Beliefs about sparsity affect causal experimentation

Anna Coenen[1], Neil R. Bramley[1], Azzurra Ruggeri[2], and Todd M. Gureckis[1]
[1]Department of Psychology, NYU, New York; [2]MPRG iSearch, Max Planck Institute for Human Development, Berlin
{anna.coenen, neil.bramley, todd.gureckis}@nyu.edu, ruggeri@mpib-berlin.mpg.de

Abstract

What is the best way of figuring out the structure of a causal system composed of multiple variables? One prominent idea is that learners should manipulate each candidate variable in isolation to avoid confounds (known as the "Control of Variables" strategy). Here, we demonstrate that this strategy is not always the most efficient method for learning. Using an optimal learner model which aims to minimize the number of tests, we show that when a causal system is *sparse*, that is, when the outcome of interest has few or even just one actual cause among the candidate variables, it is more efficient to test multiple variables at once. In a series of behavioral experiments, we then show that people are sensitive to causal sparsity when planning causal experiments.

Keywords: information search; causal learning; hypothesis testing

Introduction

To develop a causal understanding of the world, we often need to find out how multiple candidate variables affect an outcome of interest. This problem arises in everyday situations (e.g., "Which of these switches can turn on the bathroom fan?"), during scientific exploration ("Which of these chemicals affect reaction *x*?"), and plays a role in answering economic and social questions ("How do these policies affect GDP?"). Often, the quickest and most effective method of resolving the causal relationships between variables and outcomes is to conduct experiments that manipulate variables (e.g., turning switches on or off) and help to decouple causation and correlation (Pearl, 2009).

In this paper, we explore how people interact with a novel causal system to understand how it works by manipulating multiple independent variables over a series of trials. We start by introducing two strategies for causal experimentation, one of which is so well known that it is codified in contemporary STEM education standards. We then describe an analysis of an optimal experimenter (i.e., an ideal actor model) which shows how the most informative strategy for learning critically depends on a learner's knowledge about the number of causes among the variables. We then evaluate the predictions of this model in three behavioral experiments.

Test one variable at a time

The problem of disambiguating the effects of multiple variables has a long history in developmental psychology and education. Starting with Piaget (Inhelder & Piaget, 1958), many educators and psychologists have stressed the importance of *controlling or isolating variables*. One important procedural component of this approach is the experimental strategy of changing *one variable* at a time and observing its effect while holding all other variables constant. In the STEM education literature, considerable emphasis has been placed on teaching children this controlling variables strategy (e.g., Chen & Klahr, 1999; Kuhn & Brannock, 1977). In fact, it even appears in national standards for science education (National Academy of Sciences, 2013). A common finding from empirical studies is that children require extensive training to acquire the CV principle (e.g., Kuhn et al., 1995; Klahr, Fay, & Dunbar, 1993; Kuhn & Phelps, 1982). Adults and adolescents, although more likely to use the strategy spontaneously, sometimes also have a tendency to test multiple features at once instead of testing them one-by-one (Kuhn et al., 1995). Interesting exceptions have been found in more complex tasks. For example, Bramley, Dayan, Griffiths, and Lagnado (2017) tested people's intervention strategies in completely unconstrained multivariate systems (with no distinction between potential causes and outcomes) and found that participants often focused on testing one causal relationship at a time by holding most variables at a constant value.

In sum, CV is a widely regarded epistemic principle for learning about causal systems composed of multiple variables. A key advantage of a CV strategy is that it results in unconfounded data that is easy to interpret. Empirical work suggests that acquiring the ability to use the CV principle can be challenging, but adults sometimes adopt it more in complex tasks.

Test half or test multiple variables

Changing variables one-by-one has the benefit of isolating the effect of every variable without confounding influence of the others. It is therefore particularly helpful when one believes that many variables are causes of the outcome. However, consider the case in which a learner expects only very few, and perhaps just a single variable to have a causal relationship to the outcome, but is faced with a number of equally plausible candidate variables. In that case, an alternative strategy is to *test multiple variables* at once to see if any of them affect the outcome at all. For example, imagine trying to figure out which out of 20 switches in a poorly labeled basement fusebox controls the bedroom fan. An optimal strategy for finding this switch is to turn on exactly half (10) of the switches to find out which half contains the target switch and then continue halving the remaining possibilities until only one switch remains. Compared to testing switches one-by-one, this will dramatically reduce the number of trips needed for checking what effect the current switch setting has on the fan.

This *Test Multiple* or (more specifically) *Test Half* strategy has been studied by psychologists in a slightly different type of information-seeking task, often based on popular games as "Twenty questions" or "Guess who?". In these games, children or adults have to identify a target object, person or cause among a given set by asking as few yes/no questions as possible. Here too, the optimal strategy (in terms of expected information gain, see next section) is to ask about features that apply to *half* the possibilities under consideration (e.g., "Is the person female?", if the hypotheses are people and half are each sex), since it can reduce the number of possibilities more rapidly than asking about specific identities directly (e.g., Navarro & Perfors, 2011). Both children and adults have been shown to use this method successfully (Nelson, Divjak, Gudmundsdottir, Martignon, & Meder, 2014; Ruggeri & Lombrozo, 2015). Interestingly, any Test Multiple strategy would be considered an error from the perspective of the education literature (e.g., a student adopting this strategy might be coded as failing an STEM education assessment), because by changing many things at once it momentarily confounds the influence of individual variables.

The Test One and Test Multiple/Half strategies are typically studied in different kinds of psychological tasks. However, as the switch example from above illustrates, they can both be reasonable approaches for testing the causal impact of multiple variables. Next, we show how the effectiveness of each strategy depends on the structure of the task.

Sparsity determines effectiveness of strategies

As the switch example shows, an important factor that determines the effectiveness of a Test One or a Test Multiple strategy is the *sparsity* of a causal system. We define sparsity as the proportion of causes among variables (for related definitions and discussions of the importance of sparsity for hypothesis testing, see e.g., Navarro & Perfors, 2011; Langsford, Hendrickson, Perfors, & Navarro, 2014). In sparse environments (e.g., when we know that only one in 20 switches controls the fan), a learner can quickly narrow in on an effective cause by trying many variables at once. In contrast, if there are (known to be) many causes, trying many things at once will tend to be uninformative as the effect will almost always be generated and little will be learned about which variable(s) were responsible. The choice of an effective testing strategy in a particular situation is thus a question of ecological rationality.

Modeling the effect of sparsity To formalize this intuition, assume that a learner is faced with a simple causal system with N binary independent input variables, I, and a single binary outcome, o. Given the subset of input variables, $C \subseteq I$ that, when active, can cause the outcome to happen, the probability of the outcome given the current setting of inputs is

$$P(o=1) = \begin{cases} 1, & \text{if } \exists\, c \in C\ (c=1), \\ 0, & \text{otherwise} \end{cases}$$

In other words, the outcome occurs if and only if any of the input variables in C are currently active.

The learner must now decide how to manipulate the input variables to best figure out which of them are causes (i.e., which are members of C). We assume that the learner's optimal strategy lies in choosing a switch setting, $s \in S$, that maximizes the expected *Information Gain* with respect to the system. Information gain is a common metric for quantifying the value of information-seeking actions, including causal interventions. It is computed as the expected difference of a learner's current uncertainty over their hypotheses H, and their expected new uncertainty after having made an intervention on the system and observed an outcome. In this case, a learner's hypotheses H are all possible compositions of the set of causes, C, and there are two possible outcomes ($o = 1$ or $o = 0$). Thus a learner's expected information gain is

$$EIG(s|H) = SE(H) - \sum_{j=0}^{1} P(o=j|s)\ SE(H|s), \qquad (1)$$

where SE denotes the Shannon Entropy over a distribution of beliefs (Shannon & Weaver, 1949). To investigate the impact of causal sparsity, we use this model to explore how a learner's belief about sparsity affects the optimal strategy.

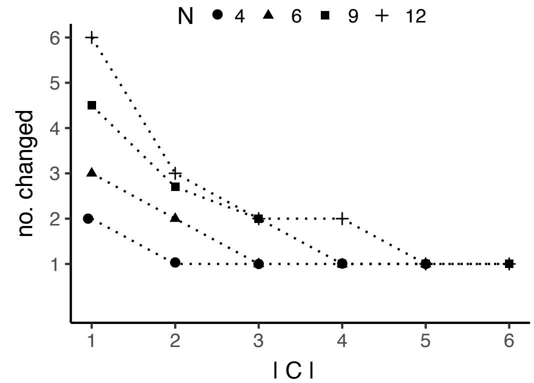

Figure 1: Effect of the number of causes, $|C|$, and the number of variables, N, on the number of variables tested.

Figure 1 shows model predictions for the number of variables an optimal learner should manipulate upon their first encounter with a system of variables, based on their knowledge about the number of causes, $|C|$, and the number of variables in the system, N, assuming a uniform belief over all remaining hypotheses. In line with the intuition outlined above, when a learner expects only one cause the model predicts a Test Half strategy. As the number of causes increases (that is, as causal sparsity decreases), the optimal number of manipulated variables decreases and quickly reaches the strategy of changing only a single variable. This relationship is modulated by the total number of variables, which increases the degree of causal sparsity and consequently the number of variables that should be manipulated.

These results show that the causal sparsity of an environment should affect a learner's strategy for manipulating binary variables to find out how they affect some outcome of interest. This means that, even in the same task, there can exist a continuum of optimal strategies with respect to the num-

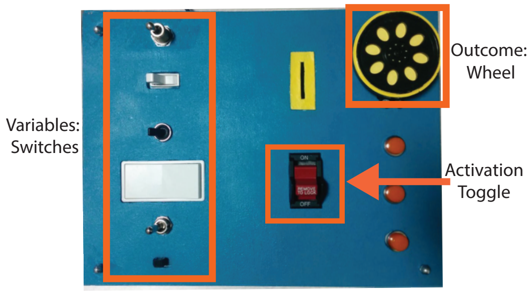

Figure 2: Wooden box used in Experiment 1.

ber of variables changed, which ranges from a Test Half to a Test One method. This observation leads to the core prediction we test in this paper. We hypothesize that when learning a causal system, people will use different strategies depending on their belief about the sparsity of the system. This result would offer a further demonstration that human intervention strategies are ecologically rational, in the sense of being well matched to the environment within which they are needed.

Experiments

We now present three experiments that investigate how knowledge about sparsity affects people's causal testing strategies. Sparsity is manipulated in two ways, both suggested by the model results shown in Figure 1. We first vary the number of causes (i.e., variables that affect the outcome) in a system (Exp. 1 & 2) and second the number of total variables available for testing (Exp. 2). We also investigate what strategies people select given no prior instructions about sparsity (Exp. 3).

Experiment 1 - manipulating number of causes

Participants 30 participants were recruited via the subject pool of New York University's Department of Psychology. Participants were paid at a rate of $5 per hour and could win an additional bonus of up to $3 (see below).

Stimuli Participants were presented with the wooden box depicted in Figure 2. The box had six different switches (inputs), a yellow wheel (output) and a red activation toggle. Each switch could be turned to the left (off) or the right (on). The activation toggle controlled whether the entire box was turned on or off. Participants were randomly assigned to one of two experimental conditions. In the *sparse* condition, only one of the switches caused the wheel to spin, whereas the remaining five switches were broken. In the *non-sparse* condition, five switches caused the wheel to spin and one switch was broken. A single working switch was sufficient to activate the wheel, and the position of the broken switches had no effect whatsoever. The wheel could only be activated if the activation toggle was currently in its on-position. Otherwise, participants were told that the box was turned off. The working and broken switches were chosen randomly for each participant. At the beginning of the experiment, participants were given six plastic tokens, each of which was worth $0.50. Participants had to pay one token every time they wanted to turn on the box via the activation toggle (see below).

Procedure Participants were first familiarized with the components on the box. They were told about the the binary (on/off) nature of the switches, and the difference between broken and working switches. Depending on the condition, participants were then told that they had to to identify the one broken switch (*non-sparse* condition) or the one working switch (*sparse* condition). Before starting the task, participants in both conditions were shown the same two demonstration trials. First, while the activation toggle was turned off, the experimenter turned all six switches to their *on* position and subsequently turned on the activation toggle, causing the wheel to spin. Second, after turning the activation toggle off again, the experimenter set all switches to their *off* state and turned the activation toggle back on, which did not cause the wheel to spin. In the main part of the experiment, participants could repeatedly test different settings of the switches to find out which one was broken/working. On each trial, they could change the switches in any way they liked while the activation toggle was off. They could then test their chosen switch setting by turning the activation toggle on and observing the effect on the wheel. Before the start of each new trial, the activation toggle had to be turned off again.

To incentivize participants to use as few trials as possible, they had to pay one of their six plastic tokens (worth $0.50 each) for each time they performed a test by inserting it into a coin slot on the box. Participants could test the box up to six times (hence the use of six tokens), but could stop whenever they thought they had identified the one broken/working switch. After their final test, they indicated to the experimenter which of the switches was broken/working. If their choice was correct, they could trade in any remaining tokens for their corresponding monetary value. If it was incorrect or they used up all their tokens, they received no bonus.

Results To characterize participants' trial-by-trial behavior at a strategy level, we used the following classification scheme. In the non-sparse condition, participants' strategies were classified as Test One if they turned on one switch on every trial, while leaving all other switches turned off. If a participant manipulated multiple switches or kept testing the same switch more than once, their strategy was classified as Other. In the sparse condition, participants' strategies were classified as Test One if participants turned on one new switch each trial, even when they left previously tested, but ineffective, switches turned on. This is because these past switches would have shown to be broken and therefore could not contribute confounding evidence on future tests. Participants' strategies were classified as Test Multiple if participants tested half or multiple of the switches. As a sequential strategy, Test Half does not have a meaningful definition for participants in the non-sparse group, who would al-

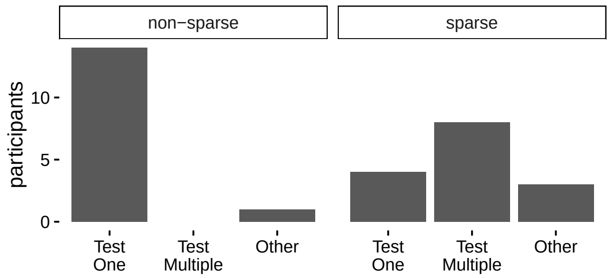

Figure 3: Strategy use in Experiment 1

ways encounter confounding evidence when changing multiple switches. Note that we also classified participants as Test One or Test Multiple if they had some interspersed trials with zero Information Gain (e.g., from repeating the same test twice), assuming that they were using a more noisy version of the respective strategy.[1]

Figure 3 shows the number of participants using each testing strategy in the two conditions. For the purposes of this figure, the classification was based on a participant's sequence of tests up to the point at which an optimal learner would have been able to correctly identify the working or broken switch (some participants made further unnecessary tests). Note that, among participants classified as "Test Multiple", everyone actually manipulated exactly half of the switches (i.e., they used the optimal strategy according to our optimal model). We kept the more general classification as Test Multiple, to stay consistent with the results presented in the next experiment. The number of participants using a Test One strategy was lower in the sparse condition (4 in 15 vs. 14 in 15, Fisher's exact $p < 0.001$). However, even in the sparse condition around a quarter of the participants decided to change one variable at a time.

In sum, as predicted by the optimal learner model presented above, Experiment 1 found that instructing participants to expect either a sparse (one cause) or a non-sparse (five causes) environment, had an effect on how they proceeded to manipulate a set of six variables. However, even in the sparse condition, we found that some use of the less effective Test One strategy persists. The following experiments explore possible explanations for this finding.

Experiment 2 - manipulating number of variables

Experiment 2 explores whether increasing the amount of sparsity by adding more variables would lead to more participants to adopt a Test Half strategy. In Experiment 1, the benefit of testing multiple variables over testing variables one-by-one was relatively modest. In fact, testing half of the variables in the sparse condition would save participants less than one step (2/3 of a step) on average, compared to testing variables individually (this difference translated to an average saving of ∼$0.33). This may have not provided sufficient incentive for participants to realize that a Test Half strategy would be more advantageous. As discussed above, one way to amplify the sparsity manipulation is to add more variables (see Figure 1). To illustrate the effect on the expected payoff from the two strategies, Figure 4 shows the average number of tests needed to find the working switch for a learner in a sparse (one cause) environment employing either a Test One or a Test Half strategy. It shows that as the number of switches increases, so does the benefit of the Test Half strategy over the Test One strategy.

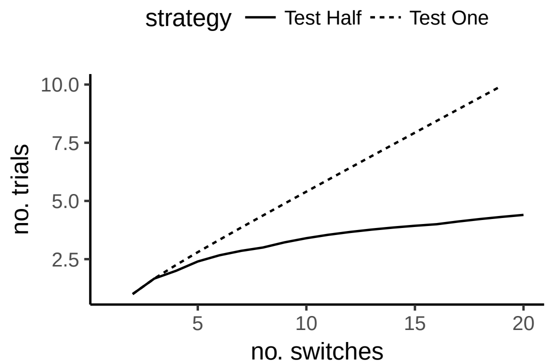

Figure 4: Expected number of trials needed to find the working switch in the sparse condition, when using a Test One or Test Half strategy.

To test if people are sensitive to the degree of sparsity, Experiment 2 manipulated the number of variables (switches). Participants on Amazon Mechanical Turk completed the same task as in Experiment 1, but were presented with either 4, 6, 10, or 20 switches (all manipulations were between-subjects). As before, they were given either sparse (one switch working) or non-sparse (one broken) instructions. Although adding variables should have no effect on behavior in the non-sparse condition, we decided to keep the manipulation to ensure that adding variables does not encourage a general increase in the number of variables participants would test on each trial. By including the 6 switches condition again, this experiment also served to replicate the results from Experiment 1 with an online sample.

Participants 120 participants were recruited on Amazon Mechanical Turk. Recruitment was restricted to AMT workers within the United States aged 18 or above. Participants were paid $0.50 for their participation, with the possibility of earning an additional bonus of up to $1 (see below).

Stimuli The task from Experiment 1 was adapted as faithfully as possible to be run on the web with some minor changes. Instead of a wheel, the outcome of interest was a light bulb, which lit up when it was turned on, and re-

[1] The precise details of the strategy classification had to be omitted from this paper for space reasons, but will appear in a longer version of this manuscript that is currently under review.

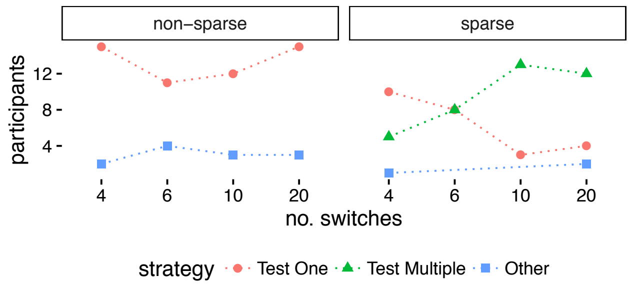

Figure 5: Strategy use in Experiment 2

mained gray otherwise. All switches were of the same kind and would turn green when on and red when off.

Procedure The experiment followed a 4 x 2 between-subjects design. Participants received different versions of the task with either 4, 6, 10, or 20 switches, and were given either the sparse or the non-sparse instructions. The procedure was the same as in Experiment 1. Participants received similar instructions and were also asked to perform two demonstration trials in which first all and then none of the switches were turned on, to show that the light bulb would turn on and stay off, respectively. The per-trial payment was adjusted depending on condition, such that participants had to pay either $0.25, $0.16, $0.1, or $0.05 per additional test in the 4, 6, 10, or 20 switches conditions, respectively. These payments were chosen so that the total potential bonus (starting at $1) would be zero if participants decided to test every single switch in isolation.

Results Figure 5 shows the frequency of the Test One, Test Multiple, and Other strategies by condition. In the non-sparse group a large majority of participants changed a single variable at a time, irrespective of the number of switches. In the sparse condition, however, the proportion of Test One users varied with the number of switches (Fisher's exact, $p < .05$), such that participants confronted with more switches (10 or 20) were less likely to test individual switches than those confronted fewer (4 or 6). This development was also accompanied by the expected increase in strategy efficiency, such that the *number of trials participants saved* on average in the sparse condition compared to the non-sparse condition increased from 0.15 trials in the 4 variable condition to 6.53 trials in the 20 variable condition, in line with the predictions in Figure 4.

These results provide further evidence that information about sparsity affects how people learn actively in multiple-variable settings. Again, participants in the sparse group were more likely to manipulate multiple variables at a time, whereas those with in the non-sparse group chose to manipulate variables one-by-one. Furthermore, participants in the sparse condition, in which there is only one cause, were sensitive to the total number of variables. The more switches were presented to participants (the more sparse the environment), the more prominent was their use of a Test Multiple strategy. Nevertheless, this experiment also replicated the finding from Experiment 1 that in the absence of a *strong* incentive to do otherwise, people have a tendency to change single variables, rather than multiple. In fact, the web sample revealed, if anything, an even stronger tendency to use a Test One strategy in the sparse condition, particularly when the number of switches was small.

Experiment 3 - no sparsity information

Experiments 1 and 2 suggest that testing one variable at a time might serve as a default strategy that is only overridden, to some degree, by knowledge about the number of causes. To explore this possibility further, Experiment 3 asked what strategy people would use to test a multi-variable system when they had no prior information about the number of causes to begin with. By giving participants vague instructions, we aimed to instill an approximately flat "prior" over all possible combinations of working and broken switches.

Participants 57 participants were recruited on Amazon Mechanical Turk. Recruitment was restricted to AMT workers within the United States aged 18 or above. Participants were paid $0.50 for their participation, with the possibility of earning an additional bonus of up to $1.

Stimuli Materials were the same as the 6-switch condition of Experiment 2. In a between-subject design participants were again randomly assigned to a switchboard that either had one broken or one working switch.

Procedure The procedure was the same as in the previous Experiment, with the exception that participants were given the *same* set of instructions about the number of causes in both conditions. Instead of being told to find the one broken or one working switch, they were instructed to "find out which switch(es) are working or broken". After the switch testing phase, participants were asked to indicate which switch(es) were working or broken, now being able to make multiple selections.

Results Figure 6 shows the proportion of participants that chose to turn on any possible number of switches on the very first trial. Data is collapsed over both conditions, since the initial instructions were the same and hence the first trial should not lead to different behaviors. The vast majority of participants (%58) chose to manipulate a single switch, with only a small number (%10) manipulating half.

This experiment verified that with no instructions about sparsity, the majority of participants chose to manipulate variables one-by-one. Note that an optimal learner initialized with a flat prior (which translates into 2^6 hypotheses, given 6 switches, each with a prior probability of $\frac{1}{2^6}$) also assigns higher expected Information Gain to testing one over testing

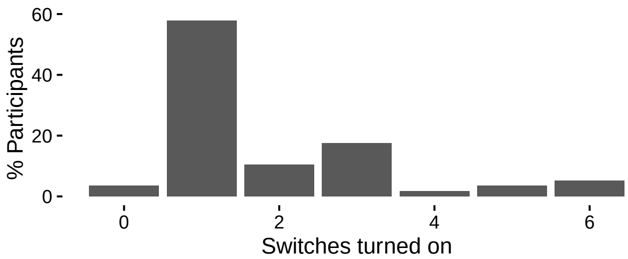

Figure 6: Number of switches tested on the first trial of Experiment 3.

multiple variables. Therefore, the Test One "default" shown by some participants in earlier experiments, could stem from them ignoring the constrained prior that they were instructed on and instead acting as if they knew nothing about the sparsity of the system. This behavior would still be in line with the optimal learner analysis presented above.

General Discussion

In a series of experiments, we found that, in line with an optimal learner model, people's strategies to manipulate multivariate causal systems take into account the causal sparsity of the system. In non-sparse environments (e.g., only one non-cause) the majority of participants adhered to a strategy of testing one variable at a time, in line with a "Controlling Variables" principle (Kuhn & Brannock, 1977). When causes were sparse (e.g., only one cause) participants were more likely to manipulate multiple (often half) of the candidate variables. We also found that increasing the degree of sparsity, by increasing the total number of variables, amplified this effect on people's strategy choices.

These findings demonstrate that people adaptively change their causal experimentation strategies in response to knowledge about the environment. Our study thus offers an example of the importance of "ecological learning" that allows people to flexibly adapt their inquiry strategies to the information structure of the task (Ruggeri & Lombrozo, 2015). This idea tallies with other recent work on causal interventions showing that people's strategy choices were made adaptively with respect to internal constraints, like cognitive load, and external factors like the match of a strategy and the task environment (Coenen, Rehder, & Gureckis, 2015). In finding that sparsity affects behavior, the experiments above also add to recent evidence from other (spatial) information search tasks, in which hypothesis sparsity was shown to affect people's hypothesis testing strategies (Hendrickson, Navarro, & Perfors, 2016).

Interestingly, we also found that even in sparse environments a proportion of participants chose to test variables individually, despite the fact that changing multiple variables would have been more efficient. This is somewhat surprising since prior work has often found that the Controlling Variables principle is difficult to teach and often violated even by adults (Kuhn et al., 1995). It is thus intriguing to think about why we found such pervasive use of a CV strategy. One possibility is that a Test One strategy is less risky than changing multiple variables under a wide range of possible prior beliefs about the system. If the underlying system is not sparse, changing multiple variables can result in ambiguous evidence and often no information gain. However, changing variables one-by-one will be informative even in a sparse environment. With some degree of uncertainty about the current environment, a learner might therefore just be better off testing one variable at a time. Another contributing factor might be the that changing one variable at a time is explicitly taught in schools as a principle of scientific experimentation (National Academy of Sciences, 2013). It is interesting to consider whether this curriculum standard might actually in some cases hinder efficient experimentation by promoting a narrow focus on the idea of testing variables individually, irrespective of situation specifics.

Acknowledgments This research was supported by NSF grant BCS-1255538 and a John S. McDonnell Foundation Scholar Award to TMG.

References

Bramley, N. R., Dayan, P., Griffiths, T. L., & Lagnado, D. A. (2017). Formalizing Neurath's ship: Approximate algorithms for online causal learning. *Psychological Review*, *to appear*.

Chen, Z., & Klahr, D. (1999). All other things being equal: acquisition and transfer of the control of variables strategy. *Child development*, *70*(5), 1098–1120.

Coenen, A., Rehder, B., & Gureckis, T. M. (2015). Strategies to intervene on causal systems are adaptively selected. *Cognitive Psychology*, *79*, 102–133.

Hendrickson, A. T., Navarro, D. J., & Perfors, A. (2016). Sensitivity to hypothesis size during information search. *Decision*, *3*(1), 62.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking*. New York: Basic Books.

Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive psychology*, *25*(1), 111–146.

Kuhn, D., & Brannock, J. (1977). Development of the isolation of variables scheme in experimental and "natural experiment" contexts. *Developmental Psychology*, *13*(1), 9–14.

Kuhn, D., Garcia-Mila, M., Zohar, A., Andersen, C., White, S. H., Klahr, D., & Carver, S. M. (1995). Strategies of knowledge acquisition. *Monographs of the society for research in child development*, i–157.

Kuhn, D., & Phelps, E. (1982). The development of problem-solving strategies. *Advances in child development and behavior*, *17*, 1–44.

Langsford, S., Hendrickson, A., Perfors, A., & Navarro, D. J. (2014). People are sensitive to hypothesis sparsity during category discrimination..

National Academy of Sciences. (2013). *Next generation science standards: For states, by states*. Washington, DC: The National Academies Press.

Navarro, D. J., & Perfors, A. F. (2011). Hypothesis generation, sparse categories, and the positive test strategy. *Psychological review*, *118*(1), 120–134.

Nelson, J. D., Divjak, B., Gudmundsdottir, G., Martignon, L. F., & Meder, B. (2014). Children's sequential information search is sensitive to environmental probabilities. *Cognition*, *130*(1), 74–80.

Pearl, J. (2009). *Causality*. Cambridge university press.

Ruggeri, A., & Lombrozo, T. (2015). Children adapt their questions to achieve efficient search. *Cognition*, *143*, 203–216.

Shannon, C. E., & Weaver, W. (1949). The mathematical theory of information.