

The Design of Discrimination Experiments

SHANKAR A. RAJAMONEY

RAJAMONE@POLLUX.USC.EDU

Computer Science Department, University of Southern California, Los Angeles, CA 90089-0781

Abstract. Experimentation plays a fundamental role in scientific discovery. Scientists experiment to gather data, investigate phenomena, measure quantities, and test theories. In this article, we address the problem of designing experiments to discriminate between two competing theories. Given an initial situation for which the two theories make the same prediction, the experiment design problem is to determine how to modify the situation such that the two theories make different predictions for the modified situation. The modified situation is called a *discrimination experiment*. We present a knowledge-intensive method called DEED for designing discrimination experiments. The method analyzes the differences in the two theories' explanations of the prediction for the initial situation. Based on this analysis, it determines modifications to the initial situation that will result in a discrimination experiment. We illustrate the method with the design of experiments to discriminate between several pairs of qualitative theories in the fluids domain.

Keywords. Experiment design, learning by experimentation, scientific discovery, theory revision, qualitative physics

1. Introduction

Experimentation, the purposeful interaction with the world to obtain additional information, plays a fundamental role in scientific discovery. Scientists experiment to gather data, investigate phenomena, measure quantities, and test theories. Experimentation is a complex process involving several interrelated activities such as design, instrumentation, measurement, and data interpretation. This article addresses the design of a specific class of experiments, called *discrimination experiments*, which discriminate between two competing theories.

Given an initial situation for which the two competing theories make the same prediction P_1 , the experiment design problem is to determine how to modify the situation into one in which the two theories will make different predictions P_1 and P_2 .¹ The modified situation is a discrimination experiment. In particular, we wish to examine this experiment design problem when a) the number of modifications to a situation is large, and b) prediction is expensive. With these constraints, a simple brute-force search of the space of possible modifications to the initial situation for a discrimination experiment is clearly inappropriate.

Instead, in this article, we propose a knowledge-intensive method for designing discrimination experiments, called DEED (for the design of discrimination experiments based upon explanation differences). Since the design of discrimination experiments is, in general, difficult, the method makes two simplifying assumptions: 1) the given initial situation is such that, though the two theories make the same prediction, the reasoning leading to the prediction (or the *explanation*) is different, and 2) the differences between the two theories are few. The DEED method analyzes the differences in the two theories' explanations of the prediction for the initial situation. Based on this analysis, it determines how to modify

the initial situation into a discrimination experiment. The first assumption ensures that the method will find some differences between the two explanations. The second assumption simplifies the analysis of the two explanations and the search for modifications to the initial situation.

The remainder of the article is organized as follows. The next section describes our representation of theories, situations, predictions, and explanations. The third section presents the DEED method for designing discrimination experiments. Finally, the last section discusses related work on experiment design and some conclusions.

2. Knowledge representation

We use a causal qualitative representation, Qualitative Process (QP) theory² (Forbus, 1984), to model physical domains.³ In this representation, continuous-valued properties of domain objects and their interactions, such as temperature and pressure, are called *quantities*. There are two types of causal relations between quantities: 1) Qualitative Proportionality. q_1 is qualitatively proportional to q_2 , denoted by $q_1 \propto_{Q\pm} q_2$, if q_1 is monotonic increasing or decreasing in its dependence on q_2 . For example, the acceleration of an object is qualitatively proportional to the force on it; when the force is increasing, steady, or decreasing, the acceleration is also increasing, steady, or decreasing, respectively. 2) Direct Influence. q_1 is directly influenced by q_2 , denoted by $1_{\pm}[q_1, q_2]$, if the derivative of q_1 equals q_2 , assuming there are no other influences on q_1 . For example, the velocity of an object is positively influenced by its acceleration; when the acceleration is positive, zero, or negative, the velocity is increasing, steady, or decreasing, respectively.

Domain objects and their interactions are represented by rules. The left-hand side of a rule specifies either an object or the conditions under which an interaction occurs, and the right-hand side of the rule describes the qualitative causal relations that hold for the object or interaction. Figure 1A shows a simple qualitative theory for fluids. The rules describe physical objects, contained liquids, evaporation, liquid flow, and heat flow. For example, the rule for evaporation states that, when a contained liquid is exposed and the pressure of its vapor is less than the saturation vapor pressure for the liquid, evaporation occurs, resulting in a decreasing influence on the evaporating liquid's amount⁴ at a rate that is qualitatively proportional to the vapor pressure.

The analysis of a domain situation is performed in three steps. In the first step, the domain theory rules are applied to the situation to determine the causal relations that hold. In the second step, since both types of causal relations—qualitative proportionalities and direct influences—specify *partial* influences, the total influence on a quantity is determined by combining all the causal relations affecting it. For example, the amount of water in a tank may be directly influenced by the rates of an inflow of water into the tank and an outflow of water from the tank:

($1 + (\text{amount-of tank-water}) \text{ in-flow-rate}$) and

($1 - (\text{amount-of tank-water}) \text{ out-flow-rate}$).

A) A Qualitative Theory for the Fluids Domain:

(physob ?obj): (physob ?obj) \rightarrow (temp ?obj) \approx_{Q+} (heat ?obj)

(CL ?subst ?can):

?cl=(CL ?subst ?can) \rightarrow (physob ?cl) \wedge (level ?cl) \approx_{Q+} (amount-of ?cl) \wedge (pressure ?cl) \approx_{Q+} (level ?cl)

(evaporation ?evap-liq):

?evap-liq=(CL ?subst ?can) \wedge (exposed ?evap-liq) \wedge (pressure (CG ?subst ?can)) < saturation-vapor-pressure
 \rightarrow ?evap-rate=(rate (evaporation ?evap-liq)) \wedge ?evap-rate > 0

\wedge ?evap-rate \approx_{Q-} (pressure (CG ?subst ?can)) \wedge (I- (amount-of ?evap-liq) ?evap-rate))

(liquid-flow ?src ?dst ?path):

?src=(CL ?subst ?can) \wedge ?dst=(CL ?subst ?can) \wedge (path ?path) \wedge (connects ?path ?src ?dst)

\wedge (flow-aligned ?path) \wedge (pressure ?src) > (pressure ?dst)

\rightarrow ?flow-rate=(rate (liquid-flow ?src ?dst ?path)) \wedge ?flow-rate > 0 \wedge flow-rate \approx_{Q+} (pressure ?src)

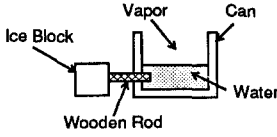
\wedge flow-rate \approx_{Q-} (pressure ?dst) \wedge (I- (amount-of ?src) ?flow-rate) \wedge (I+ (amount-of ?dst) ?flow-rate)

(heat-flow ?src ?dst ?path):

(physob ?src) \wedge (physob ?dst) \wedge (path ?path) \wedge (heat-conducts ?path ?src ?dst) \wedge (temp ?src) > (temp ?dst)

\rightarrow ?hf-rate=(rate (heat-flow ?src ?dst ?path)) \wedge ?hf-rate > 0 \wedge ?hf-rate \approx_{Q+} (temp ?src)

\wedge ?hf-rate \approx_{Q-} (temp ?dst) \wedge (I- (heat ?src) ?hf-rate) \wedge (I+ (heat ?dst) ?hf-rate)

B) A Domain Situation:**C) The Representation of the Domain Situation:**

(CL water can)

(CG water can)

(physob ice-block)

