. there will be a puddle of chicken soup rather than a puddle of coffee or wine; if it falls onto fabric, it will create a grease stain rather than a brown or red stain. A glass thermos may break, like a wine glass; a metal thermos will probably not break, like a paper coffee cup. We are not arguing that these inferences cannot be carried out properly; indeed there is a substantial literature on how that might be done; but memory alone is hardly a panacea. At the very least carrying out the inferences require integrating the memory retrieval system with some knowledge-based reasoning; unaided by other mechanisms, memory would be of little use.

The issue of similarity in comparing events is studied in depth in (Lamberts, 2004). Lamberts points out that the proper measure of similarity between two events, viewed as trajectories of objects, is rather obscure; similarity judgments are, to a substantial extent, invariant under changes of both spatial and temporal scale, and dominated by qualitative characteristics such as linear vs. curvilinear motion. The larger problem however, not mentioned by Lambert, is that the relevant measure of similarity depends on the *kind* of prediction being made. In predicting the trajectory of motion, a memory of coffee spilling from a cup is a good exemplar for chicken soup spilling from a cup; in predicting the color of a stain and its response to detergent, a memory of dripping chicken soup from a spoon would be a better exemplar. Echoing points made above, the real work is left to the engine that determines what kind of memories to search for, relative to the type of prediction being made, rather than the machinery of memory per se.

Similar issues arise in using memory-based simulation for predicting of the behaviors towers of blocks, considered in (Battaglia, Hamrick, & Tenenbaum, 2013). On first glance, this might appear a natural application for this technique; one simply remembers a similar tower and predicts a similar behavior, or a number of similar towers and have them vote on the behavior. But it is easily shown that the obvious simple measure of “similarity” between towers of blocks — match up the blocks in a one-to-one correspondence, and then measure the closeness of parameters such as the coordinates of the centers of the blocks and their dimensions — will give poor results unless the subject has seen a vast number of towers[[12]](#footnote-12), and even so the theory does not yield cognitively plausible errors. First, the critical parameters here are not the coordinates or dimensions directly; they are the position of the centers of mass of each subtower relative to the boundaries of the supporting block. Second, the remembered tower that is actually most relevant to the current tower may not match up block by block; a single block in one may correspond to multiple blocks in the other. (Figure 6).



Figure 6: Tower 1 is a better exemplar for Tower 2 than for Tower 3, despite the fact that there is a one-one matching between the blocks in Towers 1 and 3, preserving support and left-right relations, whereas the matching between Tower 1 and Tower 2 is many-many (A,B match I; D,E match K,L,M; F,G,H match N).

Another difficult problem is the need to merge multiple separate memories. Consider, for example, a subject who sees an open cup of coffee perched on top of an unstable tower of blocks. To predict what will happen, it is necessary to combine previous memories of falling towers of blocks with memories of spilling containers of liquids. In a knowledge-based system, such combinations are generally fairly straightforward, though certainly difficulties can arise; in a physics engine-based system, this can easy, if tne engine already incorporates all interactions, or laborious, if it does not; but in a memory-retrieval system it is generally highly problematic.

The third issue is that many (though not all) of the problems we have described in physics engines turn up, in altered forms, in retrieval of remembered experiences as well. The problem of finding an appropriate model is substantially reduced, since the reasoner often does not have to understand a behavior, merely to have witnessed or experienced it; likewise, the problems of discretization of time and of discontinuity and instability are certainly alleviated and possibly eliminated, since the experience necessarily “solved” them correctly.

But *all* the other problems arise, or have very close analogues. The problem of finding the correct idealization is essentially the same as the problems, discussed above, of finding the correct measure of similarity and the correct way to modify a memory. The problems of incomplete information is a well-known stumbling block in nearest-neighbor algorithms and, to a lesser extent, in case-based reasoning; if the value of a feature is not known, how can that be included in the measure of similarity. The frame problem reemerges as the problem of how to merge one situation in which things change with another in which they do not (e.g. a cup of coffee placed stably at some distance from a collapsing tower of blocks). (The problems of easy inference, of incorporating non-physical information, of ignoring irrelevant information, of range of scale, and of non-predictive tasks are also all essentially unchanged.)

In the final analysis, memory may well play an important role in physical reasoning, but it cannot solve the problems on its own; even if an organism had a perfect record of all its experiences, immediately accessible at any moment, there would still be major challenges in determining which memories were relevant at any given moment, and how to generalize and adapt them to new situations.

**4.2 Implications for embodied simulation as a theory of language** .

Within the psychological literature, simulation has been taken to play a key role not only in physical reasoning, but more broadly in the representation of meaning and the interpretation of a sentence. This idea is perhaps most fully developed in Benjamin Bergen's recent book, *Louder Than Words*, (Bergen, 2012) which suggests (p 13) “While we listen to or read sentences, we simulate seeing the scenes and performing the actions that are described." Bergen adduces a great deal of evidence that seems to fit with his view, including a range of neuroimaging studies that show activity of motor cortex in linguistic understanding.

Nonetheless it follows from the arguments that we have presented here that embodied simulation cannot serve as an adequate foundation of language comprehension. Consider for example this sample text drawn from Bergen's book.

Sailors who encountered polar bears in the nineteenth century reported seeing polar bears do something quite clever to increase their chances of a hot meal [a seal]. According to these early reports, as the bear sneaks upon its prey, it sometimes covers its muzzle with its paw, which allows it to go more or less undetected. Apparently, the polar bear hides its nose.

