How Causal Mechanism and Autocorrelation Beliefs Influence Information Search

Benjamin Margolin Rottman (rottman@pitt.edu)
Learning Research and Development Center, 3939 O'Hara St
Pittsburgh, PA 15260 USA

Abstract

When testing which of multiple causes (e.g., medicines) works the best, the testing sequence has important implications for the validity of the final judgment. Trying one cause for a period of time is important if the cause has tolerance, sensitization, delay, or carryover effects (TSDC). Alternating between the causes is important in autocorrelated environments – when the outcome naturally comes and goes in waves. Across two studies, participants' beliefs about TSDC influenced the amount of alternating; however, their beliefs about autocorrelation had a very modest effect on the testing strategy. This research helps chart how well people adapt to various environments in order to optimize learning, and it suggests that in situations with no TSDC effects and high autocorrelation, people may not alternate enough.

Keywords: causal reasoning, information search, dynamic environments

Introduction

Information Gathering and Hypothesis Testing in a Causal Environment

How do people choose which of two causes produces the most desirable outcome by repeatedly testing the two causes? Information gathering and hypothesis testing have been studied extensively in other fields. One particularly close example is how people choose which of two gambles produces the highest reward on average by repeatedly trying the two gambles (Hills & Hertwig, 2010). However, testing causes introduces a variety of complexities that need to be considered in the testing process.

For concreteness, imagine that a patient has chronic back pain and a doctor gives the patient two medicines to try. Each day the patient can try one of the two medicines, and at the end of 14 days he makes a choice about which medicine is better so the doctor write a prescription for the indefinite future. The question of interest is how the patient chooses between the two medicines on each of the 14 days.

One critical feature of causal search is that different causes have different temporal mechanisms. Tolerance is when a cause generally, or a medicine in particular, initially works but has weaker effectiveness with repeated use (e.g., caffeine). Sensitization is when a cause requires repeated use to become effective (e.g., antidepressants). Delay, somewhat similar to sensitization, is when a cause takes time before working. Lastly, carryover is when a cause continues to work even after the cause is stopped.

In the back pain scenario, if the patient believes that the medicines could have tolerance, sensitization, delay, or carryover (hereafter TSDC) effects, it would make sense to perseverate, to try Medicine 1 for a while and then switch to Medicine 2 for a while, rather than alternating between the two. For tolerance and sensitization, perseveration is necessary to see if the effectiveness changes over time. For delay and carryover, switching quickly between the medicines will make it difficult to determine whether the current medicine is responsible for the current level of back pain, whether the previous medicine is having an influence (carryover), or whether the current medicine has not started to work (delay). Trying each medicine for a period of time will give a clearer picture of which medicine is better (Laska, Meisner, & Kushner, 1983).

Another critical feature of the environment is the temporal pattern of the outcome variable (e.g., back pain) over time. In some environments the outcome is autocorrelated. For example, presumably an individual's back pain comes and goes in waves (autocorrelated). In other environments the outcome variable is random from one observation to the next. Though it is theoretically possible for an individual patient's back pain to be random from day to day, more obvious situations of independent and identically distributed (iid, or low autocorrelation) data involve a doctor trying the medicines on 14 separate patients such as in a between-subjects randomized controlled trial.

If the outcome is IID, the order in which the two medicines are tested does not matter. However, in autocorrelated environments, it is problematic to try Cause 1 for a long period of time before trying Cause 2 for a long period of time. Consider trying Medicine 1 for 7 days, and then Medicine 2 for 7 days, and the pain generally decreases over the 14 days. It is not possible to know if the decrease is due to the change in medicine or due to an underlying temporal trend because they are confounded. Instead, alternating between the medicines would distinguish the medicines and the underlying trend, leading to more accurate inference.

In sum, increased plausibility of TSDC effects should lead a learner to *perseverate* more. Believing that the baseline trend comes and goes in waves (is autocorrelated) should lead a learner to *alternate* more. There is existing evidence that people are able to detect more vs. less autocorrelation (Lopes & Oden, 1987), detect whether a causal mechanism is exhibiting tolerance or sensitization (Rottman & Ahn, 2009), and use knowledge about delay when interpreting time-series data for making causal inferences (e.g., Hagmayer & Waldmann, 2002). However, it is not known whether people are able to make use of knowledge about TSDC and autocorrelation when *planning* a testing strategy to optimize learning rather than just interpreting pre-existing data.

Motivation of Current Experiments

A previous set of studies using the medicine for back pain scenario found that most participants tended to perseverate, though about 7% of participants alternated consistently (Rottman, 2014). In this task the underlying back pain function was autocorrelated, which caused participants who perseverated to have very high error rates, grossly over or underestimating the actual difference in the effectiveness of the two medicines. For example, if the baseline pain trend increased over time, participants often concluded that the first medicine worked much better than the second. Participants who alternated, either by choice or by instruction, were much more accurate. Furthermore, perseveration vs. alternation did not make a difference when the underlying function was random from day to day.

Given that perseveration caused worse performance, why did most participants perseverate? One potential reason is that they were worried about TSDC effects and wanted to give each medicine enough time to exhibit these effects. A second reason is that they thought that back pain was random from day to day, in which case alternation would not be necessary. The current studies test whether people have the foresight to choose appropriate search strategies based on their beliefs about autocorrelation and TSDC.

There are at least two other factors that may influence this search task. If participants really imagine themselves as the patient in the scenario, they might try to test the medicine that they think is currently working the best ("exploiting"). This could lead to just a couple switches between the medicine rather than frequent alternating. A very similar strategy to exploiting is some sort of positive test strategy – to keep on testing the medicine that one thinks is working better because one erroneously thinks that testing this medicine is the best way to figure out which of the two medicines actually works best. Exploiting and positive testing are hard to empirically disentangle.

The following two studies test whether people are able to use their beliefs about autocorrelation and TSDC effects to choose more optimal search strategies. In Study 1, I approached this question by creating cover stories for which participants had different pre-existing beliefs about autocorrelation and TSDC to see if they are able to make use of these beliefs. In Study 2, I directly manipulated participants' beliefs about TSDC and autocorrelation. Both of these studies have strengths and weaknesses. The strength of Study 2 is that it has a high degree control. However, the weakness is that by manipulating people's beliefs explicitly it cannot assess how people behave in situations for which their beliefs about TSDC and autocorrelation are internally generated. In addition, Experiment 2 used 15 different cover stories to examine information search across a variety of situations for external validity.

Study 1

The purpose of Study 1 was to test whether background information that participants have about TSDC effects and autocorrelation influences their information search patterns.

Methods

Participants There were 300 participants from MTurk, about 20 participants per cover story. They were paid $1 for about 6-8 minutes, with the possibility of a bonus.

Stimuli and Design There were three conditions and five cover stories per condition (Table 1). Condition A was the base case with high autocorrelation and low TSDC. A prototypical cover story is choosing which of two gas stations has cheaper prices after visiting one or the other for 14 weeks. Presumably participants would realize that gas prices fluctuate in waves over time. However, going to one gas station on a given week should have no effect on the gas prices the next week (no opportunity for TSDC effects).

Condition B had stories with high autocorrelation and high TSDC. A prototypical example is testing two medicines on their effects on back pain within one patient over 14 days. Here TSDC effects are plausible, and back pain is likely to come and go in waves even without taking any medicine. Comparing Conditions A and B tests whether beliefs about TSDC have an influence on search strategy

Table 1: Summary of Cover Stories in Study 1.

Condition A – TSDC Low, Autocorrelation High: This is the base case to which the other two conditions can be directly compared. The stories are 1) choosing which gas station has cheaper prices after visiting one or the other for 14 weeks, 2) determining which grocery store has cheaper blueberry prices after going to one or the other for 14 weeks, 3) deciding which location to install a solar panel by testing how much electricity it generates in one or the other location over 14 days, 4) deciding on a new deodorant by trying one or the other for 14 days, 5) choosing the faster route to work by trying one or the other for 14 days.
Condition B – <u>TSDC High</u> Autocorrelation High: Participants tested two treatments on one patient over 14 consecutive days to figure out which treatment worked better. The causes could potentially have TSDC effects. The scenarios involve testing which of two medicines works better to relieve 1) back pain or 2) allergies, 3) testing two brands of vitamin supplement to increase vitamin D, 4) testing psychological reward vs. punishment to reduce thumb sucking in a child, and 5) testing yoga vs. meditation to improve mood.
Condition C - TSDC Low, **<u>Autocorrelation Low:</u>** Each of the 14 observations should be viewed as independent of the prior observation. The scenarios involve 1) choosing between two instant lottery games on 14 consecutive days to figure out which one has the higher payoff, 2) having 14 consecutive restaurant customers judge one of two teas before deciding which tea to buy for future customers 3) choosing which of two pain medicines works better by testing them on 14 different patients, 4) choosing whether reward vs. punishment works better to reduce thumb sucking in 14 children, 5) choosing whether yoga vs. meditation improves mood more in 14 separate patients.