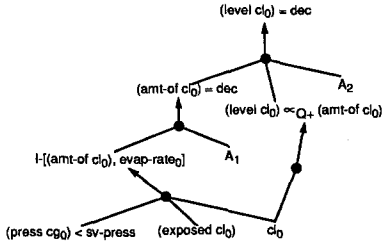
(path wooden-rod)

(exposed (CL water can))

(connects wooden-rod ice-block (CL water can))

(temp (CL water can)) > (temp ice-block)

(pressure (CG water can)) < saturation-vapor-pressure

D) An Explanation:

Key:

c₀ = (CL water can)

c₀ = (CG water can)

evap-rate₀ = (rate (evaporation c₀))

sv-press = saturation-vapor-pressure

A₁ = CWA{(amt-of c₀): I-[(amt-of c₀), evap-rate₀]}

A₂ = CWA{(level c₀): (level c₀) \approx_{Q+} (amt-of c₀)}

amt-of = amount-of

press = pressure

Figure 1. A) A qualitative theory for fluids. Rules defining physical objects, contained liquids, evaporation, liquid flow, and heat flow are shown. B) A situation in which an ice block is connected by a wooden rod to some water in a can. C) The representation of the situation. D) An explanation for the prediction that the level of the water in the can is decreasing.

The two rates must be combined to determine the total influence on the amount. If the rate of inflow is greater than, equal to, or less than that of the outflow, the total influence on the amount is positive, zero, or negative, respectively.

In the third step, inference rules are applied to determine the qualitative changes (increasing, decreasing, or steady) in the value of each quantity. Two such inference rules are

If the total direct influence on a quantity is positive, the value of the quantity is increasing.

If a quantity q_1 is qualitatively proportional to a quantity q_2 , if the value of q_2 is increasing, and if no other quantities affect q_1 , then the value of q_1 is increasing.

Since the inferences are made from partial causal relations, they are based upon *closed-world assumptions* (cwa) that explicitly state the conditions under which the inferences hold. We specify a closed-world assumption by listing all the partial causal relations on which the inference depends. For the above example involving the tank water, if the inflow rate is greater than the outflow rate, the total influence on the amount of the tank water is positive. Accordingly, based upon the first inference rule above, the amount can be inferred to be increasing, and this inference is based upon the closed-world assumption:

CWA{(amount-of tank-water):
 (1+ (amount-of tank-water) in-flow-rate),
 (1- (amount-of tank-water) out-flow rate) }

which states that the two listed direct influences are the only influences on the amount of tank water that were considered in making the inference.

Figure 1B shows a simple situation drawn from the fluids domain in which a block of ice is connected by a wooden rod to some water in an open can. Figure 1C shows the representation of this situation and the initial conditions. The application of the domain theory rules and the inference rules to the situation results in predictions such as that the level of water in the can is decreasing. Figure 1D shows a causal explanation for this prediction, which describes how it was derived from the given situational facts and two closed-world assumptions, A_1 and A_2 . The explanation states that since the water in the can is exposed and its vapor is unsaturated, evaporation occurs, resulting in a negative influence on its amount. Since there are no other influences on the amount (the closed-world assumption A_1) and the rate of evaporation is positive,⁵ the amount is decreasing. The level of the water is qualitatively proportional to its amount. Since there are no other causal relations influencing the level (the closed-world assumptions A_2), it too is decreasing.

3. The DEED Method

The DEED method determines how to modify a given initial situation into a discrimination experiment. A situation may be modified by adding or removing objects, changing their configuration, or specifying new initial conditions. Examples of these modifications to the situation shown in figure 1B are adding a tank of water or removing the ice block, closing the can or breaking the path connecting the ice block and the water, and making the initial vapor pressure equal to the saturation vapor pressure.

Let T_1 and T_2 be the two competing theories, P_1 the common prediction for the initial situation, and E_1 and E_2 the two theories' explanations of this prediction. In the discrimination experiment, let theory T_1 continue to make the prediction P_1 , while theory T_2 makes a different prediction P_2 . Let E'_1 and E'_2 be the explanations of the predictions P_1 and P_2 (for the modified situation) constructed by theories T_1 and T_2 , respectively.

A method for designing discrimination experiments must select features of the initial situation that are to be modified and must decide how to modify the selected features. The DEED method determines suitable features and the modifications to them by examining the explanations for the common prediction in the initial situation. The method addresses two questions: a) How can the explanations E_1 and E_2 be modified simultaneously to generate the explanations E'_1 and E'_2 ? b) How can the initial situation be modified so that the modifications to the explanations work? By analyzing the differences between explanations E_1 and E_2 , the method determines how to modify them to generate the explanations E'_1 and E'_2 . The modifications to the explanations determine which features of the situation are to be modified and how they are to be modified.

In the following subsections, we present an example, define the operators for modifying explanations, present the steps of the DEED method, and discuss its limitations.

3.1. An example

The theory for the fluids domain shown in figure 1A is incorrect because it fails to predict that the water's temperature is decreasing in the situation shown in figure 1B. Consider four revisions to the theory that will allow this prediction to be made:

- \mathcal{H}_2 : Adding $1 - [(\text{heat } ?\text{evap-liq}), ?\text{evap-rate}]$ to the evaporation rule. During evaporation, the heat of the evaporating liquid is conjectured to be negatively influenced by the rate of evaporation.
- \mathcal{H}_2 : Replacing $(\text{conducts-heat } ?\text{path } ?\text{src } ?\text{dist})$ by $(\text{connects } ?\text{path } ?\text{src } ?\text{dist})$ in the heat flow rule. A path connecting two physical objects at different temperatures is conjectured to be sufficient for a heat flow to occur; properties of the material of the path (e.g., conductivity) are irrelevant.
- \mathcal{H}_3 : Adding $(\text{temperature } ?\text{evap-liq}) \propto_{Q+} (\text{amount-of } ?\text{evap-liq})$ to the evaporation rule. During evaporation, the temperature of the evaporating liquid is conjectured to be qualitatively proportional to its amount.
- \mathcal{H}_4 : Adding $(\text{temperature } ?\text{cl}) \propto_{Q+} (\text{amount-of } ?\text{cl})$ to the contained-liquid rule. The temperature of a liquid contained in a vessel is conjectured to be qualitatively proportional to its amount.

The application of these four hypothesized revisions to the original theory of figure 1A yields four competing theories: $T_{\mathcal{H}_1}$, $T_{\mathcal{H}_2}$, $T_{\mathcal{H}_3}$, and $T_{\mathcal{H}_4}$ (formed under the hypotheses \mathcal{H}_1 , \mathcal{H}_2 , \mathcal{H}_3 , and \mathcal{H}_4 , respectively).

All four theories predict that, for the situation shown in figure 1B, the water in the can evaporates, causing the water's amount and level to decrease (figure 1D). In addition, all four theories also predict that the water's temperature is decreasing.⁶ The four explanations, E_1 , E_2 , E_3 , and E_4 , for this prediction are shown in figure 2.⁷ Theory $T_{\mathcal{H}_2}$ predicts that a heat flow from the contained water to the ice block through the connecting wooden rod occurs, causing the water's heat and temperature to decrease. The three other theories do not predict a heat flow, since they require the wooden rod to be heat conducting. Instead, they explain the decrease in the water's temperature as an effect of its evaporation. Theory

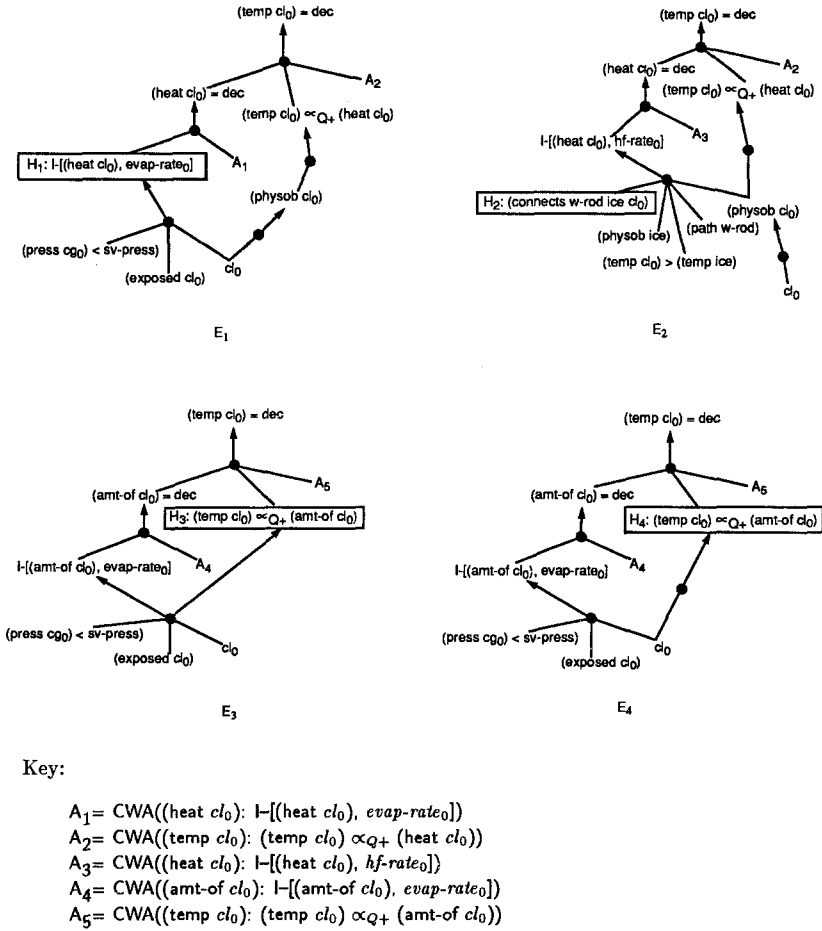


Figure 2. The four explanations E_1 , E_2 , E_3 , and E_4 constructed by the four theories $T_{\mathcal{H}_1}$, $T_{\mathcal{H}_2}$, $T_{\mathcal{H}_3}$, and $T_{\mathcal{H}_4}$, respectively, for why the water's temperature is decreasing.

$T_{\mathcal{H}_1}$ predicts that the water's heat is decreasing as a direct effect of its evaporation, since evaporation is conjectured to negatively influence an evaporating liquid's heat. Theory $T_{\mathcal{H}_3}$ predicts that the decrease in the water's amount due to its evaporation causes its temperature to decrease also, since the temperature of an evaporating liquid is conjectured to be qualitatively proportional to its amount. Finally, theory $T_{\mathcal{H}_4}$ predicts that the decrease in the water's amount due to its evaporation causes its temperature to decrease also, since the temperature of a contained liquid is conjectured to be qualitatively proportional to its amount.

Consider the problem of designing experiments to discriminate between theories $T_{\mathcal{H}_1}$ and $T_{\mathcal{H}_2}$. Since, for the initial situation (figure 1B), both theories predict that the water's temperature is decreasing, modifications to the initial situation are desired such that, for the modified situation, one theory will predict that the temperature is decreasing, while the other will predict that it remains steady (or is increasing). The initial situation may

be modified in many ways: the water may be covered, the ice block may be replaced by a wooden block, the wooden rod may be removed, more water may be added, and so forth. An examination of the initial situation alone does not determine which of these modifications will result in a discrimination experiment.

However, an inspection of explanations E_1 and E_2 shows that modifications to the initial situation such as closing the container or moving the wooden rod so that it no longer connects the ice block and the water will result in discrimination experiments. In the first case, both theories will predict that evaporation does not occur. Therefore, theory $T_{\mathcal{H}_1}$ will predict that the water's temperature remains steady in the modified situation. However, theory $T_{\mathcal{H}_2}$ will continue to predict that the water's temperature is decreasing, since the hypothesized heat flow is not affected by this modification. Similarly, in the second case, theory $T_{\mathcal{H}_1}$ will continue to predict that the water's temperature is decreasing in the modified situation, since moving the wooden rod does not affect evaporation. However, theory $T_{\mathcal{H}_2}$ will predict that the water's temperature remains steady, since the heat flow is no longer feasible.

Consider the problem of discriminating theories $T_{\mathcal{H}_1}$ and $T_{\mathcal{H}_4}$. Notice that both the above experiments fail. The first fails because both theories predict that evaporation does not occur in the modified situation and, therefore the water's temperature remains steady. The second fails because moving the wooden rod has no effect on evaporation; both theories continue to predict that the water's temperature is decreasing in the modified situation. The DEED method, by analyzing the differences in the two explanations, determines that modifications to the initial situation such as moving the wooden rod away and closing the container result in discrimination experiments for theories $T_{\mathcal{H}_1}$ and $T_{\mathcal{H}_2}$ but not for theories $T_{\mathcal{H}_1}$ and $T_{\mathcal{H}_4}$.

3.2. Operators for modifying explanations

The supports (leaves) of an explanation are of two types: situational facts and closed-world assumptions. The derivation of a prediction, given a set of supports, follows automatically from the application of the domain theory and the inference rules. Since these rules are invariant, an explanation may be modified only by changing its supports. We define two basic operators for modifying them:

1. **Defeat-Situational-Support(s)**. This operator removes a specified situational fact, s , from the explanation. In the physical situation, this operation corresponds to removing the fact along with other associated facts from the situation. Consider explanation E_1 shown in figure 2. It has three situational supports that may be defeated by applying this operator: (pressure cg_0) < sv-pressure, (exposed cl_0), and cl_0 . In the physical situation, the defeating of these supports corresponds to a) changing the initial condition to make the vapor's pressure equal to or greater than the saturation vapor pressure, b) closing the container, and c) removing the water in the container and other facts associated with the water (e.g., (exposed cl_0)).
2. **Augment-CWA-Support(CWA($q: i_1, \dots, i_n, i_{n+1}, T_{\mathcal{H}_i}$))**. This operator modifies a cwa support of an explanation. It adds the specified causal relation i_{n+1} and constructs an explanation justifying it by backward chaining on the rules of the domain theory $T_{\mathcal{H}_i}$.

A new value for q is computed by combining the previous causal relations (specified by the cwa) and the newly introduced causal relation. In the physical situation, this operation corresponds to adding the objects (configured appropriately) and the initial conditions that support the newly introduced causal relation.

Consider the application of the following instance of the operator:

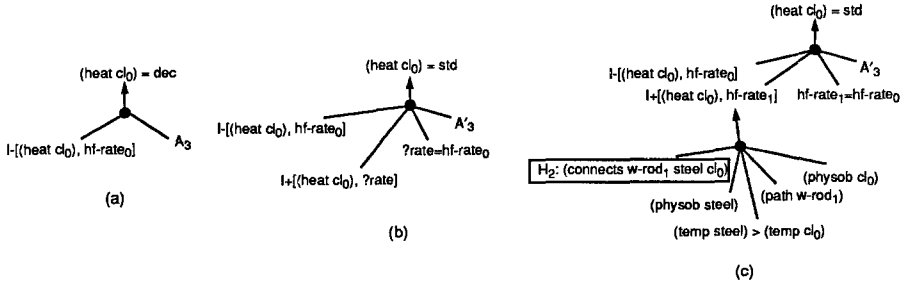
$$\text{Augment-CWA-Support}(A_3, \text{I+}[(\text{heat } cl_0), ?\text{rate}], T_{\mathcal{H}_2})$$

to the cwa support A_3 of E_2 (the relevant fragment of explanation E_2 is shown in figure 3a). Figure 3b shows the addition of the specified influence to the explanation, and the change to the cwa. The new value for the water's heat is computed by combining the two direct influences to determine the total influence, and by applying inference rules. If the rate $?rate$ is selected to equal the rate of the existing heat flow, the water's heat will remain steady.

A supporting explanation for the added influence is constructed by backward chaining over the rules of theory $T_{\mathcal{H}_2}$. The added influence unifies with a consequent of the heat flow rule:

$$(\text{I+}(\text{heat } ?\text{dst}) ?\text{hf-rate}), \text{ with } ?\text{dst} = cl_0.$$

Further backward chaining on the antecedents of the heat flow rule leads to a search for a physical object and a path that connects the object to cl_0 , and the specification



Key:

$\text{steel} = \text{steel-block}$

$\text{w-rod}_1 = \text{wooden-rod}_1$

$\text{hf-rate}_1 = (\text{rate } (\text{heat-flow steel } cl_0 \text{ wooden-rod}_1))$

$A_3 = \text{CWA}((\text{heat } cl_0) : \text{I-}[(\text{heat } cl_0), \text{hf-rate}_0])$

$A'_3 = \text{CWA}((\text{heat } cl_0) : \text{I-}[(\text{heat } cl_0), \text{hf-rate}_0], \text{I+}[(\text{heat } cl_0), \text{hf-rate}_1])$

Figure 3. Augmenting the cwa support A_3 of explanation E_2 . a) The relevant fragment of explanation E_2 from figure 2. b) Adding the specified influence to the explanation. The water's heat, after the addition of the influence, remains steady when the rate of the new influence equals that of the existing heat flow. c) Constructing an explanation to support the added influence. A steel block and a wooden rod connecting the steel block to the water in the can are added to the physical situation. The new initial conditions include 1) the temperature of the steel block is greater than that of cl_0 , and 2) the rate of the new heat flow instance equals that of the old one.

of an initial condition stating that the temperature of the selected object be greater than that of cl_0 . To find a physical object and a path, the system can either search a given database or continue backward chaining (e.g., $(\text{physob } ?\text{obj})$ can unify with the consequent of the contained-liquid rule: $(\text{physob } ?\text{cl})$). Figure 3c shows the final supporting explanation when a physical object and a path (a steel block and a wooden rod connecting the steel block to cl_0 , respectively) are found in the database.

3.3. The steps of the DEED method

Table 1 specifies the DEED method for designing discrimination experiments. The inputs to the method are the two competing theories $T_{\mathcal{H}_1}$ and $T_{\mathcal{H}_2}$, the common prediction $q=v_1$ for the initial situation, and the two theories' explanations E_1 and E_2 ; the outputs are discrimination experiments. Using the operators for modifying explanations, the method modifies explanation E_1 to E'_1 , which continues to support the prediction $q=v_1$, and simultaneously, modifies explanations E_2 to E'_2 , which supports a different prediction $q=v_2$.

Table 1. The DEED method for the design of discrimination experiments.

Design-Discrimination-Experiment($T_{\mathcal{H}_1}$, $T_{\mathcal{H}_2}$, $P: q=v_1$, E_1 , E_2)

Let

- \mathcal{S}_k be the set of situational facts supporting explanation E_k .
- \mathcal{GWA}_k be the set of closed-world assumptions supporting explanation E_k .
- $\text{Supports}(m)$ be the set of situational facts and closed-world assumptions supporting m , an internal node of an explanation.

1. For each s in $\mathcal{S}_2 - \mathcal{S}_1$ do Apply Defeat-Situational-Support(s)

2. For each $cwa(q_1: i_1, \dots, i_n)$ in \mathcal{GWA}_2 do

Find a causal relation i_{n+1} in theory $T_{\mathcal{H}_2}$ that affects q_1 such that the causal relations i_1, \dots, i_n, i_{n+1} yield a new value for q_1

For each such causal relation i_{n+1} do

If there is no $cwa(q_1: j_1, \dots, j_m)$ in \mathcal{GWA}_1

then a) Apply Augment-CWA-Support($cwa(q_1: i_1, \dots, i_n, i_{n+1}, T_{\mathcal{H}_2})$)

else b) Apply Augment-CWA-Support($cwa(q_1: i_1, \dots, i_n, i_{n+1}, T_{\mathcal{H}_2})$) selectively

3. For each situational support s in $\text{Supports}(\mathcal{H}_2) - \text{Supports}(\mathcal{H}_1)$ do

Apply Defeat-Situational-Support(s)

For each affected $cwa(q_1: i_1, \dots, i_n)$ in \mathcal{GWA}_1 do

Find a causal relation i_{n+1} in theory $T_{\mathcal{H}_1}$ that affects q_1 such that the causal relations i_1, \dots, i_n, i_{n+1} reestablish the original value of q_1

For each such causal relation i_{n+1} do

a) Apply Augment-CWA-Support($cwa(q_1: i_1, \dots, i_n, i_{n+1}, T_{\mathcal{H}_1})$) selectively

or b) Apply Augment-CWA-Support($cwa(q_1: i_1, \dots, i_n, i_{n+1}, T_{\mathcal{H}_1})$) exclusively

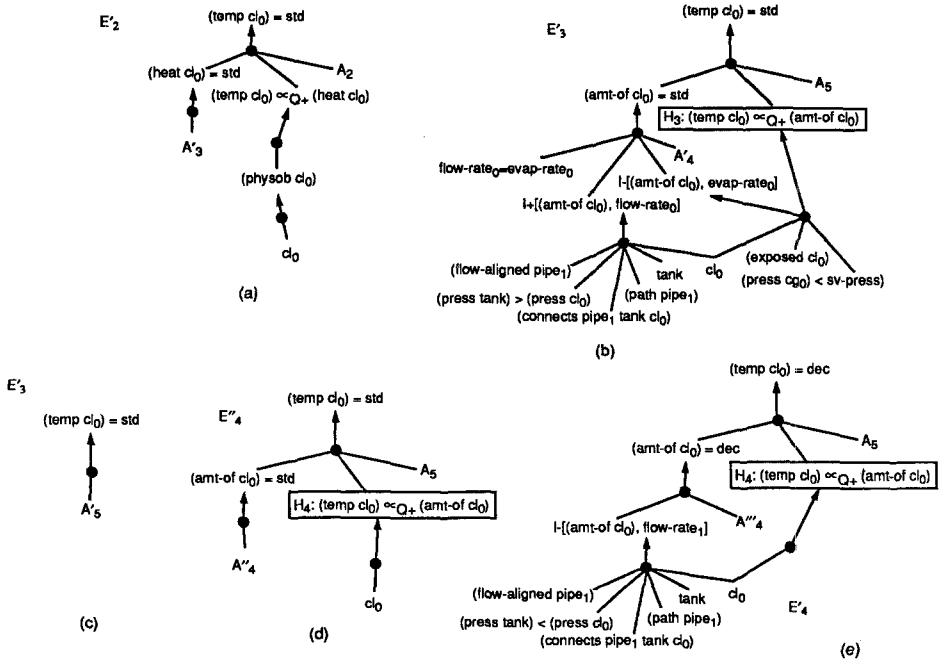
The simultaneous modification of the two explanations E_1 and E_2 to E'_1 and E'_2 can be achieved by modifying explanation E_2 to E'_2 without affecting E_1 (so that $E'_1 = E_1$ continues to support the original prediction) by either selecting for modification a support of explanation E_2 that does not appear in E_1 (steps 1 and 2a) or selecting a common support but performing the modification in such a manner that E_1 is not affected (step 2b) (e.g., the modification holds only under the conjectured revision \mathcal{H}_2). Alternatively, the modifications to E_2 may be permitted to affect E_1 , but the damage to E_1 must be repaired so that the final E'_1 supports the original prediction. To prevent the repair to E_1 from simultaneously repairing E'_2 back to E_2 , either E_1 must be selectively repaired (step 3a) (e.g., the repair holds only under the conjectured revision \mathcal{H}_1) or the method must ensure that explanation E_2 cannot be rebuilt (step 3b).

The three steps of the method design independent discrimination experiments. In step 1, a situational support s of E_2 not common to E_1 is defeated. Since the derivation steps based on the removed situational support are no longer valid in this modified situation, explanation E_2 is modified to E'_2 , which then supports a new prediction $q=v_2$. Defeating the situational support s does not affect E_1 , since s does not appear in the explanation; therefore, in the modified situation, explanation $E'_1 (= E_1)$ continues to support the prediction $q=v_1$.

Consider the application of step 1 to the explanations E_1 and E_2 of figure 2. A situational support of E_2 not common to E_1 is (connects w -rod ice cl_0). The application of the Defeat-Situational-Support operator to this support results in the removal of the support from explanation E_2 . The modified explanation E'_2 is shown in figure 4a. Explanation E_1 is not affected by the application of this operator. The modification to the physical situation simply involves removing the situational fact from the initial situation (e.g., by moving the wooden rod so that it no longer connects the ice block and the water).

Step 2 modifies the cwa supports of E_2 . For a closed-world assumption on a quantity q_1 , the method searches theory $T_{\mathcal{H}_2}$ for a causal relation i_{n+1} that affects q_1 such that q_1 's new value, based upon the combined effect of the previous causal relations and i_{n+1} , is different from its original value. The method modifies E_2 to E'_2 by augmenting the cwa with the causal relation i_{n+1} and constructing the supporting explanation for it. If explanation E_1 does not include a closed-world assumption on q_1 , the modifications to E_2 will not affect it. However, if E_1 also makes a closed-world assumption on q_1 , the modifications must be performed to selectively affect only E_2 by constructing a supporting explanation for i_{n+1} that necessarily includes an instance of \mathcal{H}_2 and, hence, is valid only under theory $T_{\mathcal{H}_2}$.⁸

Consider the application of step 2 to the explanations E_1 and E_3 of figure 2. Suppose the cwa support A_4 of explanation E_3 is selected for modification. To make the amount of water remain steady instead of decreasing, A_4 must be augmented with an influence of the form $l+[(amt-of\ cl_0), ?rate]$ with a rate $?rate$ equal to the rate of evaporation. Since explanation E_1 does not make a closed-world assumption on the amount of cl_0 , the Augment-CWA-Support operator can be applied directly to add this influence. The operator, upon searching theory $T_{\mathcal{H}_2}$, finds that $l+[(amt-of\ ?dst), ?flow-rate]$ from the liquid flow rule unifies with this form with $?dst$ bound to cl_0 , and constructs an explanation supporting the influence. Figure 4b shows the modified explanation E'_2 . Explanation E_1 is not affected by the modifications.



Key:

$flow-rate_0 = (rate\ (liquid-flow\ tank\ cl_0\ pipe_1))$
 $flow-rate_1 = (rate\ (liquid-flow\ cl_0\ tank\ pipe_1))$
 $A'_3 = CWA((heat\ cl_0):)$
 $A'_4 = CWA((amt-of\ cl_0): I-[(amt-of\ cl_0), evap-rate_0], I+[(amt-of\ cl_0), flow-rate_0])$
 $A'_5 = CWA((temp\ cl_0):)$
 $A''_4 = CWA((amt-of\ cl_0):)$
 $A''_5 = CWA((amt-of\ cl_0): I-[(amt-of\ cl_0), flow-rate_1])$

Figure 4. a) Explanation E_2 is modified by defeating the situational support ($connects\ w-rod\ ice\ cl_0$). Since evaporation no longer occurs, the water's heat remains steady. b) To nullify the decrease in the water's amount due to evaporation, an inflow of water is established, with a rate equal to that of evaporation. The water's amount and, consequently, its temperature, will remain steady. c) Explanation E_3 is modified by defeating the situational support ($exposed\ cl_0$). Since evaporation no longer occurs, the water's temperature is unaffected and, hence, remains steady. d) Explanation E_4 is affected by the modification to explanation E_3 and no longer predicts that the water's temperature is decreasing. e) Explanation E_4 is repaired by establishing an outflow to decrease the water's amount and, hence, its temperature.

Consider the application of step 2 to the explanations E_1 and E_2 of figure 2. Suppose the cwa support A_3 of explanation E_2 is selected for modification. To make the water's heat remain steady instead of decreasing, A_3 must be augmented with an influence of the form $I+[(heat\ cl_0), ?rate]$ with a rate $?rate$ equal to the rate of the heat flow. Since explanation E_1 also makes a closed-world assumption on the heat of cl_0 (A_1), the Augment-CWA-Support operator must be selectively applied to add this influence. The operator,

upon searching theory $T_{\mathcal{H}_2}$, finds that $\text{I}+[(\text{heat } ?\text{dst}), ?\text{hf-rate}]$ from the heat flow rule unifies with this form with $?dst$ bound to cl_0 , and constructs an explanation supporting the influence. Since the supporting explanation for the added influence includes an instance of \mathcal{H}_2 : $(\text{connects } ?\text{path } ?\text{src } ?\text{dst})$, namely, $(\text{connects } w\text{-rod}_1 \text{ steel } cl_0)$, the addition of the influence is valid only under theory $T_{\mathcal{H}_2}$. Therefore, explanation E_1 will not be affected by the modification. Figure 3c shows the relevant portion of the modified explanation E'_2 .

Step 3 defeats a situational support s of the conjectured revision \mathcal{H}_2 in the explanation E_2 , which does not belong to the situational supports of the conjectured revision \mathcal{H}_1 in the explanation E_1 .⁹ As in step 1, the derivation steps based on s are no longer valid in the modified situation and, therefore, the modified explanation, E'_2 , shifts to support a new prediction. However, explanation E_1 will also be affected if it is supported by s . In this case, it must be repaired such that, in the modified situation, the repaired explanation, E'_1 , continues to support the original prediction. The method identifies the cwa affected by the defeating of the situational support s . For each affected cwa, it searches theory $T_{\mathcal{H}_1}$ for a causal relation i_{n+1} that affects the quantity q_1 (on which the cwa was made) such that the combination of the causal relations remaining in the closed-world assumption and the new causal relation reestablishes q_1 's original value.

To prevent explanation E_2 from being repaired simultaneously, the repair modifications must be performed to selectively affect only E_1 by constructing a supporting explanation for i_{n+1} that necessarily includes an instance of \mathcal{H}_1 and, hence, is valid only under theory $T_{\mathcal{H}_1}$ (step 3a). Alternatively, since defeating the situational support s removed \mathcal{H}_2 from explanation E_2 , if the repair modifications on E_1 are selected such that \mathcal{H}_2 does not receive a supporting explanation under theory $T_{\mathcal{H}_2}$, the repair will not construct any explanation involving \mathcal{H}_2 and, hence, will not regenerate E_2 (step 3b).¹⁰

Consider the application of step 3 to the explanations E_4 and E_3 from figure 2. The situational support of $T_{\mathcal{H}_4}$ is cl_0 , and those of $T_{\mathcal{H}_3}$ are cl_0 , (exposed cl_0), and $(\text{pressure } cg_0) < \text{saturation-vapor-pressure}$. Suppose the situational support (exposed cl_0) is defeated. Figures 4c and 4d show the modified explanations E'_3 and E'_4 . Since defeating the situational support affected explanation E_4 , the damage to E_4 must be repaired. The only cwa support of explanation E_4 affected by the modification is A_4 . To reestablish a decrease in the water's amount, A_4 must be augmented with an influence of the form $\text{I}+[(\text{amt-of } cl_0), ?\text{rate}]$ by applying the Augment-CWA-Support operator. Upon searching theory $T_{\mathcal{H}_1}$, the operator finds two influences that unify:

$\text{I}+[(\text{amt-of } ?\text{evap-liq}), ?\text{evap-rate}]$ from the evaporation rule with $?evap-liq$ bound to cl_0

$\text{I}+[(\text{amt-of } ?\text{src}), ?\text{flow-rate}]$ from the liquid flow rule with $?src$ bound to cl_0 .

Step 3a is not applicable, since neither of these influences can be selectively added to A_4 of explanation E_4 . Step 3b fails for the first influence, since the construction of a supporting explanation leads to the simultaneous construction of a supporting explanation for an instance of \mathcal{H}_3 , namely, $(\text{temp } cl_0) \propto_{Q+} (\text{amt-of } cl_0)$. In fact, the repair results in the

regeneration of the original explanations E_4 and E_3 . Step 3b succeeds for the second influence, since a supporting explanation can be constructed by establishing an outflow of water without invoking evaporation (and, hence, without constructing a supporting explanation for an instance of \mathcal{H}_3). Figure 4e shows the repaired explanation E'_4 .

3.4. Analysis of the DEED Method

In general, the two competing theories' explanations of the prediction for the initial situation may differ in: a) their situational supports, b) their cwa supports, or c) their internal structure. In the first case, step 1 (and, sometimes, step 3) of the method will design discrimination experiments and, in the second case, step 2a will design discrimination experiments. In the third case, step 2b will design discrimination experiments when a cwa-support of one explanation can be augmented selectively without affecting the other explanation, and step 3 will design discrimination experiments when the supports of the hypotheses are different. The method fails to design discrimination experiments when the situational supports of the two explanations are identical (step 1 fails), the cwa supports of the two explanations are identical (step 2a fails), the situational and cwa supports of the instances of the two hypotheses are identical (step 3 fails), and none of the cwa-supports can be augmented selectively (step 2b fails).

The explanation modification operators used by the DEED method effect local changes to the individual supports of the explanation. Since the operators may introduce undesirable interactions or the local changes may not propagate, the global prediction may not change as desired. Therefore, the modified situations generated by the DEED method must be tested further by checking whether the actual predictions made by the two theories (obtained, for example, by simulation) for the modified situation are different.

For example, consider the application of step 2a of the method to the explanations E_1 and E_3 described earlier and illustrated in figure 4b. In this case, the application of the Augment-CWA-Support operator to cwa support A_4 of explanation E_3 resulted in the design of a discrimination experiment in which an inflow of water into the can was established to produce an increase in the water's amount to cancel the decrease due to evaporation. Although locally the application of the operator resulted in the water's amount remaining steady, if the temperature of the water entering the can is much less than that of the water already in the can, then, due to heat transfer, the water's temperature may continue to decrease, instead of remaining steady as desired.¹¹ In this case, the introduction of an inflow to keep the water's amount steady (the desired local change) affected the water's temperature (the global prediction) through alternative means (heat flow), thereby nullifying the experiment.

Theories with many differences pose several problems for the DEED method: firstly, the applicability of step 3 is restricted because the situational support to be defeated must be common to all the instances of all the differences; secondly, applying the Augment-CWA-Support operator selectively or exclusively becomes more complicated, since there are many more alternative explanations to be checked; and, finally, more of the designed experiments can be expected to fail, since there will be many more undesirable non-local interactions.

3.5. The design of experiments to discriminate between theories $T_{\mathcal{H}_1}$ and $T_{\mathcal{H}_2}$

We step through the DEED method to illustrate how it designs experiments to discriminate between the two theories. The explanations E_1 and E_2 for the prediction that the water's temperature is decreasing in the situation shown in figure 1B are shown in figure 2. In step 1, the system finds four situational supports of explanation E_2 that are not common to explanation E_1 : $(\text{temp } cl_0) > (\text{temp } ice)$, $(\text{connects } ice \text{ } cl_0 \text{ w-rod})$, (path w-rod) , and $\text{physob } ice$. The application of the Defeat-Situational-Support operator to each of these situational supports results in the design of four discrimination experiments, in which 1) the water's initial temperature is equal to or less than that of the ice block; 2) the ice block and the contained water are not connected (figure 5a¹²); 3) the wooden rod is removed; and 4) the ice block is removed. In each of these modified situations, theory $T_{\mathcal{H}_2}$ predicts that the water's temperature will remain steady (since a heat flow from the water to the ice block no longer occurs), while theory $T_{\mathcal{H}_1}$ continues to predict that it will be decreasing (since the water continues to evaporate).

In step 2, the method examines the two cwa supports of explanation E_2 :

$$A_3 = \text{CWA}((\text{heat } cl_0) : 1 - [(\text{heat } cl_0), hf\text{-}rate_0]) \text{ and}$$

$$A_2 = \text{CWA}((\text{temp } cl_0) : (\text{temp } cl_0) \propto_{Q+} (\text{heat } cl_0)).$$

The application of the Augment-CWA-Support selectively to the cwa support A_3 results in the design of an experiment in which, in order to cancel the effect of the heat flow from

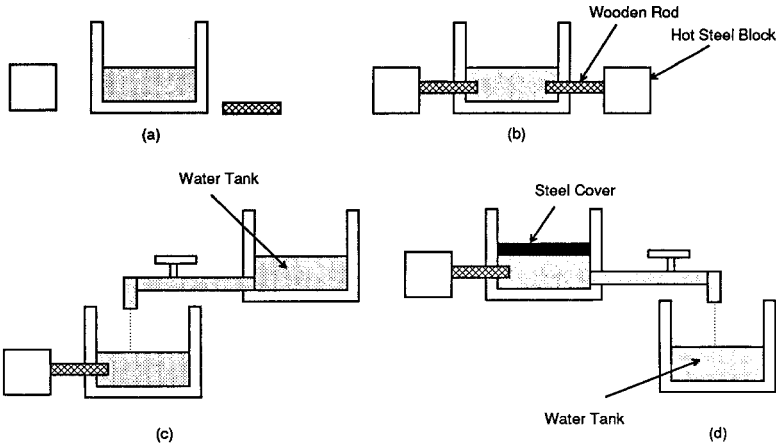


Figure 5. Experiments to discriminate between the different pairs of theories. a) $T_{\mathcal{H}_1}$, $T_{\mathcal{H}_2}$: In the modified situation, the path connecting the contained water and the ice block is broken by moving the wooden rod away. b) $T_{\mathcal{H}_1}$, $T_{\mathcal{H}_2}$: A second heat flow instance is introduced to cancel the effect of the first heat flow instance on the heat of the contained water. c) $T_{\mathcal{H}_1}$, $T_{\mathcal{H}_2}$: An inflow of water into the can is established to cancel the effect of evaporation on the amount of the contained water. d) $T_{\mathcal{H}_4}$, $T_{\mathcal{H}_5}$: The contained water is sealed with a steel cover and an outflow of water into a tank is established to decrease the amount of the contained water.

the water in the can to the ice block, a new heat flow is established by adding a hot steel block connected by a wooden rod such that its rate is exactly equal to that of the existing heat flow. Figure 5b illustrates the designed experiment. The method fails to construct discrimination experiments with the second cwa-support A_2 of explanation E_2 , since it cannot find a suitable causal relation affecting the temperature of the contained water in theory $T_{\mathcal{H}_1}$.

In step 3, the method examines the situational supports of \mathcal{H}_2 in explanation E_2 : (connects ice cl_0 w-rod). Since this situational support does not appear in explanation E_1 , the designed experiment degenerates to the one designed in step 1 in which the path connection is broken (figure 5a).

3.6. The design of experiments to discriminate between theories $T_{\mathcal{H}_1}$ and $T_{\mathcal{H}_3}$

The explanations E_1 and E_3 for the prediction that the water's temperature is decreasing are shown in figure 2. Step 1 of the method fails, since all the situational supports of explanation E_3 are common to explanation E_1 . In step 2, the method examines the two cwa supports of explanation E_3 :

$$A_4 = \text{CWA}((\text{amt-of } cl_0) : 1 - [(\text{amt-of } cl_0), \text{evap-rate}_0]) \text{ and}$$

$$A_5 = \text{CWA}((\text{temp } cl_0) : (\text{temp } cl_0) \propto_{Q+} (\text{amt-of } cl_0)).$$

The application of the Augment-CWA-Support operator to the cwa support A_4 results in the design of an experiment in which, in order to cancel the effect of evaporation on the water's amount, an inflow of water from a tank is set up such that the rate of the inflow is exactly equal to the rate of evaporation (figure 5c). In this modified situation, theory $T_{\mathcal{H}_1}$ predicts that the water's amount and, hence, its temperature will remain steady, while theory $T_{\mathcal{H}_3}$ continues to predict that the water's temperature will be decreasing. For the second cwa-support A_5 , the method fails to construct a discrimination experiment, since it cannot selectively add to explanation E_3 the only suitable causal relation, namely, $(\text{temp } ?cl) \propto_{Q+} (\text{heat } ?cl)$ from the contained-liquid rule with $?cl$ bound to cl_0 . Step 3 of the method also fails, since all the situational supports of \mathcal{H}_3 : $(\text{temp } cl_0) \propto_{Q+} (\text{amt-of } cl_0)$ are common to the supports of \mathcal{H}_1 : $(1 - (\text{heat } cl_0) \text{ evap-rate}_0)$.

3.7. The design of experiments to discriminate between theories $T_{\mathcal{H}_4}$ and $T_{\mathcal{H}_3}$

The explanations E_4 and E_3 for the prediction that the water's temperature is decreasing are shown in figure 2. Step 1 of the method fails, since all the situational supports of explanation E_3 are common to explanation E_4 . Step 2 also fails, since all of the suitable causal relations for augmenting the two cwa supports of explanation E_3 , A_4 and A_5 , cannot be established selectively as required. In Step 3, the method finds two situational supports of \mathcal{H}_3 in E_3 that are not common to the supports \mathcal{H}_4 in E_4 : $(\text{exposed } cl_0)$ and $(\text{press } cg_0) < \text{sv-press}$. Consider defeating the first one, for example, by placing a steel cover

on the water's surface.¹³ The removal of this situational fact affects the cwa-support A_4 of explanation E_4 , and it must be augmented to reestablish the decrease to the amount of the contained water. A water outflow to a tank is set up to cause the water's amount to decrease. Figure 5d illustrates the designed experiment.

4. Discussion

Rajamoney's (1989) COAST system also designs experiments to discriminate between competing theories. It uses three strategies: elaboration, which proposes experiments for easily observable quantities; discrimination, which proposes experiments by checking whether the two theories make different predictions in a given situation; and transformation, which transforms a given situation into one in which the discrimination strategy works. The DEED method described in this article corresponds most closely to the transformation strategy used by COAST. COAST's transformation strategy uses heuristic rules to modify the given situation. The DEED method, on the other hand, examines the differences in the explanations constructed by the two theories to modify the given situation. The DEED method is more general and, since the design of experiments is guided by the explanations, better focused.

Kulkarni and Simon's (1990) KEKADA system also proposes scientific experiments. To design experiments, the system applies general heuristics such as "focus on a surprising phenomenon," "try to magnify a surprising phenomenon," "break up a complex surprising phenomenon into its parts," and "determine the scope of a surprising phenomenon." These heuristics are particularly useful in the design of systematic experiments to investigate poorly understood phenomena, especially ones for which no developed theory exists. In contrast, the DEED method uses the differences in the explanations instead of heuristics to propose experiments. Our method makes extensive use of domain theories in the design of experiments (for example, to find suitable causal relations and construct supporting explanations); therefore, it is applicable primarily to developed theories of well-understood domains.

The BACON system (Langley et al., 1986) designs experiments to collect data for the empirical discovery of laws by systematically varying independent variables. Żytkow's (1987) FAHRENHEIT system designs experiments to determine the scope of these empirically discovered laws. It selectively gathers data by systematically varying independent variables until it finds an empirically justified law. Both systems are primarily data-driven and, therefore, tend to explore the domain with a large number of experiments before converging to a law. In contrast, our DEED method is knowledge-intensive: using the explanations and domain theories, it can determine which features of a situation are important, and how they should be modified. Accordingly, it generates very few experiments.

Cheng (1991) describes a model for experimentation in scientific discovery that includes finding an appropriate paradigm, setting up the experiment, and conducting the test. His system searches the experiment space with control strategies that include rules for selecting profitable paradigms and eliminating simple or irrelevant combinations of experimental setups. The system emphasizes the design of experiments to confirm or disconfirm an individual theory or hypothesis. Cheng's system may explore a large portion of the experiment space, depending on how well the heuristic control strategies work. In contrast, by focusing on the differences between the two competing theories, our DEED method can guide the search into a much smaller portion of the experiment space.

The PRODIGY system (Minton et al., 1987), a general-purpose planner with STRIPS-like operators, also incorporates a subsystem (Carbonell & Gil, 1987) that proposes experiments to refine an incorrect domain theory. On encountering one of three types of plan failures—precondition violation, operator application failure, or failed postcondition—the system designs experiments to find the culprit operator by systematically searching the execution history of the faulty plan to determine when a precondition was destroyed, or to identify the preconditions and postconditions to be added to an operator by systematically searching the different intermediate world descriptions between the present failure and previous successful applications. The two methods are similar in that the design of experiments is guided by a differential analysis—the DEED method analyzes the explanations, and the PRODIGY system analyzes the failed and successful plan execution traces. However, the PRODIGY system uses experiments to determine *how* an incorrect theory must be refined. For example, it designs and conducts experiments to find a faulty operator in a failed plan and, upon finding one, modifies it appropriately (e.g., by adding the missing precondition). In contrast, the DEED method uses a given set of revised theories to determine how a given situation must be modified to construct a discrimination experiment.

Based on earlier work on modeling discovery in humans by Shrager and Klahr (1986, 1987), Klahr et al. (1990) performed psychological experiments in which students were asked to figure out a particular command in a program to control a robot tank (called *BIG TRAK*). They found that the students could explore large experiment spaces efficiently, and some of the heuristics they appeared to be using were as follows: design experiments as different from one another as possible, focus on an aspect of a complex hypothesis, exploit surprises, and decide to validate or discriminate (from rival hypotheses) a hypothesis based upon its likelihood. While their study is broader than ours in scope, some of the experiment-design heuristics are relevant to our research. In particular, in extending our method to include strategies for designing validation and crucial experiments, the last heuristic would provide a control mechanism for selecting an appropriate design strategy for a given hypothesis.

In summary, we described a method for designing experiments to discriminate between two competing theories. The method exploited the differences in the explanations constructed by the competing theories to focus its search of the experiment space. Apart from the examples described in this article, the DEED method has been used to design experiments to discriminate between several other pairs of qualitative theories including theories of combustion, chemical reactions, osmosis, fluid flow, boiling, and dissolving.

Experimentation is a complex scientific activity that includes the design of experiments, the physical construction of the experimental setup, the control of auxiliary parameters within prescribed bounds, the design of suitable measuring devices, and the recording and analysis of measurements. This article addressed only the generation of a design specification for an experiment. Some of the limitations of the method include the inability to use quantitative information, the assumption of a small set of mutually exclusive values for a parameter (inc, dec, std), the inability to cope with non-local interactions of operators, and the inability to discriminate between theories that construct very similar or very different explanations. We expect future research to address many of these limitations.

Acknowledgments

I thank Jan Żytkow and two anonymous reviewers for helpful comments on earlier drafts of this article.

Notes

1. We assume that these different predictions are separable, that is, a measurement can distinguish them. We do not address issues such as noise, measurement accuracy and precision, and the feasibility of measurements; though these issues are important, our emphasis is on developing a basic method for the design of discrimination experiments.
2. In this article, we describe only the bare essentials of QP theory that is needed to explain the experiment design method.
3. Qualitative representations of physical domains have been extensively used by researchers in machine learning and discovery (e.g., Nordhausen & Langley, 1987; Falkenhainer, 1989; Rajamoney, 1989; Karp, 1989; O'Rorke, 1990; Falkenhainer & Rajamoney, 1989; Doyle, 1986).
4. "Amount-of" and "heat" are naive representations of the more complex notions of "mass" and "heat energy."
5. We simplify the explanation figures by omitting some details such as $evap-rate_0 > 0$.
6. In the text and figures, we abbreviate (CL water can) to cl_0 , (CG water can) to cg_0 , ice-block to *ice*, (heat-flow (CL water can) ice wooden-rod) to hf_0 , (rate (heat-flow (CL water can) ice wooden-rod)) to $hf-rate_0$, (evaporation (CL water can)) to $evap_0$, (rate (evaporation (CL water can))) to $evap-rate_0$, wooden-rod to *w-rod*, amount-of to *amt-of*, pressure to *press*, saturation-vapor-pressure to *sv-press*, increasing to *inc*, decreasing to *dec*, and steady to *std*.
7. Explanations E_3 and E_4 have been simplified by removing the qualitative proportionality relating the temperature of the water to its heat, since, according to theories T_{H3} and T_{H4} , the water's heat is constant for the given situation.
8. We assume that the hypotheses H_1 and H_2 are mutually exclusive.
9. If there are several instances of H_2 , a situational support common to all of them is selected.
10. Notice that E_2 must necessarily involve H_2 , since, otherwise, both theories will construct identical explanations.
11. Note, however, that the fluid theory described in figure 1A does not consider the thermal consequences of liquid flow.
12. The figures are for illustrative purposes only. The system actually generates specifications of experimental situations such as the one shown in figure 1C.
13. The removal of the other support is handled in a similar fashion.

References

- Carbonell, J., & Gil, Y. (1987). Learning by experimentation. *Proceedings of the Fourth International Workshop on Machine Learning* (pp. 256–266). Irvine, CA: Morgan Kaufmann.
- Cheng, P. (1991). Modelling experiments in scientific discovery. *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*. Sydney, Australia: Morgan Kaufmann.
- Doyle, R. (1986). Constructing and refining causal explanations from an inconsistent domain theory. *Proceedings of the National Conference on Artificial Intelligence* (pp. 538–544). Philadelphia, PA: Morgan Kaufmann.
- Falkenhainer, B. (1989). *Learning from physical analogies: A study in the analogy and the explanation process*. Doctoral dissertation, Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana.
- Falkenhainer, B., & Rajamoney, S. (1988). The interdependencies of theory formation, revision, and experimentation. *Proceedings of the Fifth International Conference on Machine Learning* (pp. 353–366). Ann Arbor, MI: Morgan Kaufmann.
- Forbus, K. (1984). Qualitative process theory. *Artificial Intelligence*, 24, 85–168.

- Karp, P. (1989). *Hypothesis formation and qualitative reasoning in molecular biology*. Doctoral dissertation, Department of Computer Science, Stanford University, Stanford.
- Klahr, D., Dunbar, K., & Fay, A. (1990). Designing good experiments to test bad hypotheses. In J. Shrager & P. Langley (Eds.), *Computational models of scientific discovery and theory formation*. San Mateo, CA: Morgan Kaufmann.
- Kulkarni, D., & Simon, H. (1990). Experimentation in machine discovery. In J. Shrager & P. Langley (Eds.), *Computational models of scientific discovery and theory formation*. San Mateo, CA: Morgan Kaufmann.
- Langley, P., Zytkow, J.M., Simon, H.A., & Bradshaw, G.L. (1986). The search for regularity: Four aspects of scientific discovery. In R.S. Michalski, J.G. Carbonell, & T.M. Mitchell (Eds.), *Machine learning*, (Vol. 2). Los Altos, CA: Morgan Kaufmann.
- Minton, S., Carbonell, J., Etzioni, O., Knoblock, C., & Kuokka, D. (1987). Acquiring search control rules: Explanation-based learning in the PRODIGY system. *Proceedings of the Fourth International Workshop on Machine Learning* (pp. 122-133). Irvine, CA: Morgan Kaufmann.
- Nordhausen B., & Langley, P. (1987). Towards an integrated discovery system. *Proceedings of the Tenth International Joint Conference on Artificial Intelligence* (pp. 198-200). Milan, Italy: Morgan Kaufmann.
- O'Rorke, P., Schulenburg, D., & Morris, S. (1990). Theory formation by abduction: Initial results of a case study based on the chemical revolution. In J. Shrager & P. Langley (Eds.), *Computational models of scientific discovery and theory formation*. San Mateo, CA: Morgan Kaufmann.
- Rajamoney, S., & DeJong, G. (1988). Active explanation reduction: An approach to the multiple explanations problem. *Proceedings of the Fifth International Conference on Machine Learning* (pp. 242-255). Ann Arbor, MI: Morgan Kaufmann.
- Rajamoney, S. (1989). *Explanation-based theory revision: An approach to the problems of incomplete and incorrect theories*. Doctoral dissertation, Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana.
- Shrager, J. (1987). Theory change via view application in instructionless learning. *Machine Learning*, 2, 247-276.
- Shrager, J., & Klahr, D. (1986). Instructionless learning about a complex device: The paradigm and observations. *International Journal of Man-Machine Studies*, 25, 153-189.
- Zytkow, J. (1987). Combining many searches in the FAHRENHEIT discovery system. *Proceedings of the Fourth International Workshop on Machine Learning* (pp. 281-287). Irvine, CA: Morgan Kaufmann.

Received December 12, 1990

Final Manuscript July 16, 1992