A Theoretical Analysis of Thought Experiments

David L. Hibler
Department of Computer Science
Christopher Newport University
Newport News, VA 23606.
tele: (804)-594-7065
e-mail: dhibler@pcs.cnu.edu

Abstract

The thought experiment methodology is a recently developed technique for qualitative reasoning. Thought experiments involve four main steps: (i) simplification of the original problem, (ii) solution of the simplified problem, (iii) conjecture the result of the original problem based on the solution of the simplified version, and finally (iv) verification that the conjecture is correct. This methodology has been implemented and has been applied successfully to problems involving static electricity, the motion of charged pendulums, fluid flow, and thermodynamic cycles. The purpose of this paper is to analyze this technique using probabilistic models

Introduction

Imaginary, simplified situations are often analyzed by human problem solvers in order to understand the principles behind more realistic situations. In physics, this technique is sometimes referred to as a thought experiment[Prigogine & Stengers 1984].

As an example of a thought experiment, consider how to solve the following problem.

> *A certain amount of charge is placed on a conductor. What ultimately happens to the charge? Where does it go?*

There are sophisticated ways to solve this problem, but a beginning physics student can often determine the solution in the following way.

1. We know that even a small amount of charge in the macroscopic sense involves a huge number of particles. Despite this, we **simplify** the problem to two charged particles.
2. Using the fact that like charges repel and other elementary knowledge, we **solve** the problem for two charged particles. We conclude that the particles separate until they reach the surface of the conductor. If charge cannot leave the conductor (no sparks) the answer to the simplified problem is that the charge moves to the surface and stays there.
3. We **conjecture** that the solution to the simplified problem is essentially the same as the solution to the original problem.
4. The previous step produces a result but we are, perhaps, uncomfortable with it. Did we oversimplify? Just to check, we ask ourselves if it would make any difference if we had three particles or four. Solving these related problems we discover that the result was the same. With this amount of **verification** we feel that the result very probably applies even to a large number of particles.

The above example is an informal thought experiment. It involved the following four steps: *simplification*, *solution*, *conjecture*, and *verification*. Since the verification method employed in this case was not rigorous it is termed *heuristic* verification.

We have formalized this method and use it for qualitative physics problem solving. It has been implemented in Prolog in a system called TEPS[Hibler 1992; Hibler & Biswas 1992b; Hibler & Biswas 1989]. TEPS stands for *Thought Experiment Problem Solver*.

In this paper we analyze the operation of a thought experiment problem solver. We will construct some elementary probabilistic models of the thought experiment problem solving process and determine some of its characteristics.

First, we describe the thought experiment methodology in more detail.

Next, we consider a thought experiment for which there is one right answer and n possible wrong answers. The thought experiment is verified using heuristic verification. which means that a total of s different simulations agree. We analyze the probability, $P[C \mid A]$, that the answer is correct given that all s simulations agree. This might be termed the validity of the heuristic verification. This analysis allows us to provide a condition for heuristic verification to work, to discuss the expected behavior of heuristic verification, and to consider the effect of fine versus coarse descriptions.

Finally, we analyze thought experiments in terms of the expected time, T_E, for achieving a verified result. The important quantity in this analysis is the effec-

tiveness, E_i, of each individual thought experiment. A theorem is proved which indicates the order in which thought experiments should be performed based on effectiveness. Estimation of effectiveness is discussed.

Thought Experiments

The purpose of this section is to describe the steps in the thought experiment methodology in a brief, straightforward way. The thought experiment methodology has been described in a more formal, mathematical manner elsewhere[Hibler & Biswas 1989].

The thought experiment problem solver (TEPS) which we have implemented uses a qualitative reasoning system based on Forbus Qualitative Process Theory [Forbus 1984; Hibler & Biswas 1992a]; however, thought experiment methodology is very flexible. The purpose of this paper is to analyze this methodology and not any particular implementation. We will mention some other possibilities for implementation as we discuss the steps of the thought experiment process.

Simplification

The first step in a thought experiment involves simplification. Simplifications are problem transformation rules provided to the problem solver by the system designer. Some simplifications rules are domain specific; others are domain independent. Practical simplification strategies and a partial catalog of useful simplifications are presented in [Hibler 1992].

As an example of a simplification, the thought experiment mentioned in the introduction used *Population Reduction*. The Population Reduction simplification maps any problem specification with multiple identical objects to a problem specification with only two objects. This state is simpler because there are fewer variables involved, and it is computationally much easier to determine the time evolution of the behavior of the system.

Another example of a simplification is the *Simple Stereotype* method. Simple stereotypes are easily built into modules describing objects or processes. They represent simpler versions of the object or process. Other examples of simplifications which have been implemented in TEPS include *Monte Carlo, Combined Change, Variable Blocking* and *Superposition*. Monte Carlo simplification is based on randomly sampling the state graph. Combined Change samples the state graph based on the way variables change. Variable blocking is a method for ignoring variables thought to be irrelevant. Superposition combines results from separate subproblems but has been implemented only for special cases. Some other possible simplifications which have not been implemented are discussed in [Hibler 1992].

Many abstraction techniques developed by others could be considered as simplifications in our sense. If heuristic verification is used with these techniques they might be applied in cases where the validity of the abstraction is questionable. Examples include Ontological Perspective[Falkenhainer & Forbus 1990], Structural Consolidation[Weld & Addanki; Falkenhainer & Forbus 1990], Temporal Abstraction[Kuipers 1987], and Aggregation[Weld 1986]. The Exaggeration method of Weld[Weld 1988] can definitely be considered a simplification in our sense.

Solution

The next step in a thought experiment involves "solving" the simplified model. This means that the problem solver must contain a reasoning engine which takes a problem specification and reasons about it to produce some "results".

The TEPS Reasoning Engine TEPS takes as input a specification of a qualitative state of a physical system. A qualitative state is a collection of qualitative values, one for each of the variables pertaining to the system. Qualitative values consist either of special, qualitatively significant landmark values or of intervals between two adjacent landmark values[Forbus 1984].

The dynamical behavior of the system is described by direct or indirect influences which can cause changes of qualitative state. They correspond roughly to qualitative versions of differential equations and qualitative versions of functional relationships.

Given a problem specification in TEPS we can simulate the time evolution of the system by generating a graph of qualitative states which can be reached from the original state. This is known as a reachable envisionment.

Other Reasoning Engines Other reasoning engines could be used as the basis for a thought experiment problem solver. Examples include not only other qualitative reasoning systems such as the of Kuipers [Kuipers, 1986] or de Kleer and Brown[de Kleer & Brown 1984], but also numerical simulation systems.

For a numerical reasoning system the input would consist of the initial state and the dynamical equations. The result would consist of the entire numerical simulation of the system's trajectory in phase space.

Description of Results The thought experiment problem solver is designed to answer specific equations about a physical system given an initial state for that system. We are thus not concerned about the output of the reasoning system directly because it usually does not constitute an answer to a question. What does constitute a possible answer is specified by some description function, D. D is sometimes called a description basis. We will assume that any query only has a finite number of possible answers. D can be thought of as a function which classifies the results produced by the reasoning engine into one of a finite number of categories, one for each possible answer. Thus even if our reasoning engine uses numerical simulation the description of the results is qualitative.

In the example involving charge placed in a conductor the description function takes the output of the problem solver and categorizes it by giving a list of regions which contain a nonzero amount of charge in the final state.

TEPS contains a library of description functions; however, a description function can also be input for a particular problem.

Conjecture

A conjecture is a guess about the description of the result of the original problem based on the solution obtained on the simplified version of the problem. We will assume in this paper that the conjecture is always that the description function classifies the result of the simplified problem the same way that it would have classified the result of the original problem. In other words, the same description is produced.

Verification

Verification can be rigorous or heuristic. It could even be empirical. Rigorous verification requires establishing a formal proof that the conjecture is true. This is usually difficult. Empirical verification consists of comparing the predictions with what actually occurs in the real world. Often, this approach is not practical. The most common type of verification is heuristic. With this type of verification other simplifications are tried, and the resulting conjectures are compared with the original conjecture. If they agree, we accept the conjecture as a reasonable belief.

The Validity of Thought Experiments

The first concern of our analysis is the validity of thought experiments. Do they, in fact, give the correct answer? Our certainty in any particular case will depend on the verification method used. If exact verification is available then we can be sure whether the result is correct. More has been said about exact verification elsewhere[Hibler 1992]. Many uses of thought experiments involve heuristic verification. Heuristic verification seems intuitively reasonable; however, we need to clarify some of the ideas behind it. In order to do this, we explore a probabilistic model of heuristic verification, and determine the probability that a thought experiment is correct given that it is heuristically verified. From this we determine the actual requirement for heuristic verification to be useful. We also examine the effect of independent versus dependent confirmation and of fine versus coarse descriptions on validity.

Probabilistic Models

To create a simple model of the thought experiment process we make several assumptions.

A probabilistic model requires the definition of a problem space. Let **S** be the set of all problems which the problem solver with its particular library of processes will accept. We will assume that we sample the problems using some fixed probability distribution. The probability distribution we use is left implicit and is not specified in the notation. The key assumption is that the probability distribution is fixed. Based on this sampling, we can discuss the relative frequency with which events which are functions of specific problems occur and associate probabilities with these events. In a like manner we can define probabilities on any subset of **S**.

Consider any simplification method together with its associated conjecture and verification methods. When these methods are applied to a randomly selected problem from the problem space we obtain a thought experiment which we can characterize probabilistically.

A key assumption states that solving a problem consists of performing separate, independent thought experiments until one is verified. This *independent thought experiment assumption* implies that we are ignoring the use of complementary models and inheritance in this analysis. Since the use of additional information by inheritance in general helps the performance of the problem solver it is safe to say that the simple analysis provides a conservative estimate of complexity of the problem solving task, and, in most cases, the actual results will be better.

Correctness Probability

Assume that the thought experiment process uses s simulations whose results are compared. Our description of the results classifies those results into $n+1$ possible categories. One of those categories is correct. In other words, if we had applied the classification to a full simulation of the original problem the result would have been in that category. The other n categories are wrong. We use small letters c and $w_1 \ldots w_n$ to denote the correct answer and the n wrong answers. Let C denote the event that all s simulations have the correct answer. Let A denote the event that all s simulations have the same answer. Let W denote the event that all s simulations have a wrong answer. We indicate intersections of the above events by writing the letters together.

The thought experiment is heuristically verified if all s simulations have the same answer. Thus the probability of interest is $P[C \mid A]$, the probability that the answer is correct given that all simulations agree. We know that

$$P[C \mid A] = P[C]/P[A]$$

If the simulations agree they must obviously agree and be correct or agree and be wrong. These possibilities are disjoint. Thus $P[A] = P[AC] + P[AW]$; but $P[AC] = P[C]$ so we can rewrite $P[C \mid A]$ as

$$P[C|A] = 1/(1+R) \tag{1}$$

where

$$R = P[AW]/P[C]. \tag{2}$$

The smaller R is, the better the accuracy of our thought experiment; the larger R is, the worse the accuracy.

To obtain more insight we analyze the components of R. Let a be any answer, either the correct answer c or any of the $w_1 \ldots w_n$ wrong answers. The probability of getting answer a in all s simulations is

$$P_1[a \mid (0)a]\, P_2[a \mid (1)a]\, P_3[a \mid (2)a] \ldots P_s[a \mid (s-1)a]$$

The notation $P_i[a \mid (j)a]$ denotes the probability that we obtain result a on simulation i, having had (j) a's on all the previous j simulations. We call this the probability for a confirmation of answer a. $P[a \mid (0)a]$ is just $P[a]$. Using this formula we obtain

$$P[C] = P_1[c \mid (0)c] \ldots P_s[c \mid (s-1)c] \tag{3}$$

$$\begin{aligned} P[AW] \;=\; & P_1[w_1 \mid (0)w_1] \ldots P_s[w_1 \mid (s-1)w_1] \\ & + P_1[w_2 \mid (0)w_2] \ldots P_s[w_2 \mid (s-1)w_2] \\ & \vdots \\ & + P_1[w_n \mid (0)w_n] \ldots P_s[w_n \mid (s-1)w_n]. \end{aligned} \tag{4}$$

Using equations 3 and 4 we can rewrite equation 2 as a sum of products of ratios of probabilities. It is convenient to write 2 as

$$R = r_1^s + r_2^s + \ldots + r_n^s \tag{5}$$

where r_k is the geometric mean of $P_i[w_k \mid (i-1)w_k]/P_i[c \mid (i-1)c]$ over the different simulations. i.e.

$$\begin{aligned} r_k \;=\; & ((P_1[w_k \mid (0)w_k]/P_1[c \mid (0)c]) \ldots \\ & (P_s[w_k \mid (s-1)w_k]/P_s[c|(s-1)c]))^{1/s} \end{aligned} \tag{6}$$

We must be careful to note that r_k depends on s.

Next, let us assume $r \geq r_k$ so there exists an upper bound on the r_k ratios for all s. In that case $R \leq nr^s$ so equation 1 becomes

$$P[C \mid A] \geq 1/(1 + nr^s) \tag{7}$$

This equation is sufficiently important that we reiterate what the quantities mean in words. A thought experiment is performed for which there is 1 right answer and n possible wrong answers. The thought experiment is verified using heuristic verification. which means that a total of s different simulations agree. $P[C \mid A]$ is the probability that the answer is correct given that all s simulations agree. r is a type of bound on the probability of obtaining any single wrong answer versus the probability of obtaining the right answer.

This model demonstrates certain basic points about thought experiments which we discuss below.

Heuristic Verification Requirement

The conditions for equation 7 to hold are so important that we express them as the heuristic verification requirement:

Heuristic Verification Requirement:

A sufficient requirement for the heuristic verification method to work is that on the average over all simulations the probability for a confirmation of the correct answer be greater than the probability for a confirmation of any single wrong answer by some fixed amount.

(The average mentioned is the geometric mean.)

If we assume that the heuristic verification requirement is satisfied then equation 7 holds and $r < 1$. This implies that the probability that the thought experiment is correct is bounded from below by a monotonically increasing function of the number of successful verifications. In fact, given enough verifications $P[C \mid A]$ will be arbitrarily close to 1.

On the other hand if even one of the geometric means in equation 5 is bounded from below by a quantity greater than one, then for large enough s, additional verifications make $P[C \mid A]$ worse and not better. If some of the r_k are one and any others have an upper bound less than one then $P[C \mid A]$ eventually stabilizes at a value beyond which no improvement is possible with additional verifications.

Dependent Versus Independent Confirmation

The confirmation probabilities $P_i[a \mid (i-1)a]$ will reduce to $P_i[a]$ if each simulation, i, used is independent of the previous $(i-1)$ simulations. This is often not the case, however. The simulations used in heuristic verification are often simulations for models produced by less extreme versions of the same simplification method. The production of an answer, a, using one version might be correlated with the production of a by another version. This is why the confirming P_i must be expressed in terms of conditional probabilities.

If heuristic verification uses less extreme versions of the same simplification method to verify answers then how might the conditional probabilities for the right and wrong answers behave? First, it is very plausible to believe that $P_i[c \mid (i-1)c]$ is very close to 1 if $i > 1$. In this case the preceding $(i-1)$ simulations have produced correct answers. The ith simulation involves a more realistic (less simplified) model than the preceding ones and it is based on the same type of simplification which produced correct results in these cases. Thus it would be expected to produce a correct answer. Second, $P_i[w_k \mid (i-1)w_k]$ should eventually decline as i increases simply because the models become more realistic as i increases. Unfortunately the correlation produced by using the same method might make this decline slow.

Fine Versus Coarse Descriptions

A last consideration concerns the question of whether a coarse or a fine description is more reliable. The explicit factor of n in equation 7 suggests that the smaller the number of alternatives in our description

of results the larger $P[C \mid A]$ is. This assumption is somewhat dangerous because r will depend on n also. For example, assume $n = 3$ and each simulation has the same probabilities: $P[c] = 3/9, P[w_1] = P[w_2] = P[w_3] = 2/9$. If we halve the number of categories by combining c with w_1 and w_2 with w_3 we obtain $P[c'] = 5/9$, $P[w_1'] = 4/9$, and $n = 1$. In the first case, $r = P[w_i]/P[c] = 2/3$; in the second case $r = P[w_i']/P[c'] = 4/5$. Any raising of r dominates the lowering of n if s is high enough. On the other hand, if s is small enough then coarser categories are better. For example, if $s = 1$ then $P[C \mid A]$ is just $P[c]$ and when categories are combined it is always true that $P[c'] \geq P[c]$ so coarser categories are better.

Efficiency of Thought Experiments

The next issue after correctness of thought experiments is their efficiency compared to a direct solution of the problem. We analyze thought experiments in terms of the expected time, T_E, for achieving a verified result. From this we prove a theorem showing how to achieve maximum efficiency. We then derive upper bounds on the probability that a thought experiment fails to be verified and on the average time required for a thought experiment. We next describe how to estimate efficiency parameters and give a heuristic rule for maximizing efficiency. Finally, we discuss practical efficiency issues.

Time Requirements

Let T_n be the total time required for the nth thought experiment. This time includes the time required for simplification, conjecture, and most importantly, the time required for verification. Verification may require a considerable amount of time since it usually involves solving at least one other simplified model. Let P_n be the probability that the nth thought experiment fails. We assume these probabilities are independent. Thus we assume thought experiments are not only functionally, but also statistically independent. With these assumptions, the expected time, T_E, required for a successful thought experiment is

$$T_E = T_1 + T_2P_1 + T_3P_1P_2 + T_4P_1P_2P_3 + \ldots \quad (8)$$

Thus, T_E depends on the whole sequence of thought experiments which might be performed.

The probability, P_F, that the problem solver fails to obtain any verified solution is just

$$P_F = P_1P_2 \ldots P_n, \quad (9)$$

where n is the number of possible thought experiments which may be performed.

Let us briefly consider the meaning of equations 8 and 9. If there were an infinite series of possible thought experiments n would approach infinity. If each of the P_i were bounded from above by a number less than 1 then P_F would approach zero. In this case, we would expect the problem solver to always obtain a verified solution. Since we do not actually have an infinite series the problem solver may fail. T_E represents the average time taken to either obtain a verified solution or to fail.

Thought Experiment Ordering

Our first application of equation 8 for T_E is to prove the thought experiment ordering theorem.

Let the time required to solve the problem by direct qualitative simulation be S_o; we define the effectiveness, E_i, of a thought experiment to be $E_i = (S_o/T_i)(1 - P_i)$. The significance of this definition will be seen later.

Thought Experiment Ordering Theorem

A sequence of independent thought experiments will have a minimum expected time for achieving a verified result if the sequence is performed in order of effectiveness from most effective to least effective.

Proof:

Consider thought experiment i and $i + 1$ which are adjacent in the sequence for T_E used in equation 8. If we interchange the order in which these experiments are performed the only terms in the series which change are the terms involving T_i and T_{i+1}.

$$T_E(original) = T_i(P_1 \ldots P_{i-1}) + T_{i+1}(P_1 \ldots P_{i-1})P_i + R$$

$$\begin{aligned} T_E(exchanged) &= T_{i+1}(P_1 \ldots P_{i-1}) + \\ &\quad T_i(P_1 \ldots P_{i-1})P_{i+1} + R \end{aligned}$$

By algebraic manipulation we discover that $T_E(exchanged) < T_E(original)$ if and only if $(1/T_{i+1})(1 - P_{i+1}) > (1/T_i)(1 - P_i)$. Multiplying by S_o we obtain the following exchange lemma:
Exchange lemma: $T_E(exchanged) < T_E(original)$ if and only if $E_{i+1} > E_i$.

We prove the ordering theorem by contradiction. If an optimal sequence were not in the order given by the theorem then by the exchange lemma we could improve the sequence. This would contradict the assumption that the sequence was optimal. *Q.E.D.*

Upper Bounds

Probability of Failure First, let us assume that we employ only simplification methods which are useful. If we have a finite number of simplifications, we have a finite number of P_i representing probability of failure. One of these P_i has a maximum value; call it P. Any simplification method which produces thought experiments which always tend to fail verification ($P_i = 1$) is dropped from a problem solver as useless. Therefore, P is an upper bound on the probability of failure of any simplification and this upper bound is less than one. We call this the *usefulness assumption*, equations 10, 11.

$$P_i \leq P \tag{10}$$

$$P < 1 \tag{11}$$

The usefulness assumption provides an upper bound on the probability of failure, P_F, of a thought experiment given in equation 9.

$$P_F \leq P^n \tag{12}$$

As pointed out earlier, assumption of 10 and 11 guarantees that as the number of thought experiments increases, the probability of failure, P_F, may be made as small as we please.

Average Time Next, let us determine an upper bound for T_E. This will depend on a bound for the T_i as well as a bound for the P_i. We assume that for all i,

$$T_i \leq T. \tag{13}$$

There might be some thought experiments which could go on forever without reaching a conclusion. This would be due to the fact that the qualitative simulation did not terminate. In practice any thought experiment which goes on too long can be terminated and verification considered to have failed. Thus assumption 13 is, in fact, acceptable.

Given equations 10, 13 and the fact that all quantities are positive, an upper bound for T_E is $T_E \leq T + TP + TP^2 + \ldots$. This is not an infinite series, but it is approximated by an infinite series if many different simplifications are possible. Furthermore, since all terms are positive the infinite series is certainly an upper bound. We have an ordinary geometric series and since $P < 1$ it converges. Thus

$$T_E \leq T/(1-P). \tag{14}$$

The improvement ratio over straight simulation is S_o/T_E. Using 14 we have a lower bound on this of $S_o/T_E \geq (S_o/T)(1-P)$. This bound is just an effectiveness calculated using T, and P; thus, it provides us with an interpretation of the effectiveness of a thought experiment. The significance of the effectiveness, E_i, of thought experiment i is that if every thought experiment, k, is no worse than the given experiment, i, ($S_o/T_k \geq S_o/T_i$, and $P_k \leq P_i$) then the thought experiment problem solver is faster than straight simulation by a factor of at least E_i.

Estimation of Parameters

Can we obtain estimates for parameters such as P_i and T_i which characterize thought experiments? This is one of the most important issues for practical applications. One approach is to simply assume various values for these parameters for the sake of theoretical analysis. Another approach is to attempt to make empirical estimates for them even if they are crude.

In order to make empirical estimates, we must distinguish between specific thought experiments which are attempts to solve unique problems and the simplification methods which are used in the thought experiments. Any specific thought experiment either succeeds or it doesn't. If we repeat it we always get the same result. A simplification method on the other hand, may produce a verifiable result if used in one thought experiment but not in another. If we have enough experience with a thought experiment problem solver we can collect rough statistics to indicate the frequency with which a given simplification method is useful in giving verifiable results. We can also estimate the time improvement ratio S_o/T_i for thought experiments using a given simplification. The estimates are crude because the success of a simplification method may depend on the type of problem, i.e., the characteristics of the individual problem space. If the problem solver has not encountered a similar type of problem before, the frequency based estimate may not be very reliable. Another reason for the estimates being crude is that previously encountered problems may not constitute a random or sufficiently large sample of the problem space.

Next, let us make some estimates of T_i in terms of more basic quantities. A thought experiment involves a simplification and a generalization step. The simplification step involves finding a simplified version of the problem and performing a qualitative simulation on that simplified version. Finding a simplified version takes a constant amount of time which is small compared to the time required to perform the qualitative simulation. The generalization step involves making a conjecture and verifying that conjecture. Making a conjecture takes a constant amount of time which is small compared to the time required for a qualitative simulation. Ignoring these small quantities $T_i = S_i + V_i$. S_i is the time required for qualitative simulation, and V_i is the time required for verification.

To make our estimate for T_i more useful we must make some rough estimates concerning verification. Verification usually involves performing at least one additional qualitative simulation and comparing the results to the previous simulation. Making the simplification and comparing the results would take a small amount of time compared with S_i. Our estimate for T_i becomes $T_i = nS_i'$ where n is the number of qualitative simulations performed and S_i' is the average time taken for each. A reasonable estimate for n is 2. The results of the original simulation are checked by comparing with one additional simulation. In some cases n could be 1. This would mean that the result would be compared with the results from previous failed thought experiments for the same problem. It would also make analysis more difficult as the thought experiments would no longer be independent. Whether a thought experiment succeeded or failed would depend on the ordering of thought experiments. In some cases the simplification method might specify an n greater than 2 but these are probably rare. Considering both

these effects an estimate of 2 seems reasonable. We will assume $T_i = S'_i$; the analysis would not be greatly different if 3 or some other small number were chosen.

The effectiveness of a thought experiment becomes $E_i = (S_o/S'_i)(1 - P_i)/2$. S_o/S'_i can be considered the average amount of simplification achieved by the simplification method used in the ith thought experiment. It represents an average time improvement factor in using the two simplified simulations in the thought experiment versus simulating the original problem. This parameter is important because we can often rank simplification methods by (S_o/S'_i). Examination of two simplification methods will often indicate which is more extreme and should therefore yield a larger value of (S_o/S'_i). On the other hand, knowledge required to estimate P_i may be more difficult to obtain. In this situation, the thought experiment ordering theorem suggests the following heuristic ordering rule:

Heuristic Ordering Rule:

If independent simplification methods can be ranked by degree of simplification, but no information is available about verification probabilities then a thought experiment problem solver should try the simplification methods in order of degree of simplification from strongest to weakest.

If S is an upper bound on the time any of the qualitative simulations might take then we can estimate equation 14 by:

$$T_E < 2S/(1 - P). \quad (15)$$

We want the thought experiment method to take less time than solving the original problem directly by qualitative simulation. If qualitative simulation of the original problem takes time S_o, we require $S_o/T_E > 1$. This means that we would like our upper bound to be such that $(S_o/S)(1 - P)/2 > 1$; perhaps much greater than 1. Since S is an upper bound on the S_i, S_o/S is a lower bound on the amount of simplification used in any individual thought experiment.

Practicality

Are thought experiments practical? Unless absolute verification is available even a successful thought experiment provides only a reasonable conclusion and not a certain one. Thus, this method will be used only when direct simulation is not feasible, usually because it would take too long. This is, in fact, the case with most real world problems. In order for the thought experiment problem solver to be worthwhile we must be able to find simplifications which have an effectiveness which is greater than one, preferably much greater than one. In order to achieve this we need an extreme degree of simplification with reasonable probability of verification. For example, if probability of verification is at least 1/5 then P is 4/5 and we need a degree of simplification $(S_o/S) > 10$. Experience with TEPS is too limited to take any estimates very seriously; this is a possibility for future research. With TEPS so far, the values for P are normally less than one half. It seems likely that simplifications would be dropped from a problem solver if they fail verification in the great majority of cases.

It might be argued that we need exhaustive statistical testing to determine that the effectiveness for a given simplification is adequately high to be of use. This is not true if the degree of simplification provided by the simplification method is high enough. In that case, demonstration of even a few successful verifications makes it reasonable to believe that E_i is large enough. For example, if $(S_o S_i) > 100$ then $E_i > 1$ if $(1 - P_i) > 0.02$; if $(S_o/S_i) > 1000$ then $E_i > 1$ if $(1 - P_i) > .002$; if $(S_o/S_i) > 1,000,000$ then $E_i > 1$ if $(1 - P_i) > .000002$. Since $(1 - P_i)$ represents probability of successful verification even a very few successful examples indicate that the effectiveness is high enough if the degree of simplification is really large.

Conclusions

We have provided a preliminary theoretical framework for the thought experiment methodology. This framework combined with the empirical experience with TEPS suggests that this is a viable technique. It does not replace conventional methods of qualitative reasoning but rather augments them. Further development of this technique seems desirable.

There are many possibilities for future development. Analysis of known simplification techniques and development of new ones is an important area of research. Theoretical analyses of simplifications should be attempted if possible and statistics on practical effects should be obtained. On a practical side, it would be useful to explore thought experiment problem solvers as a basis for tutoring systems. The simplified model automatically generated by the problem solver could be a basis for helping students understand the original problem.

References

Falkenhainer B. and Forbus K. D. 1990. Compositional modeling of physical systems. In *4th International Workshop on Qualitative Physics.*

Forbus K. D. 1984. Qualitative process theory. *Artificial Intelligence*, 24:85–168.

Hibler D. L. 1992. *The Thought Experiment Method: A New Approach to Qualitative Reasoning.* PhD thesis, The University of South Carolina.

Hibler D. L. and Biswas G. 1992a. Teps: applying the thought experiment methodology to qualitative problem solving. In B. Faltings and P. Struss, editors, *Recent Advances in Qualitative Physics*, pages 345–360, MIT Press.

Hibler D. L. and Biswas G. 1992b. Thought experiments as a framework for multi-level reasoning. In *Working Papers of the Sixth International Workshop*

on Qualitative Reasoning about Physical Systems, August.

Hibler D. L. and Biswas G. 1989. The thought experiment approach to qualitative physics. In *IJCAI-89*, pages 1279–1284.

de Kleer J. and Brown J. S. 1984. A qualitative physics based on confluences. *Artificial Intelligence*, 24:7–83.

Kuipers B. 1987. Abstraction by time-scale in qualitative simulation. In *Proceedings of AAAI-87*, pages 621–625.

Kuipers B. 1986. Qualitative simulation. *Artificial Intelligence*, 29:289–388, 1986.

Prigogine I. and Stengers I. 1984. *Order Out of Chaos.* Bantam Books.

Weld D. S. 1988. Exaggeration. In *Proceedings of the Seventh National Conference on Artificial Intelligence*, pages 291–295, Minneapolis, MN, August.

Weld D. S. 1986. The use of aggregation in causal simulation. *Artificial Intelligence.*

Weld D. S. and Addanki S. 1990. Task-driven model abstraction. In *Fourth International Workshop on Qualitative Physics.*