Operational definition refinement: a discovery process*

Jan M. Żytkow Jieming Zhu Robert Zembowicz
Department of Computer Science
Wichita State University, Wichita, KS 67208

Abstract

Operational definitions link scientific attributes to experimental situations, prescribing for the experimenter the actions and measurements needed to measure or control attribute values. While very important in real science, operational procedures have been neglected in machine discovery. We argue that in the preparatory stage of the empirical discovery process each operational definition must be adjusted to the experimental task at hand. This is done in the interest of error reduction and repeatability of measurements. Both small error and high repeatability are instrumental in theory formation. We demonstrate that operational procedure refinement is a discovery process that resembles the discovery of scientific laws. We demonstrate how the discovery task can be reduced to an application of the FAHRENHEIT discovery system. A new type of independent variables, the experiment refinement variables, have been introduced to make the application of FAHRENHEIT theoretically valid. This new extension to FAHRENHEIT uses simple operational procedures, as well as the system's experimentation and theory formation capabilities to collect real data in a science laboratory and to build theories of error and repeatability that are used to refine the operational procedures. We present the application of FAHRENHEIT in the context of dispensing liquids in a chemistry laboratory.

Real world discovery

Many tasks must be solved before we create an autonomous robot that can compete with an experimental scientist in real world discoveries. In this paper we concentrate on generation of high quality data; that is, on the details of procedures which lead to scientific measurements. Experimentation strategies developed by research on machine discovery (Falkenhainer & Rajamoney 1987, Zytkow & Langley 1990, Kulkarni & Simon 1988, Sleeman et al. 1989) do not concentrate on measurement and manipulation details, assuming that high quality data are obtained from a user or a simulation. In true science, however, the requests for data are based on empirical knowledge. For instance, measurements must be made at concrete times. If a measurement is made too early, before a reaction is completed or before a measuring device stabilizes, the results will be systematically wrong or the error may be large. If the measurements are made too late, the interference of a disturbing phenomenon may cause another systematic error. The exact timing of "too early and too late" may vary. A reaction may be completed at different times for varying amounts of chemicals. Similarly, the time after which a measuring device will stabilize may depend on the mass added and other circumstances. The knowledge of repeatability condition may take on the form of a complex theory.

Another problem is the knowledge of experimental error. Without this, it is difficult to evaluate theoretical findings such as empirical equations. Although many systems (BACON (Langley et al. 1987), ABACUS (Falkenhainer & Michalski 1986), and IDS (Nordhausen & Langley 1990)) include some error-related parameters, they disregard the true scientist's problems of error analysis (Żytkow, Zhu, & Hussam 1990, Zembowicz & Żytkow 1991). The error should be known for each data point used in theory generation. This may require a very large number of experiments. A better solution would be to develop a theory of error for a given measurement method.

When an error is large and data is not repeatable, the theoretical component of an autonomous discoverer would experience various difficulties and misfortunes trying to build theories from such data. Error and repeatability conditions are interrelated. Small error can be achieved under restrictive repeatability conditions.

We demonstrate how the theory of error can be determined and the repeatability conditions derived by the use of the discovery system FAHRENHEIT. None of these skills were available in previous machine discovery systems. We show how FAHRENHEIT can be adapted to carry out this task.

We will now discuss operational procedures and their refinement topics not addressed in previous pub-

*The work described in this paper was supported by Office of Naval Research under the grants No. N00014-88-K-0226 and N00014-90-J-1603.

lications in machine discovery. Two earlier papers (Żytkow, Zhu, & Hussam 1990, 1990a) reported the applications of FAHRENHEIT to discovery of a theory helpful in developing analytical instrumentation in chemistry, but the operational procedures were hard-coded into the experiment control module, and all their elements have been pre-programmed.

Interaction with the real world: operational procedures

Each independent and dependent variable must be linked to manipulations (actions) and measurements available to the experimenter. They are used to set the values of independent variables and to measure values of dependent variables obtained in response. Scientifically meaningful manipulations and measurements can be reconstructed in the form of operational definitions (Bridgman 1927, Carnap 1936). Although each scientific concept is typically defined by a coherent set of operational definitions (Żytkow 1984), in this paper we will limit our attention to a single definition of one concept. Each operational definition can be viewed as a program, the elementary instructions for which are primitive actions, instrument readings, and calculations. For instance, a procedure that defines mass difference caused by a particular action can be expressed as:

```
Mass-change(by action a) =
   Mb =: measure mass before action a
   Ma =: measure mass after action a
   Ma - Mb
```

This procedure, although defining a theoretically sound concept, requires adjustments to a particular application: the constraints on the time before and the time after, as well as the error of the procedure for a particular experimental situation, to which it is going to be applied.

We claim that each operational definition must be adjusted to the specific details of a given experiment, in preparation for the empirical discovery. The adjustment is based on a discovery process, that resembles the discovery of scientific laws. It involves the analysis of error and the determination of conditions under which experiments are repeatable.

Measurement refinement: a case study

Mass transfer experiment

As an example, let us consider an experiment illustrated by Figure 1, in which an electronic buret dispenses water in the amount controlled by FAHRENHEIT. Water is transferred to a beaker, which has been placed on a balance. Volume dispensed by the buret is the independent variable, while mass at the balance is dependent. Our task is to determine error and repeatability conditions for the Mass-change procedure provided in previous section. The experiments, even

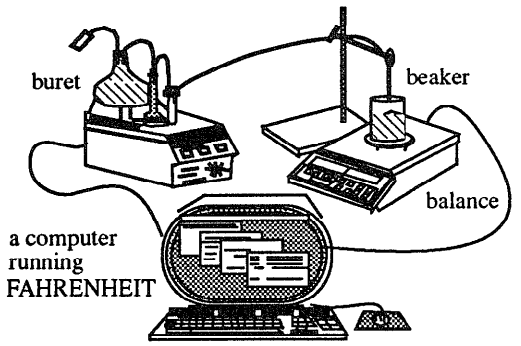

Figure 1. The liquid transfer experiment setup. The buret delivers 10ml water to the beaker at a time.

if identical from the perspective of independent variables used in theory construction, may yield different results. To analyze the differences, in all repeatability experiments, two additional independent variables can be used: time of measurement, and the ordering number of the experiment. In our example this is the number of the repeated transfer of the same amount of water and the time at which the balance measures the mass. We will concentrate on the error and repeatability of mass measurements for a single value of the volume dispensed by the buret.

At each time, the beaker is represented by the state, including:

```
substance:  values: empty, water, and the like
mass:       the reading of mass
```

The initial state of the beaker can be also controlled, by draining the beaker or by filling it with a particular amount of liquid.

Data collection and analysis

FAHRENHEIT began the repeatability and error experiment by setting three independent variables at a constant value: substance in the beaker to empty, dispensing volume to 10mL, number of dispensing to zero, while varying time from 0 to 120, by sampling the balance at the rate of about 1.25 readings per second. The first sequence of measurements thus consisted of weighing the empty beaker approximately 120 times.

As shown in Table 1, this sequence of balance readings has been recognized by FAHRENHEIT as a constant, and the standard deviation has been found. The error was not available for the first sequence of data. This is a special application of FAHRENHEIT's Equation Finder (Zembowicz & Żytkow 1991). While the error is assumed to be constant, the value of error is determined as the standard deviation of the best fit within a small set of simple equations. If the set of data is large, while the equations are simple, there is no risk to introduce overfit, which would result in an underestimated error. The initial value of error can be

time (second)	0	1	2	98	99	100
balance reading (gram)	223.827	223.827	223.827	223.829	223.829	223.829
discovered regularity: constant, m = 223.827, error: 0.002						

Table 1. Partial list of the first 120 balance readings (made in 100 seconds), the discovered regularity and error.

number of mass transfers	measurement period (in seconds) start time	end time	partition point	discovered regularity for stable readings m: mass in grams, t: time in seconds in parentheses, the error of the slope	
1	0	124	10	$m = -0.000213t + 233.685$	(5.6×10^{-6})
2	0	124	9	$m = -0.000203t + 243.517$	(5.5×10^{-6})
3	0	126	9	$m = -0.000242t + 253.454$	(5.5×10^{-6})
4	0	123	9	$m = -0.000229t + 263.398$	(5.6×10^{-6})
5	0	124	9	$m = -0.000240t + 273.339$	(5.6×10^{-6})
6	0	126	10	$m = -0.000210t + 283.286$	(5.5×10^{-6})
7	0	123	9	$m = -0.000229t + 293.235$	(5.6×10^{-6})
8	0	123	9	$m = -0.000212t + 303.182$	(5.6×10^{-6})
9	0	126	10	$m = -0.000202t + 313.131$	(5.5×10^{-6})
10	0	125	11	$m = -0.000236t + 323.082$	(5.6×10^{-6})
regularity found for partition points: $t = 9.5$, with $error = 0.6$					
regularity found for slopes: $t = -0.000222$, with $error = 9.6 \times 10^{-6}$					

Table 2. For each dispensing process, the start point of stable readings is detected (partition point) and the regularity for those readings is found (the error of the slope for each linear regularity in parentheses).

changed at a later time, if in similar but better controlled circumstances, a smaller value of error is detected. In our experiments, the value of error has been determined as 0.002, which is almost the minimum discernible value for our balance and never changed. This value has been used repeatedly as the error, when analyzing the future readings. Repeatability conditions are always satisfied by a constant, so FAHRENHEIT finds this experiment completely repeatable, without limiting conditions.

The second variable varied by FAHRENHEIT is the ordering number of dispensing. 10mL of water is dispensed each time from the buret to the beaker, roughly every 120 seconds. This dispensing process has been repeated 10 times, while balance readings were sampled continuously at the rate of about 1.25 readings per second, leading to that approximate 10×120 readings have been collected (Table 2 and Figure 2).

Each sequence of data, one at a time, is analyzed by the Equation Finder. Because the error is small, no sequence of data, as depicted in Figure 2, fits a single equation. The data partitioning mechanism is therefore invoked, finding a unique special point in each sequence (see Table 2). The initial steep raising line corresponds to the mass change during the dispensing. Each dispensing lasted about 10 seconds. The second sequence of data corresponds to a far more stable situation following each dispensing. No equation within error is detected for the first partition, while a linear equation is found for each second partition of data (Table 2), with a small error in the slope. Negative slopes indicate the decreasing mass of the beaker, which, as we know, is due to liquid evaporation. Because repeatability is satisfied only by a constant, now the data are searched incrementally from the partition point, using the initially determined error, to detect the scope of constancy. Because the slope is very small, given the error of mass as 0.002g, for about 20-30 seconds following each liquid transfer, the mass readings fit a constant. The lower bound of the constant is the partition point, while the upper bound is approximately 20-30 seconds later.

A similar constancy is found for the data immediately preceding the dispensing. Now the operational definition can use these constraints, so that the difference between the mass at the point of the first stabilized reading after dispensing and the mass of the last stabilized reading before dispensing is computed (Table 3). When FAHRENHEIT generalized the location of each partition point, the constant has been detected at 9.5 seconds with an error of 0.6 second.

Both the experiments and the theoretical derivations have been fully automated.

Theory of repeatability and error

Now that we have described details of our experiment, let us summarize the theory of repeatability and error, and the mechanism by which the initial operational definition controls the experiments.

dispensing #	1	2	3	4	5	6	7	8	9	10
vol. dispensed	9.857	9.854	9.961	9.969	9.964	9.973	9.975	9.973	9.973	9.971
discovered regularity: constant, m = 9.947, error = 0.03										

Table 3. List of ten actual dispensed volumes, their regularity and standard deviation.

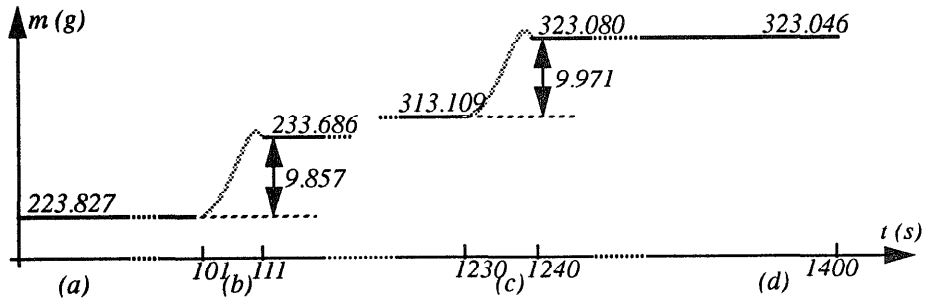

Figure 2. (a): mass of the empty beaker is constant, error is very small; (b) and (c): as water is added, mass increases. The readings stabilize after about 10 seconds; error becomes small. However, a slow long term phenomenon of evaporation (d) can be noticed by comparing the difference in mass: 323.080g at 1240s; 323.046g at 1400s.

Refinements of an operational definition

Each operational definition requires specific actions and measurements. The task is to detect limitations on the time and circumstances in which these actions and measurements are performed. Another task is to determine the error of the procedure. In our example, several constraints on time have been found, and different theories have been found for an empty beaker and a beaker which contains water. Other circumstances, such as the initial mass of the beaker before each dispensing, could be examined also, perhaps leading to more complex theories. In our example, two specific measurements required by the definition are: (1) the mass before dispensing, and (2) the mass after dispensing. The repeatability and error for both guide the refinement process.

The experiment refinement variables

In law discovery, following BACON, FAHRENHEIT distinguished between independent and dependent variables, trying to find a theory that would include all independent variables. In real world applications, the potential independent variables are those associated with manipulation procedures, while potential dependent variables are those associated with measurement procedures.

In repeatability and error analysis, all independent variables can be used to explore the space of circumstances that decrease error and improve repeatability. Variables, which are not involved in the future theory formation task, can still be very important in controlling error and repeatability. For each discovery task, therefore, all variables are divided into two categories (Figure 3). The first category includes variables for which FAHRENHEIT is supposed to build a theory and are called theory formation variables. In our experiment, this first category includes the volumes dispensed by the buret. The second category includes control variables that are of no direct interest to the current theory formation task. We call them experiment refinement variables. Some of these variables may be irrelevant to the measurements, but some of them may influence their results. If that happens for a variable V, FAHRENHEIT finds a constraint on V that limits its influence on the experiment. In our experiment, the second category includes two variables which are always available to the experimenter: the number of times the liquid is added, and the time at which the reading is conducted. If no other variables can be controlled, the last two types of variables can still be varied.

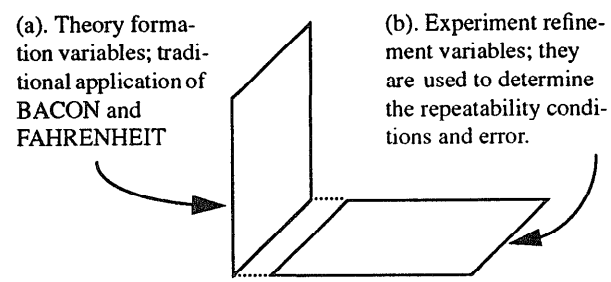

Figure 3. Two types of independent variables. The space of independent variables decomposes into two subtypes: (a). for theory formation; (b). for experiment refinement. Theory of error, which uses experiment refinement variables, must be found before FAHRENHEIT can work on the main theory for the theory formation variables.

The use of special points

FAHRENHEIT includes mechanisms to detect "special points" such as maxima, minima, big changes in slope, and so forth (Żytkow, Zhu & Hussam 1990).

Special points can be used for data partitioning. When a regularity cannot be detected for all data, FAHRENHEIT finds special points and partitions data accordingly. Then FAHRENHEIT can find regularities on boundaries of regularities, and regularities on all types of special points. Both features have been used during repeatability analysis in our example. Special points have been used as a tool for data partitioning, so that the regularity search may succeed in partitions. For instance, a regularity on a boundary of the constant regularity is used to define a repeatability condition.

The use of error

For each new variable V generated by FAHRENHEIT, such as the slope of a particular regularity, or in the boundary of a regularity, the system finds the error of each value of V. Detected errors provide important information about data, and are used to find regularities and special points. For instance, the error reported in Table 1 is used to determine all regularities for water evaporation.

When a sufficient set of values of two variables (independent and dependent) has been collected, FAHRENHEIT calls Equation Finder and then considers the best equation discovered by EF. If the error of the dependent variable is known, EF uses it to evaluate equations and estimate errors of equation parameters. If the error is not known, which is the case at the beginning of the process, EF assumes a model of constant error and estimates for each in a small class of simple equations, the measurement error from the fit. The equation with the smallest error is considered to be the best. In this way FAHRENHEIT obtains not only the best equation, but also the estimation of data error. The lucky coincidence in our experiment has been that the first sequence of measurements was performed on the empty beaker. Since the regularity is a constant for all data, this error is minimal. If the initial sequence of measurements were not so fortuitous, the initial error could be larger. But whenever the system finds a sequence of data that is constant within smaller error, it uses that smaller error in future. This seems intuitively plausible: when we discover that in special circumstances the weighing error is smaller, we change our mind on the error of the device.

Discovery of repeatability and error

In the repeatability study, FAHRENHEIT varies one variable in the category of experiment refinement variables at a time, while keeping all theory formation variables constant. In the liquid transfer experiments, FAHRENHEIT's space of experiment refinement variables includes time, the number of the dispensing, and the empty/non-empty condition of the beaker. The initial experiment is conducted with an empty beaker and zero repetition of dispensing. With the water being added to the beaker, focus is shifted to the "water in beaker" subspace, where the remainder of the balance readings take place. After each successive 10mL of water is added, the readings are collected over a period of time.

After each sequence of data has been collected, FAHRENHEIT tries to find regularities in the data. If the regularity is not constant, the FAHRENHEIT reconsiders the same data, looking exclusively for the range of constancy. The system expects that some data fit a constant within acceptable error and that the rest may either follow another regularity or increase the error. FAHRENHEIT's capability for finding multiple regularities and their boundaries is used on this task. In our approach, only a part of data is repeatable (constant) because the error is very small. FAHRENHEIT searches for a constant which is acceptable within error, and if it cannot find a satisfactory constant for the whole data set, it partitions the data and tries to find a satisfactory constant in each data subset separately. In order to determine whether the fit is satisfactory, FAHRENHEIT uses the minimum value of experimental error or error which has been detected in the first sequence of data.

The dimension of the dry beaker is important because two classes of regularities were discovered: the constant mass for the dry beaker and, the negative linear slope when water is present in the beaker. FAHRENHEIT cannot generalize these two types of regularities, so they are kept separately, and the property of the beaker, empty vs. containing water, is used in the refined procedure. It is always desirable to increase the range of repeatability and to decrease error, but there is no satisfactory universal method that will determine the tradeoff between error and repeatability and decide on the optimal combination of both. The simplest algorithm would compute the average and standard deviation for each sequence of data. This way we can maximize the range of repeatability but we also maximize the error. Usually the error computed this way is unacceptable. The selection of a viable tradeoff depends on real world considerations external to FAHRENHEIT. Scientists may prefer the smallest error within acceptable costs, while in applications we may prefer the largest repeatability range within acceptable error. FAHRENHEIT uses the minimum error of mass for all datasets it has analyzed. In our experiment, the smallest error occurred in the first sequence of data. Then the system finds the range of repeatability within that error.

Operational definition refinement

In our example, an operational procedure to measure the mass dispensed by the buret has been originally defined as:

```
Mass-change(by action a)
 = mass(after action a)-mass(before action a)
```

Following the repeatability and error analysis, that definition can be refined to:

```
Mass-change(by adding x mL of water):
   Mb =: if beaker empty, get mass at any
         time before dispensing;
         if water in beaker, get mass at
         t seconds (t < y) before
         the beginning of dispensing,
   Ma =: mass at t seconds (y+z > t > z)
         after the beginning of dispensing
   Ma - Mb

For x=10mL the error of mass-change is 0.03
```

where x is the value of the independent variable for the buret (so far only the dispensing of 10mL has been examined, but, in the BACON style, the system is ready for a generalization); y is the boundary on repeatability for the beaker which contains water: $y \leq 20 - 5(error) = 15sec$; z is obtained from the measurement of dispensing time. It has been determined that $t > z(z = 9.5 + 0.6(error) = 10sec)$, the constraint generated by dispensing, while the evaporation provides the second constraint $t < y + z$. If the experiments continued for different values of x, it would turn out that there is a regularity (proportionality) between z and x. The new operational definition contains two dependent variables, y and z. They can be treated in the same way as parameters included in the regularities, maxima, and the like. For instance, they can be generalized. For another value of the dispensed volume, y will remain constant, while z will change. Eventually, it can be discovered that $z = a \times x$, where x is the dispensed volume.

If there is leeway left by the repeatability conditions, then the experiment controller, which assigns tasks to device controllers can make its choice of measurement time for the mass reading, within constraints. This helps in experiment planning.

Conclusions

As we have shown by studying a very simple chemistry experiment, FAHRENHEIT can be used to refine an operational procedure by discovering repeatability conditions and experimental error. The refined procedure can be applied within the context of a given experiment. For other experiments, refinements may produce different results. Although the discussion was limited to one particular experiment, the same method can be applied to other laboratory situations. For example, if an acid is dispensed and the pH value is measured, how long should the pH-meter wait to read the pH value, before reagents are fully mixed by the stirrer, and then, how many readings are repeatable before liquid evaporates or another process becomes significant and influences the readings? Waiting until the reagents are fully mixed in pH reading is analogous to waiting until the balance stabilizes in a weight reading.

The presented paradigm for operational definition refinement, is a step towards application of machine discovery systems like FAHRENHEIT to real-world discovery problems. Eventually the results may reach a truly scientific quality.

References

Bridgman, P.W., 1927. *The Logic of Modern Physics.* New York, NY: Macmillan.

Carnap, R. (1936). Testability and Meaning, *Philosophy of Science, 3.*

Falkenhainer, B.C., & Michalski, R.S., 1986. Integrating Quantitative and Qualitative Discovery: The ABACUS System, *Machine Learning, 1*: 367-401.

Falkenhainer, B.C., & Rajamoney, S., 1987. The Interdependencies of Theory Formation, Revision, and Experimentation, *Proceedings of the Fifth International Conference on Machine Learning,* 353-366. Los Altos, CA: Morgan Kaufmann Publ..

Kulkarni, D., & Simon, H.A., 1988. The Processes of Scientific Discovery: The Strategy of Experimentation, *Cognitive Science, 12*: 139-175.

Langley, P.W., Simon, H.A., Bradshaw, G., & Zytkow J.M., 1987. *Scientific Discovery; An Account of the Creative Processes.* Boston, MA: MIT Press.

Langley, P.W., & Żytkow, J.M., 1989. Data-Driven Approaches to Empirical Discovery. *Artificial Intelligence, 40*: 283-312.

Nordhausen, B., & Langley, P., 1990. An Integrated Approach to Empirical Discovery. in: J.Shrager & P. Langley (eds.) *Computational Models of Scientific Discovery and Theory Formation*, 97-128, San Mateo, CA: Morgan Kaufmann Publ..

Sleeman, D.H., Stacey, M.K., Edwards, P., & Gray, N.A.B., 1989. An Architecture for Theory-Driven Scientific Discovery, *Proceedings of EWSL-89.*

Zembowicz, R., Żytkow, J.M., 1991. Automated Discovery of Empirical Equations from Data, *Proceedings of the ISMIS-91 Symposium*, 429-440, Springer-Verlag.

Żytkow, J.M. 1984. Partial Definitions in Science Compared to Meaning Families in Natural Language. *Sign, System and Function: papers of the first and second Polish-American semiotics colloquia*, 479-492, Mouton Publ..

Zytkow, J.M., 1987. Combining many searches in the FAHRENHEIT discovery system. *Proceedings of Fourth International Workshop on Machine Learning*, 281-287, Los Altos, CA: Morgan Kaufmann Publ..

Żytkow, J.M., Zhu, J. & Hussam, A., 1990. Automated Discovery in a Chemistry Laboratory, *Proceedings of the AAAI-90*, 889-894, Menlo Park, CA: AAAI Press.

Żytkow, J.M., Zhu, J. & Hussam, A., 1990a. Determining Repeatability and Error in Experimental Results by a Discovery System. In: Ras Z., Zemankova M., and Emrich M.l., (eds.), *Methodologies for Intelligent Systems 5*, 483-445, New York, NY: Elsevier Science Publ..