Bergen predicts that the reader will interpret this passage by visualizing the scene. Wwhen one considers this example carefully, however, it vividly illustrates the limitations of simulation as a theory of meaning. Consider first, the logical complexity involved in the reader's understanding here. The reader must understand that

the bear believes that

if it approaches the seal with nose uncovered

the seal may see its black nose against the white background,

so the seal is likely to know that it is coming

and the seal will flee

and the bear will not be able to catch it

whereas

if it approaches the seal with nose covered

its white fur will blend into the snowy background

so the seal will be unable to see the bear

and the seal will not flee,

and the bear will be able to catch it.

In formal, logical terminology, we thus have three epistemic operators (the bear's belief and the seal's knowledge in two different conditions), a hypothetical, a couple of temporal operators, six causal connectives, two implications, three negations, and three occurrences of the modal "will be able to". Deriving this interpretation requires reasoning about the creatures' mental processes, the physics of vision, and the geometry of covering a nose with a paw, combined with background knowledge about polar bears and seals

How can these logical operators and their hierarchical structure be represented in a simulation? A number of proposals have been made. For example, Giora et al. (2004) and Kaup et al. (2007) propose that understanding a negated sentence such as “This instrument is not sharp” or by first constructing a counterfactual simulation of a sharp instrument and then a factual one of a blunt instrument. The two simulations are marked so that subsequent cognitive processes treat them correctly; for instance, if you are looking for a sharp instrument, you do not misinterpret the first simulation as veridical.

Similarly, it has been suggested (e.g. (Weger & Pratt, 2008)) that temporal relations can be represented spatially, by visualizing events in a temporal sequence as organized spatially from back to front or from left to right. This could work for relative time relations; for instance, in understanding the sentence “Joe arrived a month ago and Amy a week ago” one could visualize a timeline with a picture of Joe’s arrival, Amy’s arrival a week ago, and the current moment in sequence from left to right. But this kind of representation will hardly work for absolute temporal relations, such as distinguishing the single sentence “Joe arrived a month ago”, from “Joe arrived a week ago”, unless there is a fixed mapping from durations to distances, such as visualizing “a week ago” as a foot to the left and “a month ago” as four feet to the left. Alternatively, temporal relations can be represented by a dynamic simulation, in which one first visualizes Joe arriving and then visualizes the current moment; but again, that will not serve to distinguish “a week ago” from “a month ago”.

Each of these proposals in itself, when applied to a sentence with a concrete proposition modified by a single logical operator, seems somewhat plausible, though there are problems. However, once one considers sentences that combine operators, then the complexity of the proposed cognitive architecture becomes extremely unparsimonious. Consider the following sentences, which combine negation, quantifiers, time, and mental states:

.

* Andy did not see the bear cover its nose.
* Andy has never seen a bear cover its nose.
* No sailors have ever seen a bear cover its nose.
* Andy knows that bears do not cover their noses.
* Sid does not know that bears do not cover their noses.
* All sailors who have been to the Arctic know that bears do not cover their noses.
* Larry believes that bears used to cover their noses but now no longer cover their noses.

It is far from obvious how to clearly distinguish the different meanings of these sentences within a simulation framework wherein time is represented as left to right, mental states are represented by simulation, and negation is represented by a sequence of the counterfactual followed by the factual situation. Note that these sentences are no more (in fact, less) complex than the situation in the original story, though the complexities are more varied.

Some of Bergen's other examples involve texts whose interpretation require the same kinds of physical reasoning that we have discussed in section 3. For instance he cites an experiment of Zwaan and Taylor (2006) in which subjects read one of the following two texts:

1) The carpenter turned the screw. The boards had been connected too tightly.

2) The carpenter turned the screw. The boards had been connected too loosely.

Subjects who read sentence (1) then find it easier to turn their hand in a counterclockwise direction, while subjects who read sentence (2) find it easier to turn their hand clockwise. But the connections between the texts and the directions of turning the screw themselves rest on a rather complex sequence of inference, combining reasoning about the physics with reasoning about the carpenter's purposes.

Even in simple concrete sentences with no complex operators, there are many important distinctions of meaning that the theory of embodied simulation does not account for. Consider, for example, the two sentences "Carol is speaking Ukranian" and "Carol is speaking Czech". Bob, who speaks neither language, can visualize Carol speaking some vaguely Slavic sounding language, but he cannot visualize these two sentences differently, in terms of any difference related to the distinction in meaning. Clearly the two sentences do not mean the same thing. They do not even mean the same thing to Bob, since he draws different inferences from them. For example, if he knows that Ann speaks Ukranian but not Czech then from the first, he concludes that Ann understands what Carol is saying; from the second he concludes that Ann probably does not. If he knows that Carol grew up in Kiev, then the first is unsurprising, and the second is more surprising. Similarly, a hearer can hardly visualize the difference between "One of Joe's cousins is coming to dinner" and "One of Joe's neighbors is coming to dinner”.

It is quite possible that processes such as visualization and motor imagery participate in some way in language understanding, but the notion that meaning consists largely in the construction of genuine simulations is not plausible.

5. Non-veridical simulation

**5.1 Partially specified models, supplemented with machinery borrowed from formal logic.** .

A critic might respond to some of the points we have raised above by claiming that simulations do not have to be veridical; they can be cartoons or schemas that stand in for the world, rather than exact simulations of physical reality. SimCity doesn't need to have a one-to-one correspondence to real world financial institutions in order to be an orderly, internally coherent proxy that stands in for the real world and teaches children something about capitalism. Johnson-Laird (1983) even argues that mental models do not have to be fully specified; they can include unspecified elements. If our critique applied only to fully-specified veridical simulations, it would not prove very much.

Once one start to introduce partially specified models, however, the distinction between models and alternatives such as inference using propositional logic starts to become hazy; the more flexible the kinds of partial specification allowed in the models, the hazier the distinction becomes. Johnson-Laird, for example, allows a symbol for "not". If one additionally allows symbols for "and" and "exists" and a scoping mechanism, the system becomes equivalent to first-order logic, and the force of “simulation” can easily become significantly weakened.

**4.2 Diagrammatic reasoning** .

Non-realistic diagrams are often useful aids to reasoning. Consider, for example, figure 7, from the Wikipedia article, "Redshift", which shows a star, a light beam, and an observer. This diagram is useful, but not veridical, given the extreme distortion of relative scale.



“Redshift and blueshift”, by [Aleš Tošovský](http://commons.wikimedia.org/wiki/User%3AAles_Tosovsky) from “Redshift”, Wikipedia

Figure 7: Star, Light, Observer

The distance from the star to the observer is shown as equal to 4 wavelengths of red light; in reality, it is 4\*1024 wavelengths. Note that the wavelength is not purely an icon, like the big green arrows next to the star; the comparative wavelength of the blue and the red lights is meaningful and important to the point being illustrated. The scale of the star and the scale of the observer are likewise hugely distorted.

Similarly, figure 8, from Feynman’s *Lectures on Physics*, illustrates molecules of gas in a piston.



Figure 8: Gas molecules in a piston

In the actual situation, there are of course about 1023 molecules of gas.

Serious problems arise in distinguishing conclusions that are true in the real situation from those that are merely artifacts of the diagram. For instance, if one took figure 7 literally, one would conclude incorrectly that, since the light is shown as goes up and down on a path with a 4 inch amplitude, a 3x5 index card will sometimes fail to block one's view of an object directly ahead (even when the card is held directly in front of the eyes), and that the card would sometimes succeed in blocking it, even when the card is held 4 inches below the line of sight (figure 9). There is nothing in figure 7 to rule this out; it must be ruled out by reasoning that lies outside this picture. Similarly the student looking at figure 8 is supposed to find it helpful in understanding that the motion of the molecules exert pressure on the inside of the piston, but they are not supposed to draw the invalid conclusion that the gas molecules collide much more often with the sides of the piston than with one another, which would be true in the situation depicted. Non-veridical simulations raise difficult problems of interpretation that can often only be solved using non-simulative reasoning.



Figure 9: Invalid inferences from figure 7

6. Alternatives to simulation

What alternatives are there? One possibility is a knowledge-based reasoning engine in which both the knowledge of the domain physics and the problem specifications are largely declaratively expressed. Several different aspects of knowledge-based physical reasoning have been developed in the AI literature (see (Davis, 2008) for an overview):

* Symbolic reasoning based on a theory of the physical domain. The theory may be composed of valid qualitative rules such “An object inside a closed container remains inside.” (Hayes, 1979; 1985; Davis, 1988; 2011); or it may be an “intuitive” or “naïve” theory such as an impulse theory of motion.
* Qualitative reasoning: A collection of algorithms developed for reasoning about relative size of physical quantities and direction of change (Forbus, 1985), (Kuipers, 1986) (Weld & de Kleer, 1989). For instance, these algorithms support the qualitative inference that, if two objects at different temperatures are in thermal contact, then heat will flow from the hotter to the colder until they reach the same temperature..
* Reasoning by analogy (Gentner & Forbus, 2011). For instance the analogy between pressure and flow in hydraulic systems to voltage and current in electronic systems allows the reasoner to conceptualize electric circuits in terms of water flow, which is more familiar.
* Meta-level reasoning; reasoning explicitly about the structure and use of theories (Lehmann, Chan, & Bundy, 2012), (Forbus, 1985) (Kuipers, 1986) (Falkenhainer & Forbus, 1988) (Weld, 1990) (Nayak, 1994). For instance, one can use a rule that states that the theory of rigid solid objects will suffice to describe wooden blocks in the context of a child playing, but not in the context of architecture.

These different approaches are largely complementary rather than in opposition; they deal with different aspects of reasoning; a complete theory would presumably involve integrating these, together with further techniques for other issues that have not yet been addressed.

Qualitative reasoning has been by far the most extensively studied, and achieved some notable successes, e.g in text understanding ,(Kuehne (2004) and analogical mapping and geomeytic reasoning (Lovett et al, (2009). The Flame system (Price, Pugh, Wilson, & Snooke, 1995) for failure mode and effects analysis in automotive electrical systems has been used by multiple companies for years. Fromherz, Bobrow, and de Kleer (2003) discuss a system for automatically generating model-based control strategies for high-end printing systems**.[[13]](#footnote-13)** Othe suggestive analogues in other forms of reasoning include non-simulative planning methods such as partial-order planning (Chapman, 1987) (Weld, 1994) and hierarchical planning (Erol, Hendler, & Nau, 1994) and on symbolic reasoning for verification of computer software (Cousot, 2012), such at the software that controls complex physical systems such as airplanes (Souris & Delmas, 2007); the latter inherently combines physical reasoning with complex logical reasoning.

Highly trained physical reasoning, such as catching a fly ball or predicting the trajectory of a billiard ball with spin, probably also involves interpolation from a database of stored examples, derived from personal experience combined with using a tight perception-action feedback loop, learned using reinforcement learning or some similar strategy.

