Betting on Transitivity in Probabilistic Causal Chains

Dennis Hebbelmann

University of Heidelberg

Momme von Sydow

University of Heidelberg and University of Munich

Author Note

Dennis Hebbelmann, Psychological Institute, University of Heidelberg

Momme von Sydow, Psychological Institute, University of Heidelberg and MCMP, University of Munich

Correspondence concerning this article should be addressed to Dennis Hebbelmann, Psychological Institute, University of Heidelberg, Hauptstr. 47-51, 69117 Heidelberg, Germany.

Email: Dennis.Hebbelmann@psychologie.uni-heidelberg.de

Published in: Cognitive Processing, 2017 online, doi: 10.1007/s10339-017-0821-x,
free full text (read only): http://rdcu.be/trhV,
Springer: https://link.springer.com/article/10.1007%2Fs10339-017-0821-x

WORD COUNT: 9844

Abstract

Causal reasoning is crucial to people's decision-making in probabilistic environments. It may rely directly on data about covariation between variables (*correspondence*) or on inferences based on reasonable constraints if larger causal models are constructed based on local relations (*coherence*). For causal chains an often assumed constraint is transitivity. For probabilistic causal relations, mismatches between such transitive inferences and direct empirical evidence may lead to distortions of empirical evidence. Previous work has shown that people may use the generative local causal relations $A \rightarrow B$ and $B \rightarrow C$ to infer a positive indirect relation between events A and C, despite data showing that these events are actually independent (von Sydow et al., 2009, 2010, 2016). Here we used an economic sequential learning scenario to investigate how transitive reasoning in intransitive situations with negatively related distal events may relate to betting behavior. In three experiments participants bet as if they were influenced by a transitivity assumption, even when the data strongly contradicted transitivity.

*Keywords***:** causal induction, causal reasoning, transitivity, causal coherence hypothesis, betting

Causal Reasoning and Decision Making

Making decisions in probabilistic environments often requires us to bet on the (non-) occurrence of an event or the change of a variable: When we decide to ride our bike to work we get the benefit of a light early workout and a clear conscience, but risk getting soaked on the way back in case the weather changes. Deciding how to invest our money in the stock market we try to foresee changes affecting the value of companies in order to buy only stocks that will increase in value. As a proxy for the likelihood of an event we can try and observe other variables that correlate with it: Looking out the window and seeing a flock of swallows flying high we might infer that it will not rain and be more likely ride the bike to work.

It has recently increasingly been emphasized that human decision making under risk or uncertainty often involves causal reasoning (Hagmayer & Meder, 2013, Hagmayer & Sloman, 2009; cf. Osman, 2010). Valid causal models enable us to make predictions that go beyond correlational approaches in at least two meaningful ways: They enable agents to make predictions about correlations they have not observed yet (through causal reasoning) and second, they allow for predictions about the effects of interventions in a causal system. When asked to predict the (non-)occurrence of an event, people should take their knowledge about the presence or absence of possible causes into account. Hagmayer and Meder (2013) have recently shown that people base their decisions, quite reasonably, on their causal beliefs: When asked to intervene in a probabilistic causal system that is fixing the value of a variable to a particular value, participants appeared to base their decisions on causal reasoning, as opposed to purely correlational information.

For a causal model to provide a useful framework for decision making it needs to fulfill two criteria: First, it needs to accurately represent the absence or presence and strength and direction of causal relations between events in the real world (*correspondence*). If the model postulates incorrect causal relations then faulty or at least non-optimal predictions and misdirected interventions can result. Secondly, it needs to provide the agent with unambiguous predictions based on valid rules of reasoning (*coherence*).

Correspondence and coherence, more generally, are respectable benchmarks of truth and they may be seen either as domain general principles in science or as assumptions of domain-specific models (cf. Arkes, Gigerenzer, and Hertwig, 2016, for a view critical of coherence as a domain-general approach). We are here concerned specifically with causal coherence in decision making.

Bayesian Probabilistic Causal Networks (Pearl, 2000; Spirtes, Glymour, & Scheines, 2001; cf. Rottman & Hastie, 2014, for a review), further referred to as Causal Bayes Nets, provide a prominent normative framework for integrating different variables and their levels of covariation into one consistent network of directed causal relations. They are constrained by prior knowledge about possible causal relations: For example, an effect cannot precede its cause and a causal network may not create any circular sets of causal relations. A Causal Bayes Net consists of nodes representing variables, one-directionally directed edges (“arrows”) between nodes pointing from causes to effects and parameters for each arrow indicating the strength and direction of the respective causal relation (positive values for *generative* causes, increasing the likelihood of an event, and negative values for *inhibiting* causes, decreasing the likelihood an event).

Transitive Reasoning in Probabilistic Causal Chains

Causal Bayes Nets (Pearl, 2000; Spirtes, Glymour, & Scheines, 2001) and, in psychology, the theory of analogous mental models (Sloman, 2005; Lagnado, Waldmann, Hagmayer, & Sloman, 2007; Waldmann, 1996; Waldmann, Cheng, Hagmayer, & Blaisdell, 2008) build on the assumption of the Markov condition, stating that any node in a causal model is conditionally independent of all upstream nodes, given its parents (Hausman & Woodward, 1999; Spohn, 2001). This entails transitivity in causal chains: If A causes B, and B causes C, without any further links connecting the three variables, then A causes C (via B). If in a probabilistic causal chain the Markov condition holds, then the strength of the global relation $A \rightarrow C$ can be inferred from the strength of the local relations $A \rightarrow B$ and $B \rightarrow C$: Using the causal strength estimate ΔP ($\Delta P_{AB} = P(B|A) - P(B|\neg A)$; Jenkins & Ward, 1965), the global ΔP can be calculated by multiplying all local ΔPs between nodes forming the causal chain (e.g., $\Delta P_{AC} = \Delta P_{AB} \times \Delta P_{BC}$). It is therefore not necessary to observe the global relation directly.

Related research on transitive reasoning in the induction of causal chains has shown that people assume a transitive causal relation based on integrating single causal links (Ahn & Dennis, 2000; Baetu & Baker, 2009). This research corroborated the hypothesis that people reasoned transitively even if no information on the global relation was shown.

Subsequent research started to investigate *intransitive* chains (von Sydow, Hagmayer, & Meder, 2016; von Sydow, Meder, & Hagmayer, 2009; von Sydow, Meder, Hagmayer, & Waldmann, 2010; see Rottman & Hastie, 2014, for an overview). Using intransitive chains allows bringing correspondence and coherence into conflict.

To illustrate this, Figure 1 shows a typical intransitive pattern of causal relations as used in the following studies. Circles A to D represent a temporal succession of events for eight individuals. Each tile position (numbered 1 to 8) represents an individual (or, here, a company). Individual 1 displays the properties A, $\neg B \neg C$, and $\neg D$. Here all local relations ($A \rightarrow B$, $B \rightarrow C$, $C \rightarrow D$) are positive. However, A and C are independent of each other, while A and D are even negatively related, thus violating transitivity.

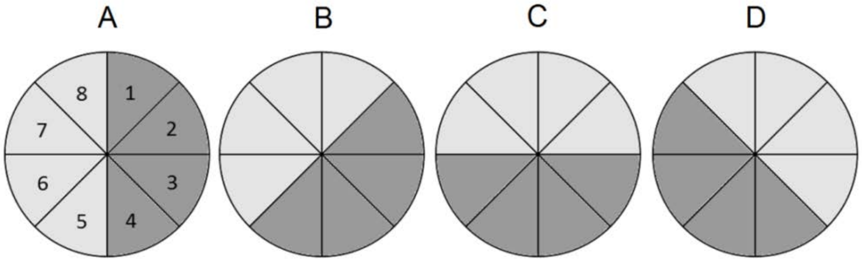


Figure 1: Structure of statistical relations between events *A*, *B*, *C*, and *D*. In this illustration, individual 1 would be “*A*, ¬*B*, ¬*C*, ¬*D*”, while individual 6 would be “¬*A*, ¬*B*, *C*, *D*”.

Von Sydow et al. (2009; 2010; 2016) suggest a *causal coherence hypothesis* that coherence-based induction may distort the perceived shown evidence about causal relations if the data violate the structural assumptions of Causal Bayes Nets. People are taken to assume, at least by default, a modular integration of single causal relations into larger causal networks, for instance implying transitivity in causal chains (cf. Waldmann, Cheng, Hagmayer, & Blaisdell, 2009). This is predicted even when evidence to the contrary is available, but people mightabandon this default belief if the mismatch between coherence-based induction and correspondence-based induction becomes quite evident.

Intransitive chains are at odds with structural implications of Causal Bayes Nets and involve a violation of the Markov condition. In the philosophical debate it has been put into question whether all causal relations necessarily adhere to the Markov condition and whether, as a consequence, chains need to be transitive (Cartwright 2001, 2006; Sober, 1988; Sober & Steel, 2012). However, even strict advocates of the Markov condition have pointed out that on the level of our *actually used* categories causal chains may not adhere to the Markov condition (Hausman & Woodward, 1999; Spohn, 2001). For instance, this may be the case if a category is the product of mixing subclasses for which different causal relations hold (Hausman & Woodward, 1999; von Sydow et al., 2016; cf. Johnson & Ahn, 2015, for other transitivity violations).

Von Sydow et al. (2009, 2010, 2016) showed that in several formats (overview format, sequential learning format) participants may infer the strength and direction of the relation $A \rightarrow C$ after learning $A \rightarrow B$ and $B \rightarrow C$, even if this is not warranted by the data presented to them: In the materials used, $A \rightarrow B$ and $B \rightarrow C$ were positive, while *A* and *C* were statistically independent from each other ($\Delta P_{AB} = \Delta P_{BC} = .5$, $\Delta P_{AC} = 0$, cf. Figure 1). If they were presented with $A \rightarrow B$ and $B \rightarrow C$ first, participants judged $A \rightarrow C$ in line with transitivity. This effect remained stable even when participants were able to directly gather information about $A \rightarrow C$.

In von Sydow et al. (2016) participants obtained overview panels where they could even have detected different subclasses of *A* with different effects on *C* ("mixing of subclasses"). Although studies of causal induction have mostly assumed homogeneity of variables, one often does not know this beforehand. Often the classes of events invovled may be heterogeneous and thus the observed correlations between such classes only mixtures from various correlations or causal relations holding for the subclasses. Take, for instance, potential causes and effects of depression. A particular attribution style seems to be a cause *A* of depression, and depression *B* has in turn been suggested to facilitate dementia *C*. Nonetheless, it may be inappropriate to reason transitively from *A* to *C* here, since there may be different kinds of depression involved (*B1* and *B2*). Perhaps only people with a hormonal imbalance have an increased risk of dementia, and not those with depression caused by a specific attribution style (on the intricate relationship between categorization and causality, cf. Lien & Cheng, 2000; Hagmayer, Meder, von Sydow, & Waldmann, 2011; Waldmann, Meder, von Sydow, & Hagmayer, 2010). Analogous to such examples, von Sydow et al. (2016) in a laboratory task investigated situations in which $A \rightarrow B$ and $B \rightarrow C$ held. But large items *A* always led to *C*, whereas small items *A* never led to *C*. Thus, using a similar correlational structure as in Figure 1, the transitive inference $A \rightarrow C$ was invalid (both events were independent). People nonetheless tended to draw transitive conclusions and were not very successful in distinguishing subclasses.

Von Sydow et al. (2010) has explored whether a focus on individual evens *A*, *B*, and *C* leads to the predicted causal distortion effects also in sequential learning scenarios.

However, in many regards the boundary conditions of the causal coherence hypothesis need further exploration. For instance it is not clear whether people continue to infer a positive distal causal relation from positive local relations if it is *actually negative* or if they completely switch to induction based on the observed data due to the obvious mismatch. We investigate non-transitive causal chains were the observed overall relation is even opposed to the local relations to test the causal coherence hypothesis. Even if correlations are based on mixing, results that approximate transitivity may occur just by chance. With regard to clear deviations, zero correlations or only *weakly* inversed (e.g., weakly negatively related) overall relations may occur more often from mixing than the investigated strongly inversed (e.g., strongly negatively related) relations. We investigated strongly inversed relations, since this provided a particularly strict test of our hypothesis that the induction in a chain of causal events is not only based on directly observed evidence, but also depends partly on inferences using the (here illegitimate) assumption of transitivity when the observed evidence is almost always opposed to the transitively predicted ones.

Furthermore, research on the causal coherence hypothesis mainly focused on people's unincentivized judgments about causal relations. Causal reasoning might however focus more on directly observed information when decisions based on it have an impact on a participant's payoff. Hence, we will investigate betting as dependent variable. Betting is often taken as a means to decrease bias in probability judgment (Andersson & Nilsson, 2015), and betting should direct participants' focus on incentives. Nevertheless, betting may reduce coherence-based distortions of evidence (because betters may pay more attention to the accuracy of their judgments), we predict that people’s bets on the occurrence of a possible effect are informed by both correspondence and coherence, two sources of information that contradict each other in this case. Their betting may either correspond to their probability judgments (probability matching; cf. Herrnstein, 1970; Vulkan, 2000) or to an optimal exploitation of their given probability judgment: If optimizers use bottom-up induction and realize the negative distal relation they should put all stakes on the negative prediction, if they use a coherence-based approach to infer the distal relation, they should put all stakes on the positive prediction.

Goals and Hypotheses

In the three experiments presented here we investigated the influence of causal beliefs on people’s decision making in an environment where transitivity is violated. In an economic sequential learning scenario participants first observed co-occurrences of four events in a non-transitive causal chain and afterwards judged the statistical relations between events.

Going beyond previous research, all three experiments examined whether causal coherence still affects participants’ judgments if the distal events in the chain are not only *independent* of each other, but their relation even runs *contrary* to the assumption of transitivity (a negative global relation when transitivity suggests a positive one and vice versa).

Additionally, all three experiments explored whether causal coherence effects can also be found with betting. In Experiment 1 we hypothesized that the causal coherence should not only influence participants’ judgments of the global relation but also the amount of money bet in line with a transitive causal model, thereby performing worse than a control group in which causal coherence should not have an effect. Here participants only had to bet in a final round. In Experiment 2 we examined whether this effect remains stable after *repeated* betting on the global relation with feedback about their performance. In Experiment 3 we examined whether the effect remains stable even if the *only* source of information on the relations is through betting on them.

Experiment 1: Betting Biases in Learning Relations

Goals and Hypotheses

In Experiment 1 we tried to replicate and expand on previous findings that causal induction is distorted by assumptions of transitivity in a trial-by-trial learning paradigm. Two major aspects were added to this line of research: First, the global relation $A \rightarrow D$ was not independent but strongly negative ($\Delta P_{AD} = -.5$). This should make the mismatch between correspondence and coherence even more salient. Second, participants were asked not only to judge the perceived global relation, but also to bet on the occurrence of D vs. $\neg D$ based on their information on A vs. $\neg A$ in a final round. This should incentivize participants to judge $A \rightarrow D$ accurately and bet accordingly. Constructing $A \rightarrow D$ not to be independent but negative made A vs. $\neg A$ a strong negative predictor for D vs. $\neg D$. A correspondence-based judgment should result in betting more money on $\neg D$ given A (or on D given non-A). A coherence-based judgment should result in participants betting more money on D given A and (or on $\neg D$ given A).

The learning trials presented are the same regardless of condition. Our three experimental conditions should lead participants to focus on the local relations, the global relation, or both the local + global relations, and we expected this to have an effect (cf. von Sydow et al., 2010). In contrast, if participants relied on correspondence alone, their judgments and bets on $A \rightarrow D$ should be the same. We induced focus by repeatedly asking for particular relations only. This manipulation is based on the general idea that causal links are learned successively rather than simultaneously (Waldmann et al., 2009). In the local-only condition, when participants were repeatedly asked to assess local relations, we expected them to learn these relations and use them most strongly to construct the overall relations by illegitimately assuming transitivity. In contrast, in the global-only focus condition, participants' judgments are expected to be more in line with the actually observed $A \rightarrow D$ relations. Finally, in the local + global conditions participants were asked to learn both local and the global relations, and here we expected still to observe some coherence-based distortions of the evidence.

Participants

We tested 84 participants (50 female, age $M = 23.6$) who were recruited at the University of Heidelberg as part of a multi-experiment session. Participants received 6€ per hour or course-credits for taking part. (Psychology students at the University of Heidelberg are required to participate in a number of experiments during their undergraduate studies and could choose to have their participation added to their "experiments account" in lieu of receiving payment.) In any case, participation was voluntary. Additionally, and regardless of the chosen form of reimbursement, participants obtained the proportion of 1 € they bet on the correct outcome in a final bet.

Material and Procedure

Participants were to observe individual companies and temporally ordered events related to each company during learning blocks. Each trial in these blocks represents an individual company (cf. von Sydow et al., 2010). For each company, participants were shown four sequential events represented by four pictures (Figure 2): Each company either buys or does not buy stocks of a second company (A vs. $\neg A$), then rises or falls on a general performance index (B vs. $\neg B$), is positively or negatively evaluated by the *Economist* (C vs. $\neg C$), and in the end either increases or decreases in stock market value (D vs. $\neg D$). The instruction stressed the temporal order of the events, which is a known cue inducing causal structure (Lagnado & Sloman, 2006). However, we neither suggested that the chain is transitive nor that specific relations were positive or negative.

The local relations between all four events were positive ($\Delta P_{AB} = \Delta P_{BC} = \Delta P_{CD} = .5$), while the global relation $A \rightarrow D$ was negative ($\Delta P_{AD} = -.5$). Figure 1 illustrates the contingencies shown to the participants. Each of the four events occurred with a probability of $P = .5$ (dark shaded segments in Figure 1). Combining four events (and their negations) results in sixteen possible trial-types, of which eight were used to produce the non-transitive chain. Figure 1 shows all types of trials used in Experiment 1. Each of the eight types of trials was used twice in each of the learning blocks, resulting in sixteen trials per block. There were twelve learning blocks, resulting in 196 learning trials.

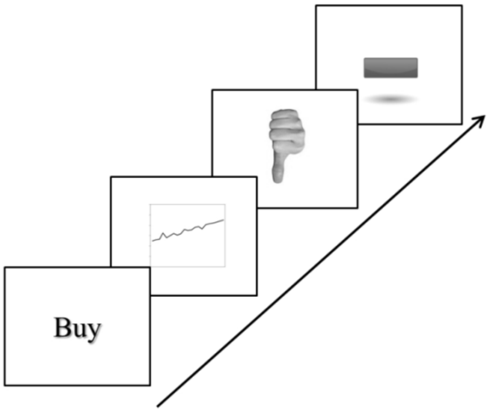


Figure 2: Exemplary trial representing one company (A, B, $\neg C$, $\neg D$).

The Experiment consisted of learning and testing blocks. Before each learning block participants were instructed which relation to focus on and were tested on only this relation afterwards. The instructions to focus for instance on $A \rightarrow B$ for the forthcoming learning block read (translated from German): “You will now see the data of several individual companies. After this block you will be asked to judge whether there is a relation between the companies’ buying decisions in Week One and their performance on the General Performance Index in Week Two; and if so, what kind of relation it is. You will always see data of all four weeks.“

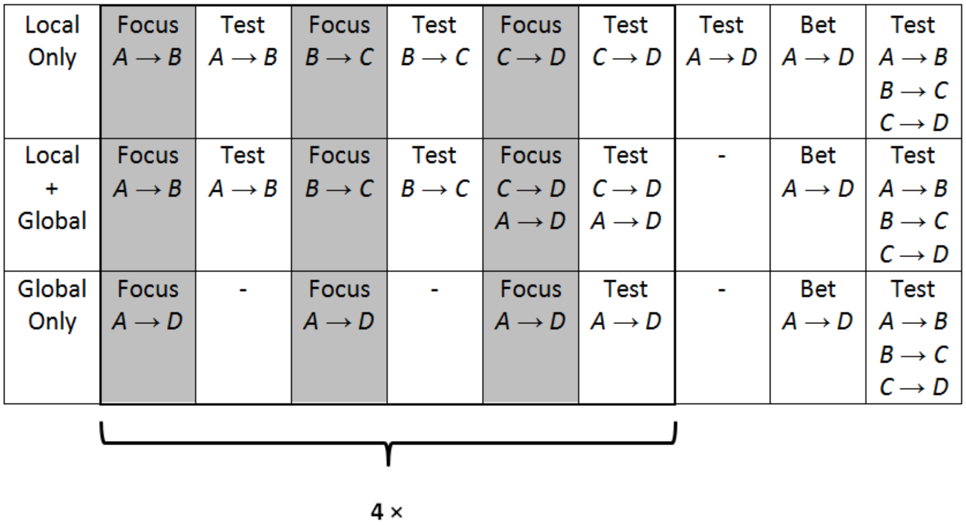
Each learning block consisted of the same 16 trials in randomized order, regardless of condition; therefore participants did not differ in the learning material presented to them. Each trial was 4 seconds long with each picture being presented for 1 second. Participants started each trial by clicking a “Next” button on the screen.

In each test phase participants judged the relation they had focused on during the preceding learning trial on a 21-point scale ranging from -100, indicating a deterministic negative relation (e.g.: “If a company is positively evaluated, then its stock market value will always decrease”), to 100, indicating a deterministic positive relation (e.g.: “If a company is positively evaluated, then its stock market value will always increase”), with a middle point of 0, indicating statistical independence.

Participants were randomly assigned to one of three conditions which only differed in learning instructions and testing (Figure 3): Participants in the *local-only* condition only focused in the learning phases (and were only tested) on the local relations *A* → *B*, *B* → *C* and *C* → *D*. We expected the local-only group’s estimates and bets on *A* → *D* to be least influenced by correspondence among the three conditions, as they were not instructed to focus on *A* → *D* and would rely heavily on coherence-based integration of local relations. Note that the local-only group still saw the same data pattern as the other two and could have learned the negative relation *A* → *D*. We expected the *global-only* group’s estimates and bets to be most in line with correspondence among the three groups, as they did not directly focus on the local relations and were not to rely on their coherence-based integration. Therefore their estimates and bets should reflect a more accurate assessment of *A* → *D*. Participants in the *local + global* condition were tested on both the local relations and the global relation *A* → *D*. We expected the local + global group's estimates and bets to fall between the two remaining groups: Because participants focused on the local relations first, we expected them to engage in coherence-based integration; but we also expected correspondence to affect the final result. This would further strengthen the assumption that coherence plays an important role in causal learning even if that stands in clear contrast to correspondence.

After the learning blocks and the test blocks all participants rated the perceived relation *A* → *D* on the same 21-point scale again. They were then told that they would see one more company drawn randomly from the ones they had seen so far during the experiment. This time they only saw the company's buying decision (*A* or ¬*A*). They had 100 Eurocents at their disposal to bet on the change of the company's stock market value (*D* vs. ¬*D*). Participants could split their money between the two options and would win the amount of money they bet on the right outcome. The outcome was shown afterwards and participants were paid the amount of money they had won on top of their usual reimbursement.

After the betting trial, all participants rated the local relations one last time.



Local Only	Focus *A* → *B*	Test *A* → *B*	Focus *B* → *C*	Test *B* → *C*	Focus *C* → *D*	Test *C* → *D*	Test *A* → *D*	Bet *A* → *D*	Test *A* → *B* *B* → *C* *C* → *D*
Local + Global	Focus *A* → *B*	Test *A* → *B*	Focus *B* → *C*	Test *B* → *C*	Focus *C* → *D* *A* → *D*	Test *C* → *D* *A* → *D*	-	Bet *A* → *D*	Test *A* → *B* *B* → *C* *C* → *D*
Global Only	Focus *A* → *D*	-	Focus *A* → *D*	-	Focus *A* → *D*	Test *A* → *D*	-	Bet *A* → *D*	Test *A* → *B* *B* → *C* *C* → *D*

Figure 3: Temporal structure of Experiment 1 (in the learning blocks we here present the foci of the three different conditions).

Results

Betting on the Global Relation. To compare participants' betting performances, we first calculated how much money each participant bet on the most likely outcome given the information about *A* vs. ¬*A* (*ideal bet*), i.e. if participants saw an instance of *A*, their ideal bet would be the amount they bet on ¬*D* and for ¬*A* it would be the amount they bet on *D*. Figure 4 shows participants' mean ideal bets by condition. A one-way ANOVA with the ideal bet as the dependent variable showed a significant main effect of condition, $F(2, 81) = 9.73$, $p < .001$. The local-only group bet less money on the ideal bet, $M = 32.1$, $SD = 27.1$, than the global-only group, $M = 66.6$, $SD = 29.2$, with the local + global group falling between the two, $M = 51.5$, $SD = 30.4$ (Figure 4). A post-hoc comparison of group means revealed a significant difference between the local-only group and the other two, *p*s < .05, and a nearly significant difference between the local + global and global-only group, $p = .05$.

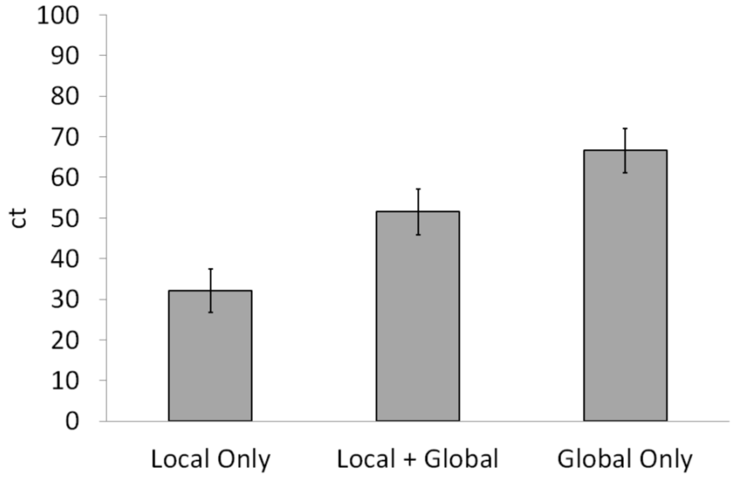


Figure 4: Mean ideal bets in ct on *D* vs. ¬*D* (± SE).

Estimates of the Global Relation. We expected participants in the local-only group to judge *A* → *D* to be positive, in line with an influence of the transitivity assumption, even though they could have seen the negative relation during 196 trials. We further predicted the global-only group to judge *A* → *D* to be clearly negative, in line with the data. As the local + global group's estimates should be informed by both correspondence and coherence we expected their estimates to fall between the other two groups. Figure 5 shows participants' mean estimates of the relevant estimates of the global relation *A* → *D*. A one-way ANOVA comparing the groups mean estimates confirms our hypothesis[1]: We found a significant main effect of condition, $F(2, 81) = 23.99$, $p < .001$. A post-hoc comparison of group means revealed significant differences between all three groups, *p*s < .01. Participants in the local-only group judged *A* → *D* to be positive, $M = 36.2$, $SD = 29.4$, the global-only group judged it to be negative, $M = –27.6$, $SD = 37.8$, with the local + global group falling between the other two, $M = 5.2$, $SD = 34.2$.

Note that the local-only group's estimates are even considerably higher than predicted by a perfectly transitive inference (which would correspond to an estimate of +12 on our scale). The global-only group's mean estimate is closer to the estimate predicted by correspondence alone (corresponding to –50 on our scale).

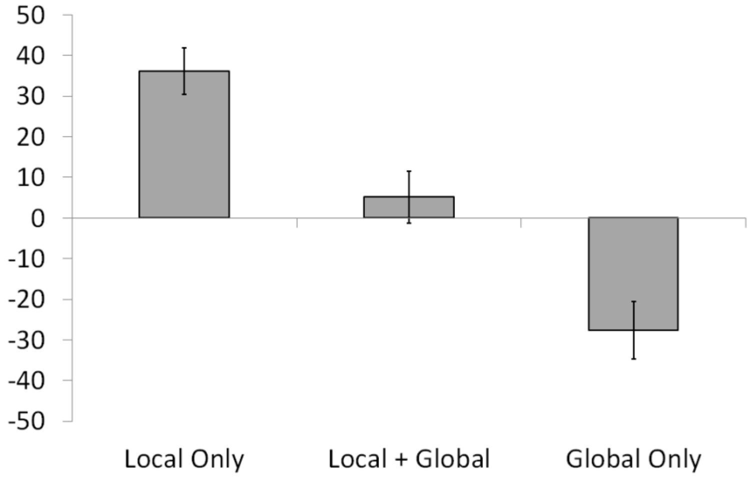


Figure 5: Mean estimates for *A* → *D* (± SE), ranging from -100, indicating a deterministic negative relation, to 100, indicating a deterministic positive relation.

[1] Although normal distribution was violated within conditions we still report the results of parametric tests as they have proven to be robust against this deviation. In all cases analyses using non-parametric tests led to comparable results to those reported.

Estimates of Local Relations. We expected the local relations to be lower in the global-only condition than in the other two conditions. This also seems to be a precondition for predicting a reduction of transitive inferences based on integrating these local links in the global-only condition. In this and all other experiments this ‘manipulation check’ was successful. Additionally, there was always no reliable difference between the local + global and the global-only condition. We here focus on presenting the main results of the study, but Online Resource 1 provides more results on estimates of local relations.

Discussion

In the first experiment we could replicate and expand von Sydow et al.’s (2010) findings. Participants stick with their assumptions of transitivity even if they are presented with *strongly* contradicting evidence during a total of 196 learning trials. However, the results in the local + global group show that this assumption is not impervious to experience: Participants’ mean judgments near the point of statistical independence may either reflect some averaging of correspondence- and coherence-based judgments or participants’ resulting confusion about the true nature of the relation. In any case, the judgment differed considerably from the value predicted by correspondence alone and also between conditions. Transitive interpretations must have had a strong impact on participants’ estimates.

In the local-only group participants’ estimates of $A \rightarrow D$ were considerably higher than predicted by an inference purely based on transitivity. Perhaps, people may not use an as fine-grained scale to convey that there is a positive relation than is suggested by a fully parameterized Bayesian model. Alternatively, this deviation may be linked to previously found deviations from the Markov condition in experimental paradigms not directly assessing transitive reasoning, but showing too positive relations in functioning chains (Rehder & Burnett, 2005).

The results of the betting trial provide first evidence that participants' assumptions of transitivity not only influence their judgments but also their decision making in a betting task: Participants in the local-only group were willing to bet most of their money in line with the belief that $A \rightarrow D$ is positive. Participants in the global-only group accurately judged $A \rightarrow D$ to be negative and bet most of their money accordingly, leading to more money bet on the most likely outcome.

Experiment 2: Extensive Learning and Repeated Betting

Goals and Hypotheses

In Experiment 2 we sought to replicate and expand on our initial findings in two important ways: First, we introduced multiple betting-trials with immediate feedback. Betting with immediate feedback should help correct inaccurate assessments of covariation (Andersson & Nilsson, 2010), therefore putting the causal-coherence hypothesis to a stronger test: Should a similar pattern of estimates as in Experiment 1 occur, this would show how persistent the effect of coherence is in the face of contradicting evidence.

Second, the results of Experiment 1 may not have been driven by assumptions of transitivity alone, but rather by prior knowledge about the stimulus material. To rule this out, we counterbalanced the first local relation being positive/negative with the global relation negative/positive, again violating transitivity.

Despite these two extensions, our hypotheses were analogous to those in Experiment 1: The local-only group's estimates and bets should be most influenced by coherence and the global only group most influenced by correspondence, with the local + global group falling between the two.

Participants

We tested 94 participants (67 female, age *M* = 23.3) who were recruited at the University of Heidelberg as part of a multi-experiment session. Participants received 6€ per hour or course credits (cf. Experiment 1). Regardless of the chosen form of show-up fee, participants received up to 3 € extra, depending on their betting performance.

Material and Procedure

Experiment 2 followed a structure similar to Experiment 1 (Figure 3), with each testing phase replaced by one betting trial as described in Experiment 1. Repeated betting on $A \rightarrow D$ should incentivize accurate learning even more, therefore putting the causal coherence hypothesis to a stronger test.

Each phase of Experiment 1 in which participants judged the relation $A \rightarrow B$ was replaced by a betting trial where participants saw *A* or $\neg A$ and were asked to bet on *B* vs. $\neg B$, etc. The same holds for other phases where participants had to judge a relation. In each betting trial participants bet 100 points they could split between the two possible outcomes. Participants won the amount of points they bet on the right outcome and received immediate feedback about the points they won. At the end of the experiment participants were paid up to 3 € on top of their usual reimbursement depending on how many points they had collected.

Participants were again randomly assigned to either the local-only, local + global, or global-only group, analogously to the design of Experiment 1. At the end of the experiment participants judged the local relations and the global relation on the same scale as used in Experiment 1.

Due to the naturalistic material used in Experiment 1 participants might have had prior beliefs about $A \rightarrow D$ being positive. Their responses in the local-only group may therefore not indicate transitive reasoning but rather participants' resorting to prior beliefs in the absence of further knowledge. To control for the effect of a general tendency to judge $A \rightarrow D$ positively we counterbalanced between participants whether $A \rightarrow B$ was positive or negative ($\Delta P_{AB} = .5$ vs. $\Delta P_{AB} = -.5$). With $A \rightarrow B$ being negative and the other local relations remaining positive, the actual intransitive relation $A \rightarrow D$ was *positive*, $\Delta P_{AD} = .5$, but the coherence-based prediction was *negative* for this relation. In both cases we expected participants in the local-only group and the local + global group to bet in line with transitivity.

Results

Betting on the Global Relation. Again all participants bet on $A \rightarrow D$ after the last learning block. To compare participants' betting performances we first calculated how much money each participant bet on the most likely outcome given the information about A vs. $\neg A$. To find out whether counterbalancing of $A \rightarrow B$ being positive or negative had an effect on participants' betting performance, we included it as a between-subject factor in the analysis. A 3 × 2 factor ANOVA with condition and counterbalancing as between-subjects factors showed no significant main effect of counterbalancing, $F(1, 92) = 3.4$, $p = .07$, as well as no significant interaction between condition and counterbalancing, $F(2, 91) = 1.7$, $p = .19$. We found a significant main effect of condition, $F(2, 91) = 19.2$, $p < .001$ (Figure 6). The local-only group bet less money on the ideal bet, $M = 43.5$, $SD = 19.5$, than the global-only group, $M = 78.5$, $SD = 24.0$, with the local + global group falling between the two, $M = 52.7$, $SD = 25.9$. A post-hoc comparison of group means revealed a significant difference between the global-only group and the other two, ps < .001, but not between the local-only and the local + global group, $p = .12$.

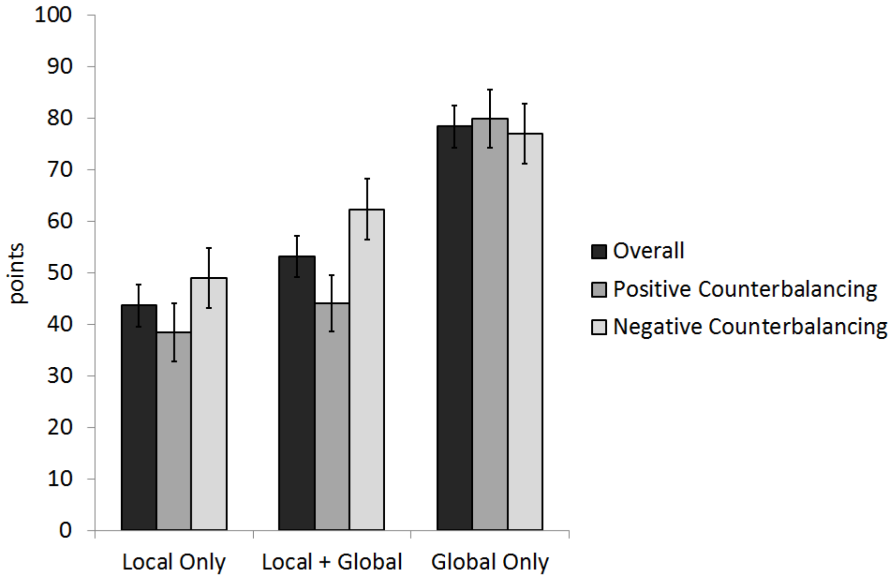


Figure 6: Mean ideal bets on *D* vs. ¬*D* (± SE), Experiment 2.

Estimates of the Global Relation. If ΔP_{AB} = –.5 participants' answers were reverse coded. We expected participants in the local-only group to judge $A \longrightarrow D$ in line with the assumption of transitivity, as predicted by the causal coherence hypothesis. The global-only group should judge $A \longrightarrow D$ in line with the data presented during learning trials (represented by negative values in Figure 7). We expected the local + global group's estimates to fall between the two other conditions as they should be driven by both correspondence and coherence. We also included the counterbalancing conditions in the analysis as prior beliefs about the strength and direction of causal relations might have influenced participants' judgments in Experiment 1. A 2 × 3 between-subject ANOVA comparing the groups' mean estimates did show a marginally significant interaction between condition and counterbalancing, $F(2, 91) = 3.0$, $p = .05$, and a significant main effect of counterbalancing, $F(1, 92) = 10.6$, $p < .01$, but not in the direction that the "prior beliefs" explanation of Experiment 1 would predict: If ΔP_{AB} = –.5, participants' estimates were *more* in line with our hypotheses (Figure 6). Again, we found a significant main effect of condition, $F(2, 91) = 6.0$, $p < .01$. Participants in the local-only group judged $A \longrightarrow D$ to be positive, *M* = 8.4, *SD* = 38.7, the global-only group judged it to be negative, *M* = –25.5, *SD* = 52.4, with the local + global group falling between the other two, *M* = –8.1, *SD* = 34.0 (Figure 7). A post-hoc comparison of group means revealed significant differences between all three groups, *p*s < .01.

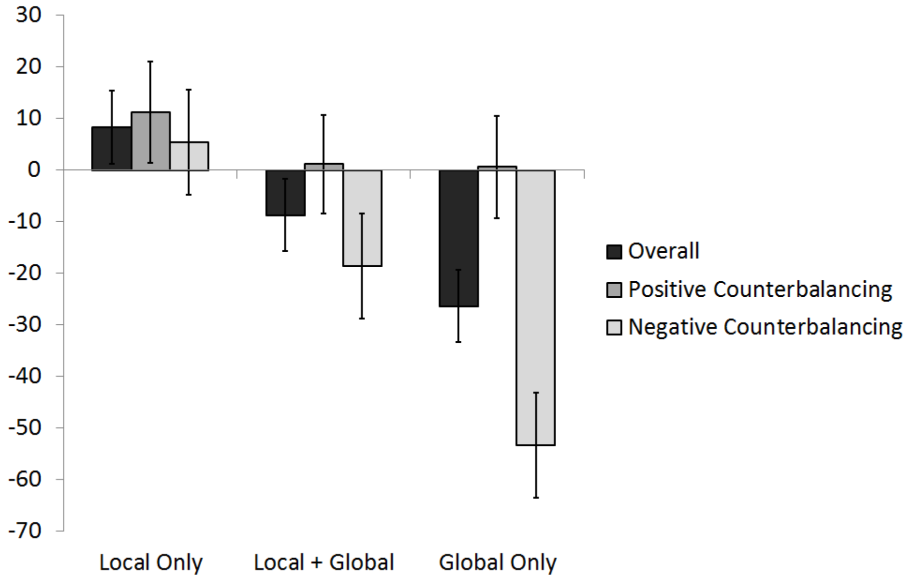


Figure 7: Mean estimates for *A* → *D* (± *SE*), Experiment 2. Estimates in the negative counterbalancing conditions were reverse-coded.

Estimates of Local Relations. Results for estimates of local relations can be found in Online Resource 1 (cf. Experiment 1).

Discussion

Experiment 2 replicates and extends the findings of Experiment 1. Even after betting on *A* → *D* three times (and receiving feedback about their performance) the local + global group failed to perform substantially better than the local-only group during betting trials, showing that the tendency to bet in line with transitivity is hard to overcome, even if it consistently leads to non-optimal outcomes. Changing the direction of the first local relation $A \rightarrow B$ ($\Delta P_{AB} = -.5$) had no substantial influence on this pattern, making it less likely that the effects in Experiment 1 are due to participants' prior beliefs about the stimulus material. Note, however, that including the counterbalancing conditions in the analysis leaves us with a rather small number of participants per cell, so these tests should be interpreted with caution. We only included them to rule out that the effect might be solely driven by participants in the positive counterbalancing condition.

Again the local + global group's betting behavior might also reflect confusion about the nature of $A \rightarrow D$ in the conditions were people may have realized a conflict between correspondence and coherence. If one is uncertain, betting 50 points on both results in each bet would be the safest bet that ensures to win at least half of the possible amount of money. However, it remains remarkable that in these conditions participants still did not seem to tend to bet on the option that has a probability of .75 of success, even after about 200 observations.

Participants' estimates of $A \rightarrow D$ show a surprising pattern in two ways: Firstly, the differences between conditions seem to be mostly due to differences in the *negative* counterbalancing condition, while mean estimates in the positive counterbalancing condition are all close to zero. This deviates from the results of Experiment 1 and Experiment 3 (see below). Secondly, in the positive counterbalancing condition, participants' estimates seem somehow detached from their betting behavior: While patterns of estimates and bets otherwise look similar within each Experiment, this is not the case in the positive counterbalancing condition in this Experiment: The global-only group in the positive counterbalancing condition bet as if they assumed $A \rightarrow D$ to be negative (cf. Figure 6), but the mean judgment for $A \rightarrow D$ was close to statistical independence. We interpret this curious result as evidence of particular participants' lack of experience with the scale. That is, participants in Experiment 1 had the chance to familiarize themselves with the rating scale before the main dependent variable was measured, whereas in Experiment 2 this was the first time the scale was presented to participants while betting was already well established. Again, as we are left with a rather small number of participants in each cell, these results that distinguish between the two counterbalancing conditions should be interpreted with caution. In any case the main effect of conditions is in line with the predictions of an influence of transitive reasoning in the local-only condition not only for betting but also for the estimates.

Experiment 3: Repeated Betting without intermittent learning phases

Goals and Hypotheses

In Experiment 3 we examined whether coherence-based integration has an influence after repeated betting on a non-transitive chain, even when participants have more opportunities to bet on $A \rightarrow D$ and are more likely to keep track of their outcomes. Again, we used the same material as in Experiments 1 and 2; however, Experiment 3 consisted only of betting trials with subsequent feedback. Where Experiment 2 replaced the judgments of Experiment 1 by betting, Experiment 3 now additionally omits the learning phases. We speculated that participants might be better able to remember past bets without intermittent learning blocks and thus keep better track of their bets. The fact that participants only learned through betting and subsequent feedback should direct their focus more toward outcomes, possibly lowering the effect of coherence.

Note that all participants saw only 24 trials over the course of the Experiments. This was partly due to the fact that a bet took longer than watching a learning trial, and that only implementing 24 bets kept the each experimental session to a reasonable length. In this regard it may seem daring to predict coherence-based effects to play a role. But as local relations were rather strong we expected that participants in the local-only and local + global group might infer strong local relations after only a few trials. In Experiment 3, participants in the local + global condition bet on $A \rightarrow D$ in a total of six trials. Finding a significant difference between the local + global group and the global-only group even after six bets would provide evidence for a strong influence of coherence.

Another addition to the previous two studies was a fourth condition. Participants in the *global-transitive* group bet in the same way as the global-only group; however, the underlying data pattern was a *transitive* chain, with $\Delta P_{AD} = .5$ ($\Delta P_{AD} = –.5$). We added this condition to test whether coherence might even have an effect on the global-only group. That is, if the two global groups differed in their bets and estimates, and the global-transitive group's results were even more in line with correspondence, this would indicate that even the global-only group might be influenced, although less strongly, by local relations.

Participants

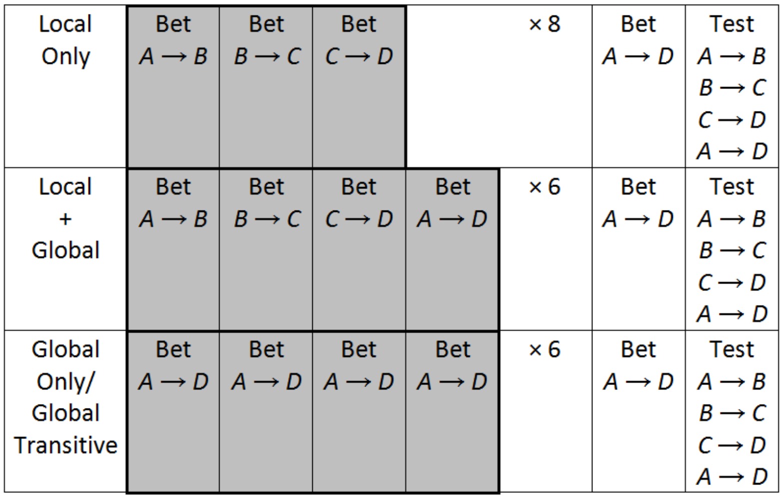
117 participants (83, age M = 23.8) who were recruited at the University of Heidelberg as part of a multi-experiment session. Participants received 6 € per hour or course credit (cf. Experiment 1) for taking part in the experiment. Regardless of the chosen form of reimbursement, participants received up to 3 € extra, depending on their betting performance.

Material and Procedure

In each trial, participants first saw (non-)occurrence of the possible cause and then were to bet on the (non-)occurrence of the effect, parallel to betting trials in Experiments 1 and 2. In each betting trial, participants bet 100 points that they could split between the two possible outcomes. After betting, they saw the whole case, followed by a slide showing their points won. At the end of the Experiment, participants were paid up to 3 € in addition to their contractual reimbursement.

Figure 8 shows the temporal structure of Experiment 3. Participants were randomly assigned to one of four conditions. In the local-only group, participants bet on local relations $A \rightarrow$

B, *B* → *C*, and *C* → *D*, eight times each. Afterward they bet on *A* → *D* once. In the local + global group, participants bet on the local relations *A* → *B*, *B* → *C*, *C* → *D*, and *A* → *D*, six times each. The global-only group and the global transitive group bet 24 times, on *A* → *D* only. Thus in all conditions participants saw the *A* → *D* relation 24 times. Again, *A* → *B* being positive or negative was counterbalanced within all conditions, resulting in positive or negative correspondence-based predictions and negative versus positive coherence-based predictions. At the end of the Experiment, participants judged the local relations and the global relation on the same scale as that used in Experiments 1 and 2.



Local Only	Bet $A \rightarrow B$	Bet $B \rightarrow C$	Bet $C \rightarrow D$		× 8	Bet $A \rightarrow D$	Test $A \rightarrow B$ $B \rightarrow C$ $C \rightarrow D$ $A \rightarrow D$
Local + Global	Bet $A \rightarrow B$	Bet $B \rightarrow C$	Bet $C \rightarrow D$	Bet $A \rightarrow D$	× 6	Bet $A \rightarrow D$	Test $A \rightarrow B$ $B \rightarrow C$ $C \rightarrow D$ $A \rightarrow D$
Global Only/ Global Transitive	Bet $A \rightarrow D$	Bet $A \rightarrow D$	Bet $A \rightarrow D$	Bet $A \rightarrow D$	× 6	Bet $A \rightarrow D$	Test $A \rightarrow B$ $B \rightarrow C$ $C \rightarrow D$ $A \rightarrow D$

Figure 8: Temporal structure of Experiment 3.

Results

Betting on the Global Relation. We compared participants' last bets on the global relation *A* → *D*. To compare participants' betting performances we first calculated how much money each participant bet on the most likely outcome given the information about *A* vs. ¬*A*.

Figure 8 shows the results for the four conditions. To be consistent with Experiment 2, we also show the results for the two counterbalancing conditions, whether *A* → *D* was positive or negative.

A 4 × 2 ANOVA showed no significant interaction between condition and counterbalancing, $F(3, 113) = 1.2$, $p = .31$, no significant main effect of counterbalancing, $F(1, 115) = 2.5$, $p = .12$, and again a significant main effect of condition, $F(3, 113) = 6.5$, $p < .01$. The local-only group ($M = 40.4$, $SD = 25.9$) and the local + global group ($M = 39.0$, $SD = 24.6$) bet less on the ideal bet than the global-only group, $M = 56.6$, $SD = 31.0$, and the global-transitive group, $M = 67.6$, $SD = 28.4$ (Figure 9), *p*s < .05.There were no significant differences between the local-only and the local + global group, $p = .85$, or between the global-only and the global-transitive group, $p = .13$.

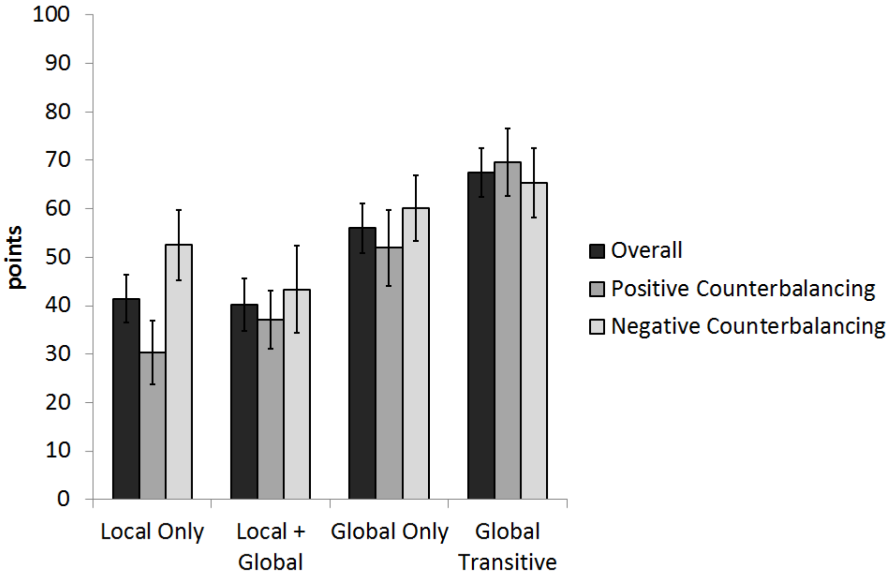


Figure 9: Mean ideal bets (points) on *D vs.* ¬*D* (± *SE*), Experiment 3.

Hence, overall the bets in the local-only condition and the local + global condition, in comparison to the global conditions, seem to have been influenced by transitive inferences, since one observes lower bets on the value that would be optimal given the observation.

Estimates of the Global Relation. We compared participants' reported estimates of the global relation $A \rightarrow D$, parallel to Experiments 1 and 2. If ΔP_{AD} =.5 participants' answers were reverse coded. A 4 × 2 ANOVA comparing the groups' mean estimates shows no significant interaction between condition and counterbalancing, *F*(3, 113) = 0.5, *p* = .72, a significant main effect of counterbalancing, *F*(1, 115) = 15.2, *p* < .01, and a significant main effect of condition, *F*(3, 113) = 8.4, *p* < .01. There were significant differences between all four groups, *ps* < .05, except for the global-only and the global-transitive group, *p* = .44 (Figure 10).

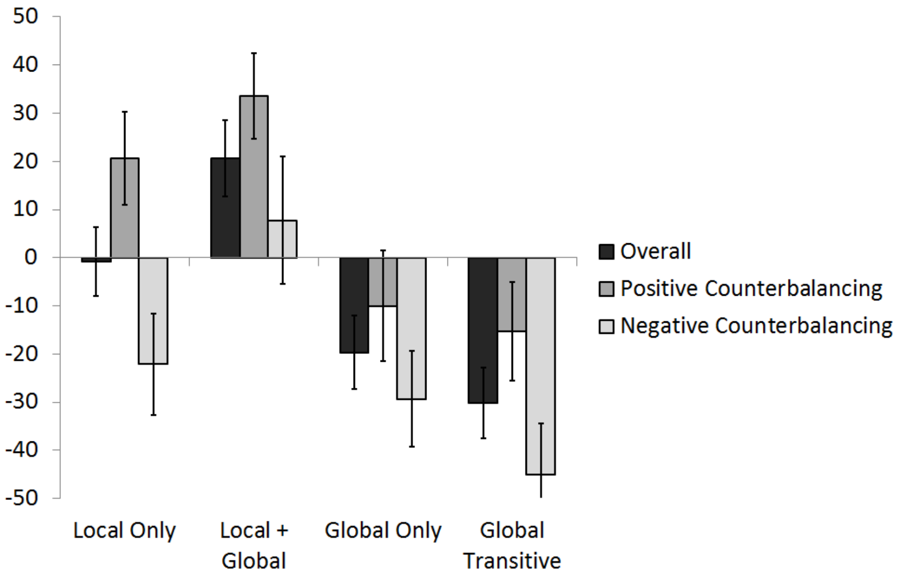


Figure 10: Mean estimates for *A* → *D* (± SE), Experiment 3. Estimates in the negative counterbalancing conditions were reverse-coded.

Estimates of Local Relations. Results for estimates of local relations can be found in Online Resource 1 (cf. Experiment 1).

Discussion

The results of Experiment 3 suggest that even in a setting where participants rely only on feedback through betting, transitive causal reasoning interferes with correspondence-based learning of intransitive causal relations. After betting on *A* → *D* six times (and receiving feedback about their performance) the local + global group failed to perform substantially better than the local-only group, showing that the tendency to bet in line with transitivity is hard to overcome, even if it leads to non-optimal outcomes. This was also reflected in participants' judgment of *A* → *D*. Again the local + global group's betting behavior might also reflect confusion about the nature of *A* → *D*, because betting 50 points on both results in each bet would be the safest bet that ensures to win at least half of the possible amount of money. Again the findings for the estimation task were less systematic. The sample size in each single counterbalancing condition might be relevant. In any case, even this dependent variable in the overall comparison of the local-only condition and global-only condition showed results broadly in line with the predictions.

General Discussion

In three studies we found compelling evidence that the causal coherence hypothesis seems to generalize to decision making in an economic context, even when using incentivized repeated betting tasks. Additionally we found that the causal coherence hypothesis does not only play a role for intransitive chains where the distal events are independent of each other, but also when the global relation strongly contradicts transitive inferences. In the experiments global relations were *strongly* negative (positive) while transitivity suggested a positive (negative) relation. Even for such strong counterevidence the results suggest that people do not seem to switch to a pure correspondence-based strategy in conditions where empirical evidence (the observed distal relation) and coherence-based inferences (independently integrating single causal relations into a chain) were inconsistent. The three main experimental groups differed considerably in their estimates of the global relations, showing that both sources of information, correspondence and coherence, play an important role in judging causal relations.

In Experiment 1 participants of the local-only group performed worse in a one-shot bet on *A* → *D*, even though they had the chance to learn about *A* → *D* in a total of 196 trials.

Experiment 2 demonstrates this tendency's strength and stability: Even after almost 200 observations and repeated bets that had led to mostly bad results for betting on transitivity, participants in the local + global condition still performed worse than the global-only group, showing no improvement over the four betting trials.

In Experiment 3 participants learned about causal relations through only a few betting trials. The betting nonetheless may have led them to focus even more on the accuracy of their predictions and on maximizing their payoffs. It should also have led participants to be more engaged during learning trials and ensured their attending to the task. Nevertheless, participants seem to have reasoned about the causal chains as if they were transitive in the local-only and the local + global group. Again, their bets were negatively influenced by transitive judgments. In the global-only group, participants still performed slightly worse than in the additionally introduced global-transitive group, possibly suggesting a (minor) influence of coherence even in the global-only group; however, this difference was not statistically significant.

The results suggest that judgments as well as bets about distal relations in a potential causal chain can be distorted in the direction implied by transitive inferences even if transitivity is violated when sequentially inducing several local relations. Depending on the conditions, we found strong distortions in the judgments and bets largely coherent with coherence-based inferences, even if direct correspondence-based inferences now went into an opposed direction while people in almost 200 trials saw evidence for this relation (Experiments 1 and 2). Even if participants' decisions lead to consistently suboptimal results and payoffs, they stick with their convictions not only when judging causal relations, but they remain willing to bet their money on it.

Participants' betting behavior showed a more consistent pattern than their estimates of $A \rightarrow D$ (cf. Experiments 2 and 3). Bets might actually be more accessible for participants than mere self-reported estimates and may thus lead to clearer results. Alternatively, participants may have got used to the betting scale by prior use, but they may have problems when suddenly having to use a bipolar causal judgment scale.

Overall the evidence provides further corroboration for the causal-coherence hypothesis claiming that people make use of a transitivity assumption (at least as a default assumption), when integrating single causal relations into larger causal chains, even if the Markov assumption is violated (cf. Cartwright, 2001, 2006; Hausmann & Woodward, 1999; Mayrhofer & Waldmann, 2015; Sober, 1988; Sober & Steel, 2014; Rehder & Burnett, 2005). We here support this idea for causal chains even for *strong and costly* evidence against the transitivity of the single relations. After early research has shown that people make a transitivity assumption *without* counterevidence (Ahn & Dennis, 2000; Baetu & Baker, 2009), it has already been shown that contradicting evidence with *zero* contingencies between distal relations can distort the judgments about distal causal relations in various conditions (von Sydow, et al. 2009, 2010, 2016). The current results suggest that even with *strongly contradicting* correlations between distal events, people are to some extent influenced by a transitivity assumption in an economic decision making context.

Arkes, Gigerenzer & Hertwig (2016) recently discussed coherence more generