

The influence of temporal distributions on causal induction from tabular data

W. JAMES GREVILLE AND MARC J. BUEHNER
Cardiff University, Cardiff, Wales

Two experiments investigated the joint influence of statistical and temporal information on causal inference from tabular data. Participants were presented with unambiguous data sets containing information about relative effect frequencies in cause-present and cause-absent situations. In addition to contingency information, the stimuli also revealed information about the temporal distribution of effects. The participants took this information into account when making causal judgments, so that the mere advancing or postponing of the effect in time was attached with causal significance, even when the cause did not increase the overall probability of the effect. These results cannot be reconciled with standard contingency accounts of causal induction.

George suffers from recurring bouts of headaches. They typically start in the morning and tend to recede around noon. If George takes aspirin, however, the headaches usually improve by mid-morning. Is aspirin effective in relieving George's headache? Intuitively, one would say *yes*: Clearly, taking the medicine improved George's well-being. One might also point out, however, that the headache would have gone away on its own anyway, so taking the drug clearly made no difference. This article addresses the questions raised by situations such as the one described above.

On what grounds would one decide whether the drug is or is not effective? Psychologists, cognitive scientists, and learning theorists agree that causal induction involves the computation of strength parameters from empirical evidence concerning the presence and absence of putative causes and effects. Exactly how this inference takes place is still a matter of debate (Buehner, Cheng, & Clifford, 2003; Cheng, 1997; Lober & Shanks, 2000; for an overview, see also Shanks, Holyoak, & Medin, 1996), but the cause-effect contingency,

$$\Delta P = P(e|c) - P(e|\neg c),$$

is at the core of most of the approaches. ΔP is the difference between the probability of the effect e in the presence of the cause c and the probability of e in the absence of c . In most models of causal learning, causal judgments are determined, at least in part, by ΔP , so that c is seen to produce e when ΔP is positive but to prevent e when ΔP is negative; when ΔP is zero, c should be judged to have no influence on e . ΔP can be calculated by relating the event frequencies in a 2×2 contingency matrix (see Table 1) to each other:

$$\Delta P = a/(a + b) - c/(c + d).$$

ΔP alone cannot help us to settle George's case, since ΔP only represents contingency and completely ignores another important cue for causation: temporal contiguity (Hume, 1739/1888; see also Young, 1995). Real-world problems, such as George's bouts of headaches, differ dramatically from standard laboratory causal-learning tasks, in that only the latter are carved into neatly defined learning trials, whereas the former present themselves as a continuous flux of events. In the absence of a clearly delineated trial structure, a contingency-based approach requires that the continuous stream of time be divided into discrete binary events entering the contingency table. Thus, in situations involving continuous time, contingency approaches such as ΔP are incapable of representing the nature of the evidence, let alone offering a solution to the inductive problem, unless they are amended with additional assumptions about cause-effect timing.

How Timing Can Influence Causal Induction

Timing determines (subjective) contingency. In a seminal article outlining an associative framework of causal induction, Shanks and Dickinson (1987) suggested that the extent of temporal contiguity between c and e could essentially modulate the perceived cause-effect contingency: An occurrence of c followed by e later on may be encoded as either (1) an instance of Cell a, if the temporal interval between c and e was sufficiently short to be evaluated as *contiguous*, or (2) one instance of Cell b and one instance of Cell c, if the interval exceeded the contiguity threshold. Shanks and Dickinson showed that the same programmed contingency ($\Delta P = .75$) between an action and an outcome gave rise to different ratings of causal effectiveness, depending on the implemented cause-effect delay. More specifically, if action and outcome were separated by more than 2 sec, the participants no longer distin-

M. J. Buehner, buehnerm@cardiff.ac.uk

Table 1
Standard 2 × 2 Contingency Matrix

| | | Effect <i>e</i> | |
|----------------|---------|----------------------|-----------------------|
| | | Present | Absent |
| Cause <i>c</i> | Present | A (<i>c e</i>) | B (<i>c ~e</i>) |
| | Absent | C (<i>~c e</i>) | D (<i>~c ~e</i>) |

Note—Cells A, B, C, and D represent the absolute frequencies of each type of event: Cause and effect, cause and no effect, no cause and effect, no cause and no effect.

guished between contingent and noncontingent conditions (see also Shanks, Pearson, & Dickinson, 1989).¹

In contrast to associationist theories, the rational approach to causal induction (Einhorn & Hogarth, 1986) states that (apparent) lack of contiguity can be overcome if the reasoner knows that the causal mechanism takes time to unfold. More specifically, for a *c* – *e* sequence to give rise to causal impressions, it is essential that the *actually elapsed time* between *c* and *e* should match the *anticipated time frame* of the causal relation in question (Buehner & May, 2002). Accordingly, Buehner and May (2004) showed that causal learning of delayed relations was just as effective as learning of immediate relations, as long as experienced delays were matched with concomitant expectations of a delay (see also Buehner & May, 2002, 2003).

Cause–effect timing—and assumptions about the extent of timing—thus crucially determine the nature of the inference. The same continuous event stream can be parsed in a number of ways, depending on the conception of cause–effect timing that is applied to the task. In our earlier example, ΔP would be considerably positive, if a time frame of *a few hours* were applied: When George takes aspirin, his headaches go away; if he does not, he continues to suffer for several hours. In contrast, ΔP would be zero if the time frame were *on the same day*: Taking aspirin makes no difference as to whether George has headaches or not. In a series of experiments, Hagmayer and Waldmann (2002) have shown that people indeed interpret exactly the same evidence in different ways, depending on their assumptions regarding the time frame of the causal relation in question.

Timing may overshadow other cues for causation. In addition to determining how events get allocated to the four cells of a contingency table, timing may influence causal induction in another important way. In certain situations, the information gained when cause–effect timing is considered is so salient and convincing that other relevant information tends to lose its influence. For a long time, it was thought that evidence of contiguity alone is sufficient to create causal impressions. Mendelson and Shultz (1976), for instance, found that young children were apparently so swayed by contiguity that they completely ignored the absence of contingency when drawing causal inferences.

In a more recent developmental study, Schlottmann (1999) showed that when various pieces of evidence need to be integrated to draw a causal inference, young children focus exclusively on temporal information and ignore considerations of causal mechanism. Buehner and McGregor (2006) have argued that Schlottmann's (1999) contiguity bias is a special case of a *time frame bias*. With an apparatus that was (a priori) more representative of delayed than of immediate causal relations, they found that participants consistently rated delayed relations as causal, even when they knew that the apparatus was set to produce effects immediately. In all these studies, the experience of a particular time frame (be it delayed or immediate) served as such a strong cue for causality that other cues (contingency or considerations of mechanism) were ignored.

An interaction between contingency and contiguity? The previous two sections outlined the modulating role of timing in causal induction: Timing can (1) disambiguate continuous data to segment it into discrete units of evidence suitable for contingency-based approaches and (2) overshadow other cues, such as contingency or mechanism, so that they are essentially ignored. A third possibility that has largely been left unexplored is that variations in contiguity and contingency can interact to influence causal assessment. Note that the disambiguating function of timing (1) also, of course, effectively constitutes an interaction, but one in which timing changes the subjective nature of the contingency. Here, we will explore the possibility of a genuine interaction between cause–effect timing and contingency in situations in which variations in timing cannot determine the extent of (subjective) contingency. None of the studies reviewed in the sections above addressed such situations.

In accumulated longitudinal data of an epidemiological nature, the allocation of individual events to the cells of a contingency table has already taken place (and cannot be altered, as in our initial example, by notions of cause–effect timing). In addition to merely considering whether or not an effect has occurred, reasoners might also take into consideration *when* the effect has occurred after the cause has been administered. Epidemiologists have recognized the importance of temporal distribution within a contingency table: According to the Henle–Koch–Evans postulates for an epidemiological proof of causation, the disease not only has to have a higher incidence in those exposed to the cause, as compared with those not so exposed, but also should temporally “follow exposure to the hypothesized causative agent with the distribution of incubation periods as a log-normal shaped curve” (Evans, 1993, p. 192). To our knowledge, these postulates have not yet been embraced or scrutinized by psychologists. More specifically, no study we are aware of has investigated the influence of temporal distributions within contingency tables. In the remainder of this article, we will explore to what extent lay people are sensitive to such temporal distributions.

To illustrate the point, consider a study in which one group is administered an experimental treatment and another is administered a placebo treatment at time t_1 ; the

groups are then monitored over a number of days for the occurrence of the effect. Records are made not only of *whether* the effect occurred in individuals of both groups, but also of *when* it occurred within the time frame of the study (say, t_1, t_2, \dots, t_n ; see Figure 1 for an example). Thus, both contingency and contiguity information is easily accessible, and the data are completely unambiguous. We were interested to find out to what extent contingency and temporal information interact to influence causal induction from summary data. To that effect, we created stimulus materials that varied with respect to their contingencies (when computed across a focal time frame) and with respect to the distribution of effects over time within that focal time frame. If people attach significance only to contingencies, variations in temporal distributions should make no difference. If participants base their judgments entirely on the level of contiguity between causes and effects, variations in the overall contingency should make no difference. Our prediction, however, was that reasoners are sensitive to both contingency and temporal information; consequently, temporal distributions should

influence causal assessments in concert with variations in contingency.

EXPERIMENT 1

We investigated how statistical and temporal data interact to determine causal judgments from both contingent and noncontingent data sets. We presented participants with information that had accumulated over 5 consecutive days, taking care to explicitly separate information about cause-present and cause-absent cases, so as to render the data completely unambiguous. This was achieved by creating a cover story about a biotech laboratory investigating the effect of radiation on the survival of bacterial cultures. A set of bacterial cultures was exposed to a one-time radiation treatment and then were monitored over the next 5 days, with lab assistants recording whether and when individual cultures died off. One advantage of this cover story was that it was comparatively devoid of a priori notions of causal power and time frames associated with the causal relations of interest. Both immediate and delayed

This study investigated the effect of **Delta**-Rays on **BAC-4** bacteria

BAC-4 bacteria exposed to **Delta** treatment on day 0

| | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 |
|----|-------|-------|-------|-------|-------|
| 1 | | | | | |
| 2 | | | | | |
| 3 | | | | | |
| 4 | | | | | |
| 5 | | | | | |
| 6 | | | | | |
| 7 | | | | | |
| 8 | | | | | |
| 9 | | | | | |
| 10 | | | | | |
| 11 | | | | | |
| 12 | | | | x | |
| 13 | | | | | |
| 14 | | | | | |
| 15 | | | | | |
| 16 | | | | | |
| 17 | | | | | |
| 18 | | | | | x |
| 19 | | | | x | |
| 20 | | | | | |
| 21 | | | | | |
| 22 | | | x | | |
| 23 | | | | x | |
| 24 | | | | | |
| 25 | | | | | x |
| 26 | | | | | |
| 27 | | | | | x |
| 28 | | | | | |
| 29 | | | x | | |
| 30 | | x | | | |
| 31 | | | | | |
| 32 | | | | | |
| 33 | | | | | |
| 34 | | | | | |
| 35 | | | | | |
| 36 | | | | | |
| 37 | | | | | x |
| 38 | | | | | |
| 39 | | | | | |
| 40 | | | | | |

BAC-4 bacteria with **no exposure** to any radiation treatment

| | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 |
|----|-------|-------|-------|-------|-------|
| 1 | | | | | |
| 2 | | | | | |
| 3 | | | | | |
| 4 | | | | | |
| 5 | | | | | |
| 6 | | | | | |
| 7 | | | | | |
| 8 | | | | | |
| 9 | | | | | |
| 10 | x | | | | |
| 11 | | | x | | |
| 12 | | | x | | |
| 13 | | | | | |
| 14 | | | | | |
| 15 | | | | | |
| 16 | | | x | | |
| 17 | | | | | |
| 18 | | x | | | |
| 19 | | | | x | |
| 20 | | | | | |
| 21 | | | | | |
| 22 | | | | | |
| 23 | | | | | |
| 24 | | | | | |
| 25 | | | | | |
| 26 | | | | | |
| 27 | x | | | | |
| 28 | | | | | |
| 29 | | | | | |
| 30 | | | | | |
| 31 | | | | | |
| 32 | | | | | x |
| 33 | | | | | |
| 34 | | | | | |
| 35 | | | | | |
| 36 | | | | | |
| 37 | | | | | |
| 38 | | | | x | |
| 39 | | | | | |
| 40 | | x | | | |

Please rate the influence of Delta Rays on BAC-4 bacteria.

-100 ----- 0 ----- +100

Very effective at killing bacteria/no effect/very effective at promoting survival

Your rating? _____

Are there any other observations you can make about the treatment from the data?

Figure 1. Example stimulus materials for Condition C, Experiment 1.

effects of radiation on the life of organisms fit our intuitive understanding of biotechnology. Instructions further explained that bacteria may, of course, naturally die, irrespective of radiation, and that the effect, if any, of radiation on these cultures is yet unknown. Thus, it was necessary to compare the experimental data sets with control data sets that involved the same strand of bacteria but ones that had not been exposed to any radiation.

Each condition thus consisted of an experimental and a control data set, with each data set containing information about how many out of 40 cultures had died over the course of 5 days and when, within this period, they had died. Contingency information was conveyed through the *relative frequencies* of effects within the experimental and the control groups, and temporal information was conveyed through the *distribution* of effects over the 5-day period. We manipulated contiguity information by varying the distribution parameters of the experimental data sets, so that the effects either tended to occur closer to Day 1 or closer to Day 5 or were randomly distributed over the 5 days. In the control data sets, the temporal distribution of effects was always random.

With this setup, we thought it would be especially informative to include noncontingent conditions. In noncontingent conditions, the relative frequencies of effects are, by definition, identical in both groups. Studies involving real-time learning have often reported an outcome density bias, so that participants erroneously attach increasing effectiveness to the cause as both $P(e|c)$ and $P(e|\neg c)$ increase (e.g., Shanks, 1985). Subsequent work has revealed that such nonnormative judgments arise when participants fail to notice that relative frequencies are identical between the groups (Buehner et al., 2003). The tabular stimulus format makes the equality of frequencies especially noticeable and, thus, should lead to noncausal judgments throughout that are not subject to an outcome density bias.

However, contiguity biases may occur: If experimental and control groups vary with respect to the temporal distri-

bution of effects, participants may pick up on this feature, irrespective of the objective zero contingency. More specifically, contiguity may be interpreted as signifying generative causal power: Although the effect would have happened at some point anyway, the occurrence of the cause advances in time. Delay, in contrast, may be construed to convey preventive power: Although it cannot avert it, the cause postpones an effect to a later time. A contiguity effect of this sort has been reported in Anderson and Sheu (1995) and Wasserman and Neunaber (1986), but both of these studies employed real-time learning, and thus it is not clear to what extent contiguity genuinely influenced causal assessment over and above modulating the subjective perceived contingency, as outlined in the previous section.

Method

Participants. Forty-seven undergraduate students from Cardiff University participated in order to fulfill part of a course requirement. Due to an experimenter error, 5 of these received incorrect materials and were subsequently dropped from the analysis.

Design. The design comprised three within-subjects independent variables: $P(e|c)$, $P(e|\neg c)$, and *contiguity*. $P(e|c)$ and $P(e|\neg c)$ could take on values of .25, .50, .75, or 1.00, and *contiguity* had three levels: strong, random, and weak. We combined these variables to produce the 18 scenarios listed in Table 2. The dependent variable was the causal rating provided by the participant.

Materials. Contingency and contiguity information was presented in tabular format via a paper-and-pencil task. Each condition was implemented on a separate page, displaying two tables, listing information pertaining to $P(e|c)$, and $P(e|\neg c)$. Each table comprised 40 rows (for an overall sample size of 80) and six columns. The table conveyed contingency and contiguity information in the following way: The presence of e was indicated by an X in the relevant row, with the column indicating *when* X had occurred (on Days 1–5). Thus, for instance, for $P(e|c) = .25$, 10 out of the 40 rows carried an X; for $P(e|c) = .50$, 20 out of 40; for $P(e|c) = .75$, 30; and for $P(e|c) = 1.00$, all 40 rows carried an X. Analogously, the same frequencies were used to implement the various values of $P(e|\neg c)$.

In order to create the different levels of contiguity, different frequency density distributions were used to map out the Xs within the tables corresponding to $P(e|c)$. For the *strong contiguity* condition,

Table 2
Experimental Design and Data From Experiments 1 and 2

| Scenario | $P(e c)$ | $P(e \neg c)$ | ΔP | Contiguity | Experiment 1 ($n = 38$) | | Experiment 2 ($n = 35$) | |
|----------|----------|---------------|------------|------------|------------------------------|-----------|------------------------------|-----------|
| | | | | | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| A | .25 | .25 | .00 | strong | -9.18 | 31.42 | -1.26 | 35.19 |
| B | .50 | .25 | .25 | strong | -27.13 | 33.50 | -46.60 | 30.15 |
| C | .75 | .25 | .50 | strong | -47.55 | 35.02 | -79.43 | 13.22 |
| D | 1.00 | .25 | .75 | strong | -67.50 | 30.49 | -89.63 | 14.03 |
| E | .25 | .25 | .00 | weak | 18.26 | 33.48 | 21.77 | 31.40 |
| F | .50 | .25 | .25 | weak | -8.84 | 39.72 | -10.86 | 38.26 |
| G | .75 | .25 | .50 | weak | -7.20 | 53.38 | -33.26 | 35.79 |
| H | 1.00 | .25 | .75 | weak | -26.84 | 55.76 | -53.46 | 40.66 |
| I | .50 | .50 | .00 | strong | -8.82 | 25.36 | -25.14 | 24.06 |
| J | .75 | .75 | .00 | strong | -26.82 | 36.89 | -36.49 | 24.78 |
| K | 1.00 | 1.00 | .00 | strong | -27.34 | 34.65 | -45.46 | 32.60 |
| L | .50 | .50 | .00 | weak | 15.74 | 23.93 | 13.29 | 16.04 |
| M | .75 | .75 | .00 | weak | 22.18 | 36.46 | 23.49 | 22.38 |
| N | 1.00 | 1.00 | .00 | weak | 22.95 | 41.95 | 26.26 | 22.53 |
| O | .25 | .25 | .00 | random | 2.89 | 27.20 | 5.91 | 21.44 |
| P | .50 | .50 | .00 | random | 1.42 | 14.75 | 1.57 | 21.69 |
| Q | .75 | .75 | .00 | random | -1.97 | 16.71 | -9.14 | 10.47 |
| R | 1.00 | 1.00 | .00 | random | -1.18 | 16.25 | -7.86 | 12.14 |

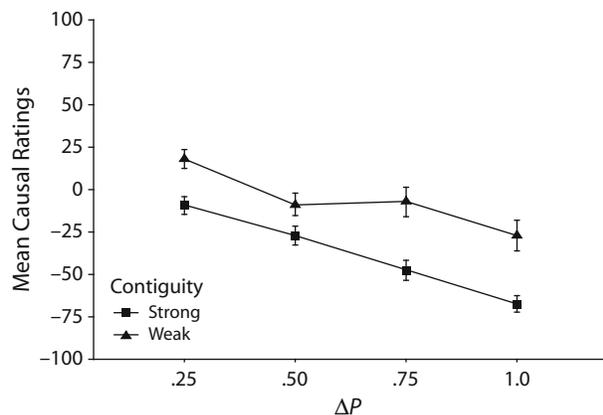


Figure 2. Experiment 1: Mean causal ratings in conditions with different contingencies. $P(e|\neg c) = .25$ in all conditions. Error bars represent standard errors. Negative ratings indicate high causal effectiveness.

the Xs were distributed so that 40% of the effects occurred on Day 1, 30% on Day 2, 20% on Day 3, 10% on Day 4, and 0% on Day 5. For example, in Condition C, there were 30 effects in total on Day 1, $0.4 \times 30 = 12$ Xs occurred; $0.3 \times 30 = 9$ Xs occurred on Day 2; $0.2 \times 30 = 6$ Xs occurred on Day 3; $0.1 \times 30 = 3$ Xs occurred on Day 4; and 0 Xs occurred on Day 5. For the *weak contiguity* condition, the reverse distribution was used: 0% on Day 1, 10% on Day 2, 20% on Day 3, 30% on Day 4, and 40% on Day 5. Finally, noncontingent problems, in addition to weak and strong contiguity, were also presented in a *random* format, employing a random distribution of the Xs throughout the table. The temporal positions of effects in the tables corresponding to $P(e|\neg c)$ were also distributed randomly within the table. Within each table, the order of rows was randomized to prevent the participants from easily spotting the contiguity pattern. Figure 1 shows an example stimulus.

A cover story explained that a biotechnology laboratory was investigating the effects of various radiation treatments on the survival of bacterial cultures. The instructions stated that radiation could promote the survival of cultures, prematurely kill the cultures, or not influence the survival of bacterial cultures. Furthermore, it was explained that bacterial cultures may, of course, die of natural causes, irrespective of radiation, and that this natural survival rate varies between different types of bacteria. Therefore, a control group of cultures, not exposed to any radiation, had always been included for each condition. Each of the 18 experiments had been conducted over a 5-day period, and laboratory assistants had recorded in a table when a culture had died. (See the Appendix for full instructions.)

Eighteen different bacteria and ray names were used for the 18 conditions. For each condition, the participants were presented with one table showing the results of 40 treated cultures and a second table showing the results of an untreated control group. The death of a culture constituted the effect and was indicated in the table by an X at the appropriate point in time. The participants had to rate the effect of each treatment on a scale ranging from -100 to $+100$, where $+100$ meant that the treatment was very effective at promoting survival of the bacteria, 0 meant that the treatment had no effect, and -100 meant that the treatment was very effective at killing the bacteria. They were also invited to provide free comments on each condition, if they so wished. The materials were assembled into individual booklets, with a different random order of conditions for each booklet.

Procedure. The booklets were available from the departmental resource room. The participants collected and completed the booklets on their own time. Upon return of a completed booklet, they were awarded 30 min of participation hours toward their course

requirement. Completion of the booklet took approximately 30 min, and the participants were encouraged to divide this time equally between conditions.

Results

Four participants had to be excluded from the analysis due to failure to comply with the instructions, resulting in a total sample size of 38 for the analysis. A significance level of .05 was adopted for all the analyses. We first analyzed the influence of contiguity on conditions with different contingencies. Figure 2 illustrates that the participants successfully discriminated between the four levels of contingency, with higher contingencies eliciting higher judgments of causal effectiveness. Figure 2 also suggests a clear influence of contiguity information, so that conditions with high contiguity received consistently higher ratings than did their low-contiguity counterparts. A 2 (contiguity) \times 4 (contingency) repeated measures ANOVA corroborated these impressions: There were main effects of contiguity [$F(1,37) = 36.26$, $MS_e = 3,668.84$, $h_p^2 = .50$] and contingency [$F(3,111) = 22.37$, $MS_e = 105,102.04$, $h_p^2 = .38$]; the interaction between these factors was marginally significant [$F(3,111) = 2.736$, $MS_e = 2,236.198$, $p = .047$, $h_p^2 = .07$].

Figure 3 shows the influence of contiguity on noncontingent conditions. As can be seen, contiguity also affected causal ratings of noncontingent conditions: conditions involving strong and weak contiguity elicited higher and lower ratings, respectively, of causal effectiveness, as compared with the random conditions with no systematic contiguity information. Overall, causal ratings were largely unaffected by variations in outcome density, although the effect of high and low contiguity appears to be particularly pronounced in the two conditions involving the highest outcome density. A 3 (contiguity) \times 4 (outcome density) repeated measures ANOVA revealed a main effect of contiguity [$F(2,74) = 30.25$, $MS_e = 54,377.36$, $h_p^2 = .45$]. The effect of outcome density failed to reach significance [$F(3,111) = 0.99$, $MS_e = 1,134.00$, n.s.,

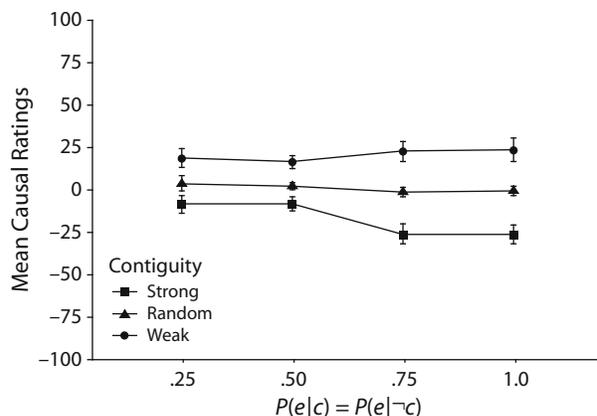


Figure 3. Experiment 1: Mean causal ratings in noncontingent conditions. $P(e|\neg c) = P(e|c)$ in all conditions. Error bars represent standard errors. Negative ratings indicate high causal effectiveness.

This study investigated the effect of **Gamma**-Rays on **PG13** bacteria

PG13 bacteria exposed to **Gamma** treatment on day 0

| | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 |
|----|-------|-------|-------|-------|-------|
| 1 | | | x | | |
| 2 | x | | | | |
| 3 | | x | | | |
| 4 | | x | | | |
| 5 | | | x | | |
| 6 | | x | | | |
| 7 | x | | | | |
| 8 | | | x | | |
| 9 | x | | | | |
| 10 | x | | | | |
| 11 | | | | | |
| 12 | x | | | | |
| 13 | | | x | | |
| 14 | | x | | | |
| 15 | | | | | |
| 16 | | x | | | |
| 17 | | | | | |
| 18 | | | x | | |
| 19 | x | | | | |
| 20 | x | | | | |
| 21 | | | | | |
| 22 | | | | x | |
| 23 | x | | | | |
| 24 | | | | x | |
| 25 | x | | | | |
| 26 | | x | | | |
| 27 | | x | | | |
| 28 | | | | | |
| 29 | | | | | |
| 30 | | | | | |
| 31 | x | | | | |
| 32 | | | | | |
| 33 | | | | | |
| 34 | x | | | | |
| 35 | | | | x | |
| 36 | | x | | | |
| 37 | x | | | | |
| 38 | | x | | | |
| 39 | | | x | | |
| 40 | | | | | |

PG13 bacteria with **no exposure** to any radiation treatment

| | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 |
|----|-------|-------|-------|-------|-------|
| 1 | | | | x | |
| 2 | | | | | |
| 3 | | x | | | |
| 4 | | | | | x |
| 5 | | | | | |
| 6 | x | | | | |
| 7 | | | | | |
| 8 | | | | | |
| 9 | | | | | |
| 10 | | | | | |
| 11 | | | | | |
| 12 | | | x | | |
| 13 | | | | | |
| 14 | | | | | |
| 15 | | | | | x |
| 16 | | | | | |
| 17 | | | | | |
| 18 | | | | | |
| 19 | | | | | |
| 20 | | | | | |
| 21 | | | | | |
| 22 | x | | | | |
| 23 | | | | | |
| 24 | | | | | |
| 25 | | | | | |
| 26 | | x | | | |
| 27 | | | | | |
| 28 | | | | | |
| 29 | | | | | |
| 30 | | | | | |
| 31 | | | | | |
| 32 | | | | | |
| 33 | x | | | | |
| 34 | | | | | |
| 35 | | | | | |
| 36 | | | | x | |
| 37 | | | | | |
| 38 | | | | | |
| 39 | | | | | |
| 40 | | | | | |

Please rate the influence of Gamma Rays on PG13 bacteria.

-100 ----- 0 ----- +100

Very effective at killing bacteria/no effect/very effective at promoting survival

Your rating? _____

Are there any other observations you can make about the treatment from the data?

Figure 4. Example stimulus materials for Condition C, Experiment 2.

$h_p^2 = .03$]. However, the contiguity \times outcome density interaction was highly significant [$F(6,222) = 4.39, MS_e = 1,814.39, h_p^2 = .11$]. To better understand the nature of this interaction, we analyzed the effect of outcome density separately for each level of contiguity. For conditions involving levels of strong contiguity, variations in outcome density produced a reliable linear trend [$F(1,37) = 7.83, MS_e = 9,979.63, h_p^2 = .18$] and a marginally significant cubic trend [$F(1,37) = 3.90, MS_e = 2,440.85, h_p^2 = .10$] in the direction of an outcome density effect. The weak contiguity and random conditions, in contrast, were unaffected by variations in outcome density [$F(3,111) = 0.640, MS_e = 436.50, n.s.,$ and $F(3,111) = 0.451, MS_e = 194.58, n.s.,$ respectively].

Fewer than 50% of the participants provided qualitative feedback beyond the numerical ratings. Most of the time comments were made on particular conditions only, rather than consistently for all the conditions. Although we did not systematically analyze these qualitative data, we noticed that some comments reflected how the participants tried to integrate contingency and contiguity information

into their ratings. Of particular interest were comments made on the noncontingent conditions: “no real difference but more earlier deaths in treatment group,” “kills bacteria earlier than no exposure but doesn’t kill many more bacteria over the 5 days,” or “kept most bacteria alive for 2 days, though all did die eventually.” Clearly, statements such as these reflect the fact that the participants noticed the overall zero contingency but also attached significance to the difference in temporal distribution between the experimental and the control groups.

Discussion

Experiment 1 demonstrated that people are sensitive to both contingency and contiguity information expressed in tabular format. When both parameters varied (see Figure 2), the participants were sensitive to both manipulations; when contingency was kept constant at a zero level (see Figure 3), the participants were sensitive only to variations in contiguity. What is particularly remarkable about the latter finding is that it produced systematic over- and underestimations of zero contingencies (relative to

the random controls) when contiguity was strong or weak, respectively; the participants correctly evaluated zero contingencies as noncausal only when the distribution of effects in the experimental group was just as random as that in the control group. The informational difference between the systematic and the random *distributions* of event frequencies in the experimental and control groups was sufficient to induce impressions of causality, even in the absence of a difference in total frequencies. In general, Experiment 1 did not show any evidence for an outcome density bias, apart from the linear trend in the strong contiguity conditions, which indicated that the contiguity effect was particularly strong when outcome density was high.

The stimulus format we adopted in this experiment, combined with the absence of time pressure inherent in our procedure, ensured a maximum transparency of the statistical parameters in each condition. Rather than focusing exclusively on event *frequency* information, the participants also paid close attention to event *distribution* over time. This was particularly apparent in the noncontingent conditions, where strong and weak contiguity turned otherwise noncausal assessments into causal and preventive ratings, respectively.

EXPERIMENT 2

As far as we are aware, the influence of frequency distribution on the evaluation of tabular data is novel to the field and has never been reported before. In order to ensure that the effects found in Experiment 1 are robust and not dependent on the particular stimulus format we employed, we decided to replicate Experiment 1, with a slight change in stimulus format. In earlier experiments in which a tabular format was employed only contingency was varied, but not distribution over time (e.g., White, 2003). In such experiments, participants were presented with only three rows of data; more specifically, information about the presence and absence of the effect was listed in only one row, as compared with five rows each for the cause-present and cause-absent data in Experiment 1. It might have been harder to keep track of the exact number of effect frequencies in our experiment, as compared with earlier studies. The aim of Experiment 2 was, therefore, to make information about the total frequency of effects in each group more salient, with the goal of testing the robustness of contiguity effects.

Method

Participants. Forty-six undergraduate students from Cardiff University participated to fulfill part of a course requirement. Due to experimenter error, 7 of these received incorrect materials and were excluded from the analysis.

Design, Materials, and Procedure. The design, materials, and procedure were identical to those in Experiment 1, with one notable difference: The presence and total number of effects were made more salient. This was achieved by selectively shading those cells in the table in which an effect had occurred. The cells to the right of effect-present cells (i.e., those representing the following days) were likewise shaded. Thus, the total number of effects across the 4-day period could easily be determined by counting the gray cells in the column representing the last day. Figure 4 displays an example stimulus.

Results and Discussion

Four participants failed to comply with the instructions and were removed from the data set, resulting in a total sample size of 35 being entered into the analysis. Figure 5 shows causal ratings for conditions with varying levels of contiguity. As in Experiment 1, the participants clearly distinguished between the contingencies and were strongly influenced by contiguity information. A 2 (contiguity) \times 4 (contingency) repeated measures ANOVA revealed main effects of contiguity [$F(1,34) = 67.35$, $MS_e = 87,120.43$, $h_p^2 = .67$] and contingency [$F(3,102) = 121.88$, $MS_e = 90,261.14$, $h_p^2 = .78$]. The contiguity \times contingency interaction was not significant [$F(3,102) = 2.49$, $MS_e = 1,573.42$, $h_p^2 = .07$].

The influence of contiguity on noncontingent conditions is depicted in Figure 6. Again, conditions with strong and weak levels of contiguity elicited higher and lower causal ratings, respectively, relative to the random control conditions. Furthermore, Figure 6 suggests that the impact of contiguity information increases with outcome density. A 3 (contiguity) \times 4 (outcome density) repeated measures ANOVA revealed a main effect of contiguity [$F(2,68) = 96.27$, $MS_e = 81,617.72$, $h_p^2 = .74$]. There was also an effect of outcome density [$F(3,102) = 11.12$, $MS_e = 6,819.04$, $h_p^2 = .25$], qualified by a contiguity \times outcome density interaction [$F(6,204) = 11.67$, $MS_e = 4,469.36$, $h_p^2 = .26$]. Trend analyses performed on each level of contiguity individually revealed that outcome density produced reliable linear trends in the strong contiguity [$F(1,34) = 40.84$, $MS_e = 36,259.21$, $h_p^2 = .55$] and the random [$F(1,34) = 12.91$, $MS_e = 4,737.20$, $h_p^2 = .28$] conditions, whereas it produced a cubic trend [$F(1,34) = 7.57$, $MS_e = 1,193.42$, $h_p^2 = .18$] in the weak contiguity condition.

Overall, Experiment 2 replicated the patterns found in Experiment 1, despite our change in stimulus format. The modification in the materials was aimed at increasing the salience of frequency information contained in each table: Comparing the number of shaded cells in the final columns allowed an easy assessment of the contingen-

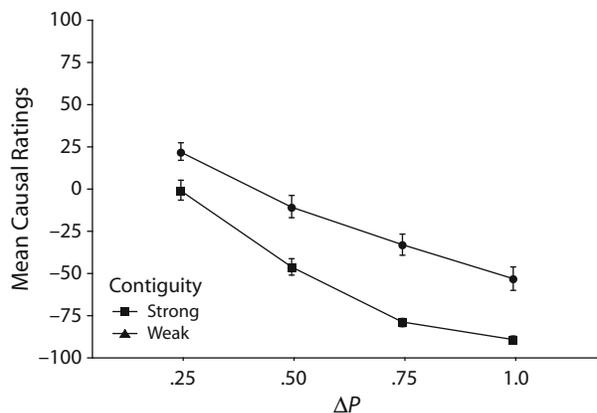


Figure 5. Experiment 2: Mean causal ratings in conditions with different contingencies. $P(e|-c) = .25$ in all conditions. Error bars represent standard errors. Negative ratings indicate high causal effectiveness.

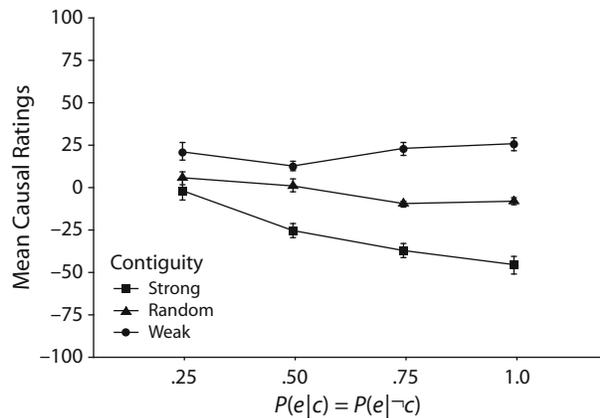


Figure 6. Experiment 2: Mean causal ratings in noncontingent conditions. $P(e|\neg c) = P(e|c)$ in all conditions. Error bars represent standard errors. Negative ratings indicate high causal effectiveness.

cies, similar to the instance–list procedure used by White (2003). The persistence of contiguity effects in light of a stimulus format that ensured that event frequency information was easy to read off suggests that our findings are indicative of a robust and general phenomenon: Contiguity and contingency both determine causal reasoning.

GENERAL DISCUSSION

Our aim was to investigate the influence of contiguity information on causal judgment in situations in which variations in contiguity could not alter subjective impressions of contingency. Two experiments have demonstrated that contingency and contiguity act in concert to influence causal induction from tabular data. Conditions involving identical contingencies were rated higher when contiguity was strong than when it was weak. Moreover, noncontingent conditions were interpreted as generative or preventive when contiguity was strong or weak, respectively. This pattern of findings was replicated under conditions whose aim was to enhance the accessibility of contingency information. Across both experiments, the tabular format we employed ensured that contingency could be assessed unambiguously and independently of variations in contiguity.

Our results show that current models of causal judgment are inadequate, since they focus exclusively on the role of contingency and, thus, cannot even represent the problems investigated in this article. Our experiments employed a minimal extension of a standard procedure in causal induction research. The simple inclusion of information about the temporal distribution of events, in addition to event frequencies, changed the pattern of results usually obtained when such information is omitted or left unspecified. Data distributions with a peak near the administration of the cause boosted impressions of causal power, whereas a peak distant from the cause weakened such impressions. In addition to this timing-induced attribution shift, temporal event distribution also managed to completely change the nature of the inference gleaned from noncontingent data. Zero contingencies with a random distribution of

events over time were evaluated as noncausal (mirroring effects obtained when temporal information has been absent in visual summaries; see Buehner et al., 2003). Adding a temporal distribution with its peak near or far from the administration of the cause, however, qualitatively changed the quality of the inference toward postulation of a generative or a preventive causal force.

We have thus—for the first time, as far as we are aware—demonstrated a genuine interaction between contingency and contiguity in causal judgments for situations in which information about both factors was unambiguously available. Previous studies had, of course, already investigated the influence of timing on causal learning. In most cases, it was demonstrated that delays impair causal learning (e.g., Shanks et al., 1989; but see Buehner & May, 2002), but in these studies, the role of timing was restricted to a *modulating* one, so that it reduced the overall impression of subjective contingency and, consequently, the causal rating. In Shanks et al.'s (1989) experiments, for instance, variations in contingency ($\Delta P = .75$ vs. $\Delta P = 0$) were rendered completely inconsequential with delays longer than 2 sec. In other words, once a critical temporal threshold had been exceeded, contingency evidence was no longer assessed—presumably, because it was subjectively reduced to zero. Similarly, timing has been shown to *overshadow* other relevant evidence or knowledge. The majority of studies in this area have employed ambiguous or workload-intensive tasks. In Schlottmann (1999), for example, young children (with comparatively less working memory) were confronted with multiple pieces of evidence and focused solely on cause–effect timing, rather than integrating the entire evidence toward a causal inference.

Our method differed from those in previous studies in two important ways. First, assessment of contingency information was not compromised by information about contiguity, due to the tabular format we employed. Second, the task was framed in a context that was compatible with multiple time frames; consequently, there would have been no a priori time frame bias, as in Schlottmann (1999) or Buehner and McGregor (2006), nor could a subjectively adopted time frame alter the extent of contingency, as in Hagmayer and Waldmann (2002). Information about both cause–effect contingency and contiguity was easily retrievable from the stimulus materials, and one did not influence or compromise the other. Instead, both cues acted in concert to influence causal inference.

The Role of Temporal Distributions

The influence of temporal distributions in noncontingent conditions is particularly interesting when a comparison is made with normative accounts of causation. Any normative model (see, e.g., Cheng, 1997) postulates that the absence of contingency signals the absence of causation.² On a probabilistic level, a lack of contingency implies that the cause makes no difference to the occurrence of the effect: The effect is just as likely when the cause is present as when it is absent. In noncontingent conditions involving random distributions of events over time, participants followed this normative principle and correctly inferred that the cause made no difference to the occurrence

of the effect. When the temporal distribution of effects in the experimental groups of noncontingent conditions had a discernible peak either near or far from the administration of the cause, however, participants did not think that the cause made no difference to the occurrence of the effect, even though, overall, the effect occurred just as often in the experimental as in the control group. The mere *temporal regularity*—that is, that effects tended to cluster soon or late after administration of the cause—was sufficient to create impressions of generative or preventive causality, even though there was no *statistical regularity*, at least not when considered over the entire data range of 5 days.

The temporal regularity effect not only is in conflict with current normative psychological theories of causation, but also contradicts epidemiological constructs of causality. Although epidemiology, unlike psychology, has already recognized the informational value of temporal distributions (Evans, 1993), it lists regular temporal distributions as only one of several requirements that must be fulfilled to license causal conjecture. Most important, epidemiological principles of causation state that, among other things, a disease has to have a higher prevalence and incidence in individuals exposed to the cause than in those not so exposed. In other words, the relative frequency of the effect has to be higher in the presence of the cause than in its absence. Expressed in terms of contingency, ΔP has to be greater than zero, just as in psychological theories of causation. Evidently, our participants ignored these considerations and, at least in the noncontingent conditions, focused exclusively on the temporal distribution of effects.

Future Perspectives

Real-time studies (e.g., Anderson & Sheu, 1995; Wasserman & Neunaber, 1986) have reported interactions between probability and timing that are qualitatively similar to our findings: Contiguous contingencies were given higher causal ratings than were equivalent alternatives that were less contiguous, and the mere advancing or postponing of an effect that would occur anyway was sufficient to create impressions of generative or preventive causal power. Because real-time paradigms often lack a clear trial structure, calculation of contingency values is impossible. Consequently, one cannot claim that such studies show an interaction between contingency and contiguity, as do our experiments. The difference in paradigms notwithstanding, could the similarity between pattern of results from both types of studies hint at a common underlying psychological process? Anderson and Sheu argued that a rate-based model provides the best psychological account for causal induction in real time. Rate-based models, by definition, synthesize probability and timing considerations and, thus, can, in principle, address interactions of the kind we reported. Future research will have to determine whether rate-based approaches can be extended to handle tabular data and whether the interactions between temporal distribution and probability that we reported can be explained as a synthesis of time and probability, in line with a rate-based approach.

AUTHOR NOTE

This article has been developed from W.J.G.'s final-year research project under the supervision of M.J.B. This article was written while M.J.B. was on study leave at the Max Planck Institute for Evolutionary Anthropology in Leipzig and at the Max Planck Institute for Human Cognitive and Brain Sciences in Munich. Correspondence concerning this article should be addressed to M. J. Buehner, School of Psychology, Cardiff University, Park Place, Cardiff CF10 3AT, Wales (e-mail: buehnerm@cardiff.ac.uk).

REFERENCES

- ANDERSON, J. R., & SHEU, C.-F. (1995). Causal inferences as perceptual judgments. *Memory & Cognition*, **23**, 510-524.
- BUEHNER, M. J., CHENG, P. W., & CLIFFORD, D. (2003). From covariation to causation: A test of the assumption of causal power. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 1119-1140.
- BUEHNER, M. J., & MAY, J. (2002). Knowledge mediates the timeframe of covariation assessment in human causal induction. *Thinking & Reasoning*, **8**, 269-295.
- BUEHNER, M. J., & MAY, J. (2003). Rethinking temporal contiguity and the judgement of causality: Effects of prior knowledge, experience, and reinforcement procedure. *Quarterly Journal of Experimental Psychology*, **56A**, 865-890.
- BUEHNER, M. J., & MAY, J. (2004). Abolishing the effect of reinforcement delay on human causal judgments. *Quarterly Journal of Experimental Psychology*, **57B**, 179-191.
- BUEHNER, M. J., & MCGREGOR, S. (2006). Temporal delays can facilitate causal attribution: Towards a general timeframe bias in causal induction. *Thinking & Reasoning*, **12**, 353-378.
- CHENG, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, **104**, 367-405.
- EINHORN, H. J., & HOGARTH, R. M. (1986). Judging probable cause. *Psychological Bulletin*, **99**, 3-19.
- EVANS, A. S. (1993). *Causation and disease: A chronological journey*. New York: Plenum.
- HAGMAYER, Y., & WALDMANN, M. R. (2002). How temporal assumptions influence causal judgments. *Memory & Cognition*, **30**, 1128-1137.
- HUME, D. (1888). A treatise of human nature. In L. A. Selby-Bigge (Ed.), *Hume's treatise of human nature*. Oxford: Oxford University Press, Clarendon Press. (Original work published 1739)
- LOBER, K., & SHANKS, D. R. (2000). Is causal induction based on causal power? Critique of Cheng (1997). *Psychological Review*, **107**, 195-212.
- MENDELSON, R., & SHULTZ, T. R. (1976). Covariation and temporal contiguity as principles of causal inference in young children. *Journal of Experimental Child Psychology*, **22**, 408-412.
- SCHLOTTMANN, A. (1999). Seeing it happen and knowing how it works: How children understand the relation between perceptual causality and underlying mechanism. *Developmental Psychology*, **35**, 303-317.
- SHANKS, D. R. (1985). Continuous monitoring of human contingency judgment across trials. *Memory & Cognition*, **13**, 158-167.
- SHANKS, D. R., & DICKINSON, A. (1987). Associative accounts of causality judgment. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 21, pp. 229-261). San Diego: Academic Press.
- SHANKS, D. R., HOLYOAK, K. J., & MEDIN, D. L. (Eds.) (1996). *The psychology of learning and motivation: Vol. 34. Causal learning*. San Diego: Academic Press.
- SHANKS, D. R., PEARSON, S. M., & DICKINSON, A. (1989). Temporal contiguity and the judgment of causality by human subjects. *Quarterly Journal of Experimental Psychology*, **41B**, 139-159.
- WASSERMAN, E. A., & NEUNABER, D. J. (1986). College students' responding to and rating of contingency relations: The role of temporal contiguity. *Journal of the Experimental Analysis of Behavior*, **46**, 15-35.
- WHITE, P. A. (2003). Making causal judgments from the proportion of confirming instances: The pCI rule. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 710-727.
- YOUNG, M. E. (1995). On the origin of personal causal theories. *Psychonomic Bulletin & Review*, **2**, 83-104.

NOTES

1. Shanks and Dickinson presented the modulating role of contiguity within a contingency account as an alternative to their preferred explana-

tion along the lines of associative learning.

2. Some special cases of noncontingent data, such as ceiling effects when generative causal influence is evaluated, imply that information about causality cannot be computed, rather than that causality is absent.

APPENDIX
Instructions for Experiment 1

A biotechnology lab investigates the effects of various types of radiation on the survival of bacterial cultures. Bacterial cultures may naturally die after a number of days, but the effect (if any) of radiation treatment on their survival is not yet known. It could be that radiation promotes the survival of bacterial cultures, or that it prematurely kills bacterial cultures. Alternatively, radiation may not have any influence whatsoever on the survival of bacterial cultures.

The laboratory experimented with 18 different species of bacteria. To study the effects of radiation treatment on a species, 40 bacterial cultures of that species were selected and had the treatment applied once. Each group of 40 cultures was then observed over a five-day period after the treatment had been applied. Laboratory assistants recorded the survival or death of each culture in a table.

Bacterial cultures may of course also die of natural causes, irrespective of exposure to radiation. Therefore for each treatment condition, 40 other cultures, that have not had any radiation treatment applied to them, were also observed over the five-day period as controls. The treated bacteria can then be compared to the untreated bacteria to see what the effect of the treatment was. Different types of bacteria may differ with respect to how well they survive in the absence of radiation.

The following tables show the results from the experiments. The table on the left (shaded) displays the cultures where the treatment has been applied. The table on the right (unshaded) displays the controls where no treatment has been applied. Each table shows the list of 40 cultures, and the five days over which they were observed. An X indicates that the culture has died on that day. Although it sounds obvious to say, a bacterial culture can only die once. If it dies on day 1 for example it cannot then die again on day 4.

Please examine the data in the tables and then rate how effective the treatment is, using a scale from -100 to +100.

-100 means that the treatment is extremely effective at killing bacteria.

0 means that the treatment has no effect on the survival or death of the bacteria.

+100 means that the treatment is extremely effective at promoting survival of the bacteria.

Use intermediate numbers if your judgment is in between these extreme values.

Try to spend the same amount of time on each of the following 18 problems.

Example

Here you can see that there is an x alongside culture 1 in the day 1 column, indicating that culture 1 died on day 1. Culture 2 survived the first two days but then died on day 3. There is no x in the culture 3 row, indicating that it did not die but survived the whole 5 days. Culture 4 died on day 2.

| | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 |
|---|-------|-------|-------|-------|-------|
| 1 | x | | | | |
| 2 | | | x | | |
| 3 | | | | | |
| 4 | | x | | | |

Here it can be seen that all cultures in this section died. Some died early on whereas others died later in the five-day period. Remember that in each case we do not know if the death of a culture is due to the treatment or due to natural causes.

So for the questions that follow, examine the data in the tables, compare the treated condition with the control, and then provide your rating of the treatment.

| | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 |
|---|-------|-------|-------|-------|-------|
| 1 | | x | | | |
| 2 | x | | | | |
| 3 | | x | | | |
| 4 | | | | x | |
| 5 | | | | | x |