Condition C had stories with low autocorrelation and low TSDC effects. A prototypical example is a doctor testing two back pain medicines on 14 sequential patients (each patient gets only one medicine once). There should be no autocorrelation or TSDC effects across 14 patients because there is no plausible way that one patient's pain level or medicine should influence another patient's pain level. Comparing Conditions A and C tests for an influence of autocorrelation beliefs on the testing strategy.

There is no fourth condition because it is difficult to conceive of situations in which each observation is independent from the previous one (low autocorrelation) yet an intervention at one time could have some TSDC effect at a later time. It is not that such a case is impossible (see Experiment 2), but that it would be hard to devise a natural situation that participants would confidently interpret as having low autocorrelation and high TSDC effects.

There were 3 reasons for having 5 stories per condition. First, if only one story was used per condition, any differences between condition could be due to the different story. Thus I took the approach of sampling from a broader range. Second, using a variety of stories introduces variability in the cover stories within and across conditions (e.g., degree of belief in autocorrelation), which is useful for correlational analyses. Third, the cover stories allow for a degree of external validity not typically afforded to many reasoning studies.

Procedures and Manipulation Checks Participants were randomly assigned to one of the 15 cover stories. After reading the story they answered two questions about whether the outcome (e.g., pain, mood) was autocorrelated or not. Two measures were used because there are no validated instruments about autocorrelation beliefs, and autocorrelation beliefs can be queried multiple ways. Question 1 asked whether the outcome scores were closely related (9) to the prior observation or not (1). Question 2 showed participants a graph with low, medium, and high autocorrelation and participants judged which graph reflected their beliefs about the outcome on a 1-9 scale. Even though the two measures were not strongly correlated, r=.27, p<.001, they behave similarly for all the analyses, so they are averaged for simplicity. The manipulation worked as intended. Participants believed that autocorrelation was higher in Condition A (M=5.4, SD=1.8) than C (M=3.5, SD=2.0), $t(202)$=7.01, p<.001, d=.98, and there was no difference between A and B (M=5.6, SD=1.7), $t(197)$<1.

Then, participants in Condition B were asked to rate whether the causes would have TSDC effects. These four questions were not asked in Conditions A and C. For example, it does not make sense how one patient's medicine would have a tolerance or carryover effect on another patient's back pain (Condition C). Asking participants to make such a judgment could encourage unintended beliefs about the scenario to accommodate the question. The only exception was that these questions were asked of the deodorant story in Condition A; this story was included in Condition A instead of B because it was guessed that deodorant would be viewed to have low TSDC effects.

Beliefs about tolerance, sensitization, delay, and carryover were all significantly (p<=.01) but weakly (rs in the range of .23-.27) correlated; the only exception was that tolerance and delay were uncorrelated, r=-.06, p=.53). Even though they were weakly associated, they are all expected to have the same influence on alternation (and indeed they all show the same pattern when analyzed separately), so for conceptual convenience they were averaged. Participants were worried about the possibility of TSDC effects within Condition B; the average rating was 5.12, right at the middle of the 9-point scale, "somewhat likely". Average ratings for individual scenarios ranged from 4.62 to 5.78. The deodorant story had an average rating of 3.35, verifying that it did belong in Condition A.

Next, participants were tasked with figuring out which of the two options produced a better outcome. Participants received 14 sequential choices between the two options. After they chose one option they saw the outcome score (e.g., pain, mood, etc.). When they were ready they made the next choice.

The outcome score after each choice was determined in the following way. There was a baseline function that participants did not know about. One of the options increased the score of the baseline function by exactly 5 points whenever it was chosen, and the other did not change the score from the baseline function. So, at any given choice, one option always worked exactly 5 points better than the other, but participants could not directly experience the 5 point difference because they had to choose between the two options. The outcome scores were given numerically, and disappeared when the next choice was made; they did not see a graphical plot over time.

In Conditions A and B, the baseline function was a compilation of three sine waves with different amplitudes and frequencies. This function is highly autocorrelated and gradually fluctuates in unpredictable waves. In Condition C, 14 observations from the function were sampled, but then randomized so that the data would support the interpretation that the observations were independent, not autocorrelated.

After making the 14 choices participants were instructed to identify the better option (e.g., the higher option for the Vitamin D scenario and the lower option for the back pain scenario). They also rated how much better it was; 5 points was the correct answer counterfactually. Participants knew in advance that they would earn a 20, 15, 10, or 5 cent bonus for a judgment within 2, 4, 6, or 8 points on either side of the correct answer, respectively.

Finally, participants rated the extent to which they exploited and used a positive test strategy. They were asked: "When I thought that one medicine was working better than the other, I would continue to use that medicine"…"in order to reduce my pain during the 14 days" (exploitation) and "in order to figure out whether it really works better or not to choose the best medicine for the future" (positive testing).

Exploitation and Positive Testing had a correlation of .70, and were averaged to create one composite measure (EPT). The reason for asking these questions was to understand why certain participants perseverated. However, there is a challenge in interpreting these sorts of questions in which subjects introspect about their reasons for behaving in a particular way; it is possible that they use the questions to justify their behavior even if it was not actually the cause of the behavior. Participants had higher ratings for exploiting and positive testing in Condition B (M=5.3, SD=2.4) than A (M=4.1, SD=2.8), $t(197)$=3.3, p=.001, and C (M=3.9, SD=2.6), $t(195)$=4.1, p<.001, and no difference between A and C, $t(202)$<1. This difference could be interpreted two ways. First, it could be interpreted as a confound, that there was some inherent difference in Condition B that lead to exploiting and positive testing. Another interpretation is that participants perseverated more in Condition B, and consequently rated these questions higher as justification. The analyses will account for both of these possibilities.

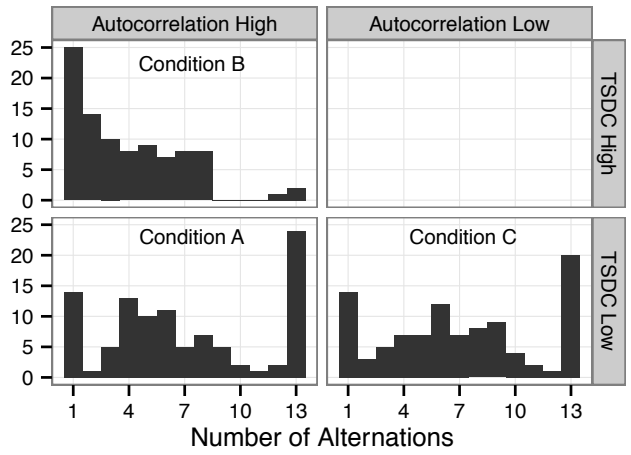

Figure 1: Histograms of Alternation in Study 1.

Results

The dependent variable of interest was the number of times that participants alternated between the two choices. Given that there were 14 choices, there were 13 opportunities to alternate; 9 participants are ignored in future analyses for alternating zero times suggesting disengagement in the task.

Figure 1 shows histograms of the number of alternations. In Condition B, almost all participants alternated less than chance (7 out of 13 possibilities), and there were almost no participants who consistently alternated. The most common strategy was to try one option (often for exactly 7 days), and then try another option.

In contrast, Conditions A and C have trimodal distributions suggesting three strategies: alternating exactly once, exactly 13 times, or something in the middle – alternating roughly randomly or every 2 to 3 observations. Because of the three distinct strategies, some of the following analyses use multinomial regression with three outcome categories: 1, 2-12, and 13.

Accuracy The prior research already demonstrated that alternating reduces error in participants' judgments of which option is better and by how much. This finding was replicated. Within the autocorrelation high conditions (A and B), more alternation is associated with less absolute (logged) error, b=-.11, p<.001, r^2=.15, but within the autocorrelation low condition (C), the number of alternations had no influence on the amount of error, b=-.01, p=.48, and the interaction is significant b=.09, p=.002. This finding highlights the importance of alternating in autocorrelated environments.

Autocorrelation Beliefs Conditions A and C provide the minimal pair to test the influence of autocorrelation beliefs, so only these two conditions are analyzed here. The critical question is whether participants who believed that the outcome was more likely to be autocorrelated alternated more. In Figure 1 there is no obvious difference between Conditions A vs. C. A multinomial logistic regression assessed whether autocorrelation beliefs were associated with different distributions across the three basic strategies (1 vs. 2-12 vs. 13 alternations). The 2-12 strategy was set as the reference, so the regression tests whether autocorrelation beliefs are associated with a change in the ratio of 1 vs. 2-12 alternations, and in the ratio of 13 vs. 2-12 alternations. The regression did not find any influence of participants' beliefs about autocorrelation on alternation: b=.04, p= .73 for 1 vs. 2-12, b=.02 p=.85 for 13 vs. 2-12. To investigate whether the autocorrelation beliefs had an influence on the amount of alternations within the range of 2-12 (n=127), a Gaussian regression was run within this subset of data. There was a marginal positive effect of beliefs about autocorrelation on number of alternations, b=.17, p=.10; if there is any effect of autocorrelation beliefs is very small, r^2=.02.

TSDC beliefs There is a large difference in the pattern of alternations between Conditions A and B (Figure 1). As already explained, when comparing Conditions A and B there is a possible confound of exploiting and positive testing (EPT). This possible confound was accounted for two ways.

First, multinomial regression was run to predict number of alternations in Condition A vs. B while statistically controlling for EPT. Higher scores on EPT where associated with a higher ratio of 2-12 vs. 1 alternation, b=.39, p<.001. The reason for this increase is that EPT often requires trying Option 1, then trying Option 2, and sometimes switching back to Option 1 if it is judged to be more beneficial; in contrast, a common strategy that does not involve exploiting is trying Option 1 for (roughly) 7 days and then Option 2 for 7 days. EPT was also associated with a higher ratio of 2-12 alternations relative to 13 alternations, b=1.02, p<.001; alternating at every opportunity necessarily means not exploiting. Above and beyond EPT, Condition B (relative to Condition A) was associated with a higher ratio of 1 alternation relative to 2-12, b=.94, p=.03, and was also associated with a higher ratio of 2-12 alternations than 13,

b=1.64, *p*=.05. In sum, this analysis suggests that participants alternated less in Condition B even after controlling for EPT beliefs; a likely reason is the possibility of TSDC effects.

The second way to test of an effect of TSDC beliefs but removing the possible confound of EPT between Conditions A and B, involved testing whether TSDC beliefs were correlated with the amount of alternation only within Condition B (and also including the deodorant story), for a total of 118 participants. Because the distribution for Condition B in Figure 1 does not show the characteristic tri-modal distribution of the other two conditions, a Poisson regression was run (due to the skew). This regression a did not reveal an effect of TSDC, *b*=.02, *p*=.47. These findings do not change if both TSDC and autocorrelation beliefs are simultaneously used as predictors. In sum, there is mixed evidence as to whether TSDC beliefs influence the amount of alternation.

Discussion

Study 1 assessed whether beliefs about autocorrelation and TSDC that participants have going into the search influence the search pattern. It would make sense for high autocorrelation beliefs to lead to many alternations, and for high TSDC beliefs to lead to low alternation.

This study successfully activated beliefs of high vs. low autocorrelation for different scenarios, yet these beliefs had minimal if any influence on participants' alternation patterns. There is mixed evidence on whether TSDC effects had an influence on the testing strategy. There was a significant difference in the search patterns between the high vs. low TSDC stories, even after accounting for a possible confound of exploitation and positive testing. However, beliefs about TSDC within the high TSDC condition had no influence on search strategy.

The strength of Study 1 was that it assessed the influence of participants' own beliefs about TSDC and autocorrelation in a variety of learning challenges, but the weakness is lower control. Study 2 attempted a more controlled and stronger manipulation of TSDC and autocorrelation beliefs to see whether people can use this knowledge, when made more explicit, to choose appropriate search strategies.

Study 2

Methods

Participants There were 201 participants from MTurk, and they were again paid $1 with a bonus of up to 20 cents.

Stimuli and Design For increased control, only the back pain cover story was used. The design was a 2 (TSDC: high vs. low) x 2 (Autocorrelation: high vs. low).

Procedures The procedures were very similar to Study 1 except for the following differences. Beliefs about autocorrelation were manipulated by presenting participants with 14 days of pain scores sequentially before starting to test the 2 medicines for another 14 days. The pain scores (both in the initial 14 days and during the 14 days of testing) were either autocorrelated using the same wave-like function from Study 1, or they were from the same function but randomized across the 28 days just like in the autocorrelation low condition from Study 1. Participants then judged the amount of autocorrelation in the back pain level using the same two measures from Study 1. The two measures were moderately correlated, *r*=.49, *p*<.001, and both received significantly higher scores in the high condition: Question 1, *M*=5.8 vs. 4.0, *t*(196.07)=6.71, *p*<.001, *d*=.95; Question 2, *M*=7.0 vs. 3.0 *t*(198.96)=12.89, *p*<.001, *d*=1.82.

After experiencing the initial 14 days, participants were told that they visit a doctor who tells them about the two medicines that they can test for 14 days. The doctor conveys information about the medicines that manipulates whether TSDC beliefs were high or low. Participants were told that the medicines start to work in 30 minutes (low) vs. 1-2 days (high), that they continue to work for 12 hours (low) vs. 1-2 days (high), and that they either do not (low) vs. may start to work better or worse after repeated use (high). In order to move forward with the study they had to correctly answer three questions about TSDC to verify that participants had read these instructions. The 14 days of testing proceeded the same way as in Study 1.

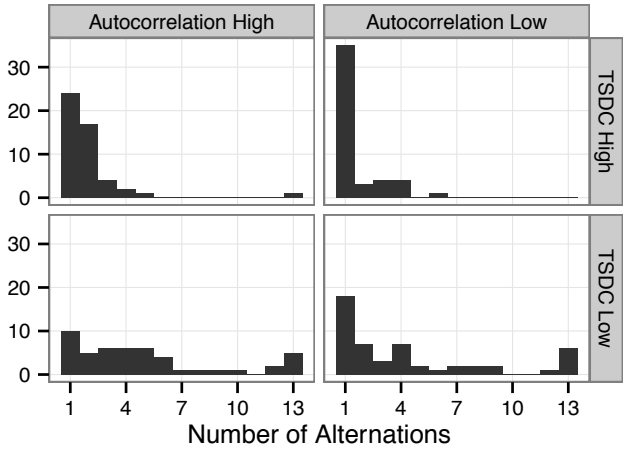

Figure 2: Histograms of Alternation in Study 2.

Results

TSDC and Autocorrelation Beliefs Figure 2 shows histograms of the number of alterations by condition. Casual inspection reveals that 1) most participants across all conditions alternated fairly little, 2) there appears to be considerably more alternation in the low than high TSDC condition, and 3) it is less evident whether the autocorrelation manipulation influenced the test strategy.

A multinomial regression predicting alternation pattern (1 vs. 2-12 vs. 13) was run with autocorrelation and TSDC conditions as predictors. Increasing TSDC beliefs increased the ratio of 1 vs. 2-12 alternations, *b*=1.35, *p*<.001, and (marginally) increased the ratio of 2-12 vs. 13 alternations, *b*=1.86, *p*=.08. Increasing beliefs about the amount of

autocorrelation increased the ratio of 2-12 vs. 1 alternation, b=.95, p=.003, but it did not have an influence on ratio of 2-12 vs. 13 alternations, b=.26, p=.67.

For the sake of robustness, a Poisson regression was also run to model the entire distribution rather than just the three categories of alternation. Higher TSDC beliefs caused less alternation, b=-.97, z=10.80, p<.001. Higher autocorrelation beliefs caused a marginally higher amount of alternation, b=.15, z=1.88, p=.06.

General Discussion

Two studies found that people use knowledge of tolerance, sensitization, delay, and carryover effects (TSDC), particularly when information about such effects was provided explicitly instead of implicitly, for deciding how to test which of two causes produces a better outcome. In contrast, the studies found minimal use of beliefs about autocorrelation in the environment for deciding how to test the causes.

Performance in the TSDC low autocorrelation high condition in Study 2 is especially problematic. In this condition participants have all the information they need that would warrant frequent switching (they know that TSDC effects are not plausible and they know that the pain function has waves), yet there was still no consistent pattern of alternating, which would have produced more accurate judgments and higher payoffs. Two likely culprits are a desire to exploit, and positive testing, though it is very difficult to empirically distinguish these two strategies. Indeed, out of participants in Study 2 who eventually said that one medicine was better than the other (not equal), 40% tested both medicines exactly 7 times, 47% tested the medicine they eventually thought worked better more than 7 times, and only 13% tested the medicine that they eventually thought worked better less than 7 times. Whenever one tests one option more than the other (exploitation) it inherently limits the number of possible alternations. One direction for future research is to examine how people test two causes in a situation for which there is no possibility of exploiting or the possible outcomes are equally desirable.

How will these results, particularly the low rates of alternation, translate into the real world? One plausible hypothesis is in the real world people would want to exploit even more, which could further reduce the probability of choosing the correct medicine. On this topic, it will be important to further probe whether implicit vs. explicit beliefs moderate the extent to which knowledge about TSDC and autocorrelation is used. In Study 2, knowledge about TSDC and autocorrelation were made more explicit, but in many contexts in the real world it would not be so explicit (Study 1).

How do people test two causes when they also have the possibility of trying neither? In the current study participants were forced to try one or the other. Abstaining from either cause, if used appropriately, could help participants distinguish increasing and decreasing trends from TSDC effects.

Another direction for future study is separately examining beliefs about tolerance, sensitization, delay, and carryover effects on the testing strategy. In the current research they were grouped together because they should all have the same influence on the amount of alternation. Additionally, even if they are in fact separable, they are interrelated. For example, sensitization is similar to delay, the difference being that delay is the time of onset after a single dose or intervention, whereas sensitization is an accumulation across doses or interventions. Furthermore, in many situations a learner may have a belief that TSDC effects are plausible, but not have specific beliefs about any of them. Thus, as a first investigation it seemed sensible to group them together. But it will also be important to investigate how people reason about combinations of them.

Lastly, an optimal learner model of autocorrelation and TSDC would help clarify exactly how one should test two causes. However, such a model will be challenging to build because of the many possible functional forms of TSDC.

More broadly, most research on causal learning has investigated situations with low autocorrelation and for which TSDC effects are implausible. Yet, in many causal decisions autocorrelation is high and TSDC effects are plausible. Understanding how people reason in these complex environments is critical for predicting human behavior and providing support or training where people have suboptimal habits, such as physicians and patients working together to identify the optimal medicine.

References

Hagmayer, Y., & Waldmann, M. R. (2002). How temporal assumptions influence causal judgments. *Memory & Cognition*, *30*(7), 1128–37.

Hills, T. T., & Hertwig, R. (2010). Information search in decisions from experience. Do our patterns of sampling foreshadow our decisions? *Psychological Science*, *21*(12), 1787–92. http://doi.org/10.1177/0956797610387443

Laska, E., Meisner, M., & Kushner, H. B. (1983). Optimal crossover designs in the presence of carryover effects. *Biometrics*, *39*(4), 1087–1091.

Lopes, L. L., & Oden, G. C. (1987). Distinguishing Between Random and Nonrandom Events. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *13*(3), 392–400.

Rottman, B. M. (2014). Information Search in an Autocorrelated Causal Learning Environment. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 1321–1328). Austin, TX: Cognitive Science Society.

Rottman, B. M., & Ahn, W. (2009). Causal learning about tolerance and sensitization. *Psychonomic Bulletin & Review*, *16*(6), 1043–9. http://doi.org/10.3758/PBR.16.6.1043