**6.1 Intuitive physics** .

A large literature, often referred to as the intuitive physics literature, demonstrates that people often make large, systematic errors in solving physical problems. For example: Siegler (1976) demonstrated that people make errors in simple balance beam problems; some subjects systematically used only the comparative weights and ignored placement whereas others used only placements and ignored weights. Cohen (2006) and many others have studied human misjudgments in scenarios involving collisions of two objects, particularly in estimating relative mass from the perceived change in velocity. McCloskey (1983) studied errors in predicting the trajectory of an object dropped by someone in motion, the trajectory of an object released after being moved on a circular track, and many similar errors. He proposes that many of these can be explained if subjects are using an impetus theory, similar to Aristotelian theory, rather than a Newtonian theory of forces and momentum. Keil (2003) studied errors in explaining the workings of such devices as helicopters. Carey (1992) demonstrated that small children do not distinguish between weight and density and have related misconceptions, such as believing that a sufficiently small piece of an object had zero mass. The list could be extended at great length.

Often, these errors have been attributed to an “intuitive”, presumably inaccurate and nonsimulative physics (McCloskey, 1983). An alternative possibility is that such errors are not in fact due to misunderstanding the physics, but better explained purely in terms of misperception. For instance, Sanborn, Masinghka, and Griffiths (2013) argue that the patterns in errors in judging collision experiments can be explained in terms of a reasoner who has a perfect understanding of the law involved, but misestimates the velocities of the objects. It has also been demonstrated, both in the “intuitive physics” literature and in the “physics engine” literature, that these errors are highly dependent both on the context and on the task. For example, both McCloskey himself (1983) and, as discussed above, Smith and Vul (2013) showed that subjects’ errors in predicting how a moving object will fall depend strongly on how those predictions are elicited and what else is happening in the physical situation.

In principle, it might be possible to develop either knowledge-based theories, or physics engines with an erroneous physics, that incorporate the kinds of misunderstanding proposed in “intuitive physics” theories. The ease with which this can be done depends on the type of error involved. It is reasonably easy to state an impetus theory, corresponding well to some of the errors that McCloskey describes, in either a knowledge-based theory or a physics engine, particularly if the physics engine is allowed to base its projection for the next state on the entire past history of the object rather than just the current state.[[14]](#footnote-14) Such a solution would be at odds with the Newtonian approaches adopted by Sanborn et al and Battaglia et al, but consistent with the spirit of using a physics simulator.

By contrast, it would be much more challenging to create either a knowledge-based system or a physics engine that is confused about the distinction between density and mass in the way that Carey (1992) described. Such a theory is almost necessarily inconsistent; indeed it is precisely the inconsistency in subjects’ responses that gives the strongest evidence for this confusion. However, the problem of how a reasoner of any kind, human or automated, can or should reason with an inconsistent theory is poorly understood, despite numerous attempts to formulate it, such as forms of paraconsistent logic (Priest, Tanaka, & Weber, 2013). While many open questions remain, the large body of “intuitive” physical errors unquestionably creates difficulties for theories based on the premises that humans are using scientifically correct theories in reasoning.

7. Conclusion: The Limits of Simulation

With infinite computational power, infinite time and unlimited memory, perhaps anything could be simulated, from the bottom up, much as Pierre Simon Laplace (Laplace, 1814) once imagined,

An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes.

But in the real world, computational power is finite, decisions must be made in finite time, and memory resources are limited. According to (Laughlin & Pines, 2000), it is effectively impossible to accurately compute an exact solution to Schrödinger's equation for systems of more than about ten particles. Neither the mind nor an automated problem solver can hope to model the interactions of everyday objects, let alone the interactions of complex agents like human beings, by simulations that originate at a quantum or even a molecular level.

Instead, real-world simulations that run in realistic time with reasonable resources must use approximations and idealization at a higher level of abstraction. In many cases, setting up the simulation ― choosing the approximations and idealizations appropriate to the problem ― and interpreting the output of the simulation ― deciding how accurate the results of the simulation are likely to be, which parts of the simulation are valid, and which parts are artifacts of the idealization ― can be much more difficult than executing the simulation. It is simply naïve to imagine that, for example, a human cognitive system could infer all that it needs to know about physics by running a Newtonian physical simulator on every entity that it encountered. With the many examples that we have reviewed in this paper, we hope to have made clear that the full-simulation view of cognition is entirely unrealistic and by the same token that simulation, however precise, is equally unlikely to solve the problems of automated commonsense physical reasoning.

At this juncture, it is difficult or impossible to quantify what fraction of the physical reasoning in human cognition or in a general artificial intelligence is or could be carried out using simulation. We have argued, however, that the range of examples we have presented in this paper ― many constructed as simple variants of problems considered in the pro-simulation literature ― suggests that there are significant limits to the use of simulation. In particular, although we have suggested that simulation is effective for physical reasoning when the task is prediction, when complete information is available, when a reasonably high quality theory is available, and when the range of spatial or temporal scale involved is moderate, in many other cases simulation is to some extent problematic. In particular, physical reasoning often involves tasks other than prediction and information is often partial.

Moreover, even when optimal conditions hold, there are many cases in which it would appear that alternative non-simulative modes of reasoning are likely to be easier, faster, or more robust. Finally, setting up and interpreting a simulation requires modes of physical reasoning that are not themselves simulation. For all these reasons, we suggest that non-simulative forms of reasoning are not an optional extra in either automated physical reasoners or cognitive theories but are centrally important.

One reason that researchers have overstated the role of simulation in cognitive models is that in the majority of studies of physical reasoning in the psychological literature, subjects are asked to carry out task of prediction and in the overwhelming majority they are presented with complete specifications of the situation.[[15]](#footnote-15) However, there is little reason to suppose that that at all reflects the frequency of these forms of reasoning in ecologically realistic settings; it may well result from artificial constraints that arise in setting up a controlled experiment.

We concur with Hegarty (2004) in seeing room for hybrid models that combine simulation with other techniques, such as knowledge-based inference,[[16]](#footnote-16) especially in the circumstances that we have enumerated here, such as circumstances in which information is incomplete. In our view, however, simulation may play a much less central role than Hegarty envisioned In particular, in a general intelligence, human or artificial, practically any reasoning that involves simulation will probably involve some degree of non-simulative reasoning, to set up the simulation and to check that the answer is broadly reasonable. Conversely, simulation can often be used to check the results of non-simulative reasoning; if you have inferred such and such but cannot actually imagine it, that is probably a sign of trouble. Outside of carefully constructed laboratory situations, and certain forms of specialized expert reasoning, we think it is relatively rare that simulation is used in isolation in human cognitive reasoning.

In the final analysis, it seems plausible that simulation plays some role in human cognitive systems, and there can be no doubt that it finds uses in automated reasoning. But if our arguments are correct, the scope of simulation in both cognition and automated reasoning is far more restricted than generally thought. AI researchers working in common sense and automated reasoning might do well to study what alternative mechanisms human beings use, beyond simulation, in order to achieve a high degree of success in everyday reasoning about the world.

**Acknowledgements**

Thanks to Peter Battaglia, Benjamin Bergen, Thomas Bräunl, Alan Bundy, Jonathan Goodman, Philip Johnson-Laird, Casey McGinley, Andrew Sundstrom, and Ed Vul for helpful discussions; and to Nils Thuerey for information about simulation for film CGI.

# Bibliography

Barsalou, L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*(4), 577-660. From http://journals.cambridge.org/download.php?file=%2FBBS%2FBBS22\_04%2FS0140525X9900214Xa.pdf&code=f67f9de0fca9e317f111c69b82140018

Battaglia, P., Hamrick, J., & Tenenbaum, J. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences, 110*(45), 18327-18332.

Benger, W. (2008). Colliding galaxies, rotating neutron stars, and merging black holes -- visualizing high dimensional datasets on arbitrary meshes. *New Journal of Physics, 10*. doi:http://dx.doi.org/10.1088/1367-2630/10/12/125004

Bergen, B. (2012). *Louder than Words: The New Science of How the Mind Makes Meaning.* New York: Basic Books.

Boeing, A., & Bräunl, T. (2007). Evaluation of real-time simulation systems. *GRAPHITE.* doi:http://dx.doi.org/10.1145/1321261.1321312

Boeing, A., & Bräunl, T. (2012). Leveraging Multiple Simulators for Crossing the Reality Gap. *12th International Conference on Control, Automation, Robotics, and Vision.* Guangzhou, China. From http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=06485313

Bridson, R. (2008). *Fluid Simulation for Computer Graphics.* Wellesley, Mass.: A.K. Peters.

Carey, S. (1992). The origin and evolution of everyday concepts. *Cognitive Models of Science, 15*, 89-128.

Chapman, D. (1987). Planning for conjunctive goals. *Artificial Intelligence, 32*(3), 333-379. doi:http://dx.doi.org/10.1016/0004-3702(87)90092-0

Cohen, A. (2006). Contributions of invariants, heuristics, and exemplars to the visual perception of relative mass. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 574-598.

Cousot, P. (2012). Formal verification by abstract interpretation. In A. Goodloe, & S. Person (Ed.), *4th NASA Formal Methods Symposium.* *7211*, pp. 3-7. Norfolk, Virginia: Springer, Lecture Notes in Computer Science. From http://www.di.ens.fr/~cousot/publications.www/Cousot-NFM2012.pdf

Davis, E. (1988). A logical framework for commonsense predictions of solid object behavior. *AI in Engineering, 3*(3), 125-140. doi:http://dx.doi.org/10.1016/0954-1810(88)90029-5

Davis, E. (1998). The naive physics perplex. *AI Magazine, 19*(4), 51-79. From http://www.aaai.org/ojs/index.php/aimagazine/article/view/1424/1323

Davis, E. (2008). Physical Reasoning. In F. van Harmelen, V. Lifschitz, & B. Porter, *Handbook of Knowledge Representation* (pp. 597-620). Elsevier.

Davis, E. (2011). How does a box work? *Artificial Intelligence, 175*(1), 299-345. doi:http://dx.doi.org/10.1016/j.artint.2010.04.006

Davis, E. (2013). Qualitative spatial reasoning in interpreting text and narrative. *Spatial Cognition and Compujtation*. From http://www.cs.nyu.edu/faculty/davise/papers/cosit.pdf

de Kleer, J., & Brown, J. (1985). A qualitative physics based on confluences. In D. Bobrow, *Qualitative Reasoning about Physical Systems* (pp. 7-84). Cambridge: MIT Press. doi:http://dx.doi.org/10.1016/0004-3702(84)90037-7

Dennett, D., & Viger, C. (1999). Sort-of symbols? *Behavioral and Brain Sciences, 22*(4), 613. doi:http://dx.doi.org/10.1017/S0140525X99262140

Erol, K., Hendler, J., & Nau, D. (1994). UMCP: A sound and complete procedure for hierarchical task-network planning. *Proceedings, International Conference on AI Planning Systems*, (pp. 249-254).

Falkenhainer, B., & Forbus, K. (1988). Setting up large-scale qualitative models. *Procceedings of the National Conference on Artificial Intelligence (AAAI-88).* From http://www.aaai.org/Papers/AAAI/1988/AAAI88-054.pdf

Forbus, K. (1985). Qualitative process theory. In D. Bobrow, *Qualitative Reasoning about Physical Systems* (pp. 85-168). Cambridge, Mass.: MIT Press. doi:http://dx.doi.org/10.1016/0004-3702(84)90038-9

Fromherz, M., Bobrow, D., & de Kleer, J. (2003). Model-based Computing for Design and Control of Reconfigurable Systems. *AI Magazine, 24*(4), 120.

Gentner, D., & Forbus, K. (2011). Computational models of analogy. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(3), 266-276. doi:10.1002/wcs.105

Giora, R., Balaban, N., Fein, O., & Alkabets, I. (2004). Negation as positivity in disguise. In H. Colston, & A. Katz, *Figurative Language Comprehension: Social and Cultural Influences* (pp. 233-258). Maswah, N.J.: Erlbaum.

Goldman, A. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading.* Oxford: Oxford University Press.

Hairer, E., Lubich, C., & Wanner, B. (2006). *Geometric numerical integration structure-preserving algorithms for ordinary differential equations.* New York: Springer.

Hamrick, J., Battaglia, P., & Tenenbaum, J. (2011). Internal physics models guide probabilistic judgment about object dynamics. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society.* From http://www.academia.edu/1884257/Internal\_physics\_models\_guide\_probabilistic\_judgments\_about\_object\_dynamics

Hayes, P. (1979). The naive physics manifesto. In D. Michie, *Expert Systems in the Microelectronic Age.* Edinburgh: Edinburgh University Press. From http://aitopics.org/publication/naive-physics-manifesto

Hayes, P. (1985). Naive physics 1: Ontology for liquids. In J. Hobbs, & R. Moore, *Formal Theories of the Commonsense World* (pp. 71-107). Norwood, NJ: Ablex. From http://www.issco.unige.ch/working-papers/Hayes-1978-35.pdf

Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Science, 8*(6), 280-285. doi:http://dx.doi.org/10.1016/j.tics.2004.04.001

Johnson-Laird, P. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness.* Cambridge, Mass.: Harvard University Press.

Johnston, B. (2009). *Practical Artificial Intelligence.* Ph.D. Thesis, University of Technology, Sydney.

Johnston, B., & Williams, M. (2007). A generic framework for approximate simulation in commonsense reasoning systems. *International Symposium on Logical Formalizations of Commonsense Reasoning.* From http://epress.lib.uts.edu.au/research/bitstream/handle/10453/2532/2007001119.pdf

Kapitaniak, M., Strzalko, J., Grabski, J., & Kapitaniak, T. (2012). The three-dimensional dynamics of the die throw. *Chaos, 22*(4). doi:doi: 10.1063/1.4746038.

Kaufman, D., Vouga, E., Tamstorf, R., & Grinspun, E. (2012). Reflections on simultaneous impact. *SIGGRAPH.* doi:http://dx.doi.org/10.1145/2185520.2185602

Kaufmann, H., & Meyer, B. (2008). SImulating educational physical experiments in augmented reality. *SIGGRAPH Asia.* doi:http://dx.doi.org/10.1145/1507713.1507717

Kaup, B., Lüdtke, J., & Maienborn, C. (2010). "The drawer is still closed": Simulating past and future actions when processing sentences that describe a state. *Brain and Language, 112*(3), 159-166. doi:http://dx.doi.org/10.1016/j.bandl.2009.08.009

Kaup, B., Yaxley, R., Madden, C., & Zwaan, R. L. (2007). Experiential simulations of negated text information. *Quarterly Journal of Experimental Psychology, 60*(7), 976-990. From http://www.tandfonline.com/doi/full/10.1080/17470210600823512

Keil, F. (2003). Folkscience: Coarse interpretations of a complex reality. *Trends in Cognitive Science, 7*(8), 368-373.

Kuehne, S. (2004). *Understanding natural language description of physical phenomena.* Evanston, I.: Northwestern University.

Kuipers, B. (1986). Qualitative simulation. *Artificial Intelligence, 29*(3), 289-338. doi:http://dx.doi.org/10.1016/0004-3702(86)90073-1

Lamberts, K. (2004). An Exemplar Model for Perceptual Categorization of Events. *The Psychology of Learning and Motivation, 44*, 227-260.

Laplace, P. (1814). *A Philosophical Essay on Probabilities.*

Laughlin, R., & Pines, D. (2000). The theory of everything. *Proceedings of the National Academy of Sciences of the United States of America, 97*(1), 28-31. From http://www.jstor.org/stable/121737

Lehmann, J., Chan, M., & Bundy, A. (2012). A higher-order approach to ontology evolution in physics. *Journal of Data Semantics*, 1-25. From http://link.springer.com/article/10.1007%2Fs13740-012-0016-7#

Levillain, F., & Bonatti, L. (2011). A dissociation between judged causality and imagined location in simple dynamic scenes. *Psychological Science, 22*(5), 674-681. doi:doi:10.1177/0956797611404086

Lovett, A., Tomei, E., Forbus, K., & Usher, J. (2009). Solving Geometric Analogy Problems through Two-Stage Analogical Mapping. *Cognitive Science, 33*(7), 1192-1231.

Marcus, G., & Davis, E. (to appear). How robust are probabilistic models of higher-level cognition? *Psychological Science*.

Markman, K., Klein, W., & Suhr, J. (2009). *Handbook of Imagination and Mental Simulation.* New York: Psychology Press.

McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the perspective of artificial inteligence. In B. Meltzer, & D. Michie, *Machine Intelligence 4* (pp. 463-502). Edinburgh: Edinburgh U. Press.

McCloskey, M. (1983). Intuitive Physics. *Scientific American, 284*(4), 122-130.

Mombauri, K., & Berns, K. (2013). *Modeling, simulation and optimization of bipedal walking.* Springer.

Moulton, S., & Kosslyn, S. (2009). Imagining predictions: Mental imagery as mental emulation. *Philosophical Transactions of the Royal Society (B): Biological Sciences, 364*, 1273-1280. doi:http://dx.doi.org/10.1093/acprof:oso/9780195395518.003.0040

Murman, S., Chan, W., Aftosmis, M., & Meakin, R. (2003). An interface for specifying rigid-body motions for CFD applications. *41st AIAA Aerospace Sciences Meeting.* From http://people.nas.nasa.gov/~wchan/publications/aiaa2003-1237.pdf

Murphy, T. (2013). The First Level of Super Mario Bros. is Easy with Lexicographic Ordering and Time Travel ... after that it gets a little tricky. *SIGBOVIK.* From http://www.cs.cmu.edu/~tom7/mario/mario.pdf

Nayak, P. (1994). Causal Approximations. *Artificial Intelligence, 70*(1-2), 277-334. doi:http://dx.doi.org/10.1016/0004-3702(94)90108-2

Nichols, S., & Stich, S. (2003). *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds.* Oxford: Clarendon.

Nichols, S., Stich, S., Leslie, A., & Klein, D. (1996). Varieties of offline simulation. In P. Carruthers, & P. Smith, *Theories of Theories of Mind.* Cambridge: Cambridge University Press. From http://cogprints.org/376/

Nyga, D., & Beetz, M. (2012). Everything robots always wanted to know about housework (but were afraid to ask). *IEEE/RSJ International Conference on Intelligent Robots and Systems.* From http://ias.cs.tum.edu/\_media/spezial/bib/nyga12actioncore.pdf

Price, C., Pugh, D., Wilson, M., & Snooke, N. (1995). The FLAME System: Automating electrical failure mode and effects analysis (FEMA). *IEEE Reliability and Maintainability Symposium*, (pp. 90-95).

Priest, G., Tanaka, K., & Weber, Z. (2013). *Paraconsistent logic.* (E. Zalta, Ed.) Retrieved from Stanford Encyclopedia of Philosophy: http://plato.stanford.edu/archives/fall2013/entries/logic-paraconsistent/

Rahimian, A., Lashuk, I., Veerapaneni, S., Chandramowlishwaran, A., Malhotra, D., Moon, L., . . . Biros, G. (2010). Petascale Direct Numerical Simulation of Blood Flow on 200K Cores and Heterogeneous Architectures. *Supercomputing*, (pp. 1-11). doi:http://dx.doi.org/10.1109/SC.2010.42

Sanborn, A., Mansinghka, V., & Griffiths, T. (2013). Reconciling Intuitive Physics and Newtonian Mechanics for Colliding Objects. *Psychological Review, 120*(2), 411-437.

Saxe, R. (2005). Against simulation: the argument from error. *Trends in Cognitive Sciences, 9*(4), 174-179. doi:http://dx.doi.org/10.1016/j.tics.2005.01.012

Siegel, R. (1976). Three aspects of cognitive development. *Cognitive Psychology, 8*(4), 481-520. doi:http://dx.doi.org/10.1016/0010-0285(76)90016-5

Smith, K., & Vul, E. (2013). Sources of uncertainty in intuitive physics. *Topics in Cognitive Science, 5*, 185-199. doi:10.1111/tops.12009

Smith, K., Battaglia, P., & Vul, E. (2013). Consistent physics underlying ballistic motion prediction. *Proceedings of the 35th Annual Meeting of the Cognitive Science Society.* From http://www.academia.edu/3689250/Consistent\_physics\_underlying\_ballistic\_motion\_prediction

Smith, K., Dechter, E., Tenenbaum, J., & Vul, E. (2013). Physical predictions over time. *Proceedings of the 35th Annual Meeting of the Cognitive Science Society.* From http://www.academia.edu/3689251/Physical\_predictions\_over\_time

Souris, J., & Delmas, D. (2007). Experimental assessment of Astree on safety-critical avionics software. In F. Saglietti, & N. Oster (Ed.), *Computer Safety, Reliability, and Security, 26th Intl. Conf.* *4680*, pp. 479-490. Nuremberg, Germany: Springer, Lecture Notes in Computer Science.

Stewart, D. (2000). Rigid-body dynamics with friction and impact. *SIAM Review, 42*(1), 3-39. From http://www.jstor.org/stable/2653374

Tabiei, A., & Nilakantan, G. (undated). *Reduction of acceleration induced injuries from mine blasts under infantry vehicles.* Cincinatti: Dept. of Aerospace Engineering and Engineering Mechanics, University of Cincinatti. From http://www.ase.uc.edu/~atabiei/pdfs/J.pdf

Torralba, A., Fergus, R., & Freeman, W. (2008). 80 Million Tiny Images: A Large Dataset for Non-Parametric Object and Scene Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 30*(11), 1958-1970.

Vondrak, M., Sigal, L., & Jenkins, O. (2008). Physical simulation for probabilistic motion tracking. *Computer Vision and Pattern Recognition.* From http://robotics.cs.brown.edu/projects/dynamical\_tracking/

Weger, U., & Pratt, J. (2008). Time flies like an arrow: space-time compatibility effects suggest the use of a mental timeline. *Psychonomic Bulletin and Review, 15*(2), 426-430. From http://link.springer.com/article/10.3758/PBR.15.2.426

Weld, D. (1990). Approximation Reformulations. *Processings of the National Conference on Artificial Intelligence (AAAI-90).* From http://aaaipress.org/Papers/AAAI/1990/AAAI90-062.pdf

Weld, D. (1994). An introduction to least-commitment planning. *AI Magazine, 15*(4), 27-61. From http://www.aaai.org/ojs/index.php/aimagazine/article/view/1109

Weld, D., & de Kleer, J. (1989). *Qualitative Reasoning about Physical Systems.* San Mateo, Calif.: Morgan Kaufmann.

Zicker, S., & Veloso, M. (2009). Tactics-based behavioural planning for goal-driven rigid body control. *Computer Graphics Forum*. From http://www.cs.cmu.edu/~mmv/papers/09cgf-stefan.pdf

Zwaan, R. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. Ross, *The Psychology of Learning and Motivation* (Vol. 43, pp. 35-62). New York: Academic Press. From http://www.academia.edu/749405/The\_immersed\_experiencer\_Toward\_an\_embodied\_theory\_of\_language\_comprehension

Zwaan, R., & Taylor, L. (2006). Seeing, acting, understanding: motor resonance in language comprehension. *Journal of Experimental Psychology, 135*, 1-11. From http://ureca.recherche.univ-lille3.fr/uploads/File/CNCC/JEPGen\_06.pdf

1. Many can be found on YouTube by searching on the name of the engine plus "Bug"; e.g. https://www.youtube.com/watch?v=Jl-A5YaiWi8 shows a bug with gyroscopic motion in ODE and https://www.youtube.com/watch?v=vfl33Tn0pYc shows a number of bugs in Skate 3. [↑](#footnote-ref-1)
2. Specifically, the theory of collisions is problematic (Kaufman, Vouga, Tamstorf, & Grinspun, 2012) (Stewart, 2000). [↑](#footnote-ref-2)
3. No possible representation can separately represent each instant of time in a continuous model. It is possible, nonetheless, for an automated reasoner to use representations to refer to a continuous model and axioms or inference rules that are valid over a continuous model; and these are qualitatively different from representations, axioms, and rules based on discrete models. For instance, a differential equation refers to a continuous model of time whereas a difference equation refers to a discrete model of time. [↑](#footnote-ref-3)
4. There has been work in physics on using discrete models of time. However, since these require a time quantum of about 10-43 seconds, these are irrelevant to the discretization of time used in simulations. [↑](#footnote-ref-4)
5. It is not known to be decidable for rotating rigid objects. [↑](#footnote-ref-5)
6. The probability of a successful simulation is much higher if we go backward in time from the eye to Polaris, along the lines of the "ray-tracing" algorithms of computer graphics. If we use the maximum angular resolution of the human retina in the dark, and trace a line backward, the probability that the line will intersect Polaris is about 4\*10-11. (Thanks to Rob Fergus and Yann LeCun for their assistance.) [↑](#footnote-ref-6)
7. A salient counterexample is the program described in (Zicker & Veloso, 2009) which constructs plans for playing soccer and miniature golf by a heuristic search through the space of actions, using simulation to predict the behavior of the ball when kicked or putted. However, the vocabulary of actions and of physical behaviors is small and high precision is required. It is not likely to yield a generally applicable solution to the problem of planning. [↑](#footnote-ref-7)
8. There is, of course, a large AI literature on incomplete information, irrelevant information, and the frame problem, but these challenges have largely left unaddressed in the modeling and simulation literature [↑](#footnote-ref-8)
9. In the literature of cognitive psychology the word “simulation” is used very loosely with many different meanings. Our discussion here is limited to two well-known theories. [↑](#footnote-ref-9)
10. We are not of course proposing that cognitive processes are open to introspection; we are pointing out the difference between single simulations, which are confirmed by introspective experience, and probabilistic simulations, which are not. [↑](#footnote-ref-10)
11. The machine learning techniques of “nearest-neighbors” and “case-based reasoning” are both implementations of this general idea. [↑](#footnote-ref-11)
12. With a large enough data set , nearest neighbors can be surprisingly effective, even for tasks as complex as image categorization (Torralba, Fergus, & Freeman, 2008). [↑](#footnote-ref-12)
13. Admittedly, rule-based systems as applied to physical reasoning have had only limited success in either automated commonsense reasoning or cognitive modeling and in particular few successful practical applications; possible reasons for this are discussed in (Davis, 1998). [↑](#footnote-ref-13)
14. The technical term for this is “hysteresis”; there are actually legitimate physical theories that work this way. [↑](#footnote-ref-14)
15. There is also a tendency in the cognitive science literature to focus on idealized physical models that are characteristic of textbooks or engineered devices, such as balance beams, perfectly bouncing balls, and gears rather than those that occur in natural settings like camping trips or kitchens. [↑](#footnote-ref-15)
16. The COMIRIT program combines tableau-based logical reasoning with simulation-based physical reasoning (Johnston, 2009); however, all of the system’s physical knowledge is in the simulator, so this is integration of a rather limited kind. [↑](#footnote-ref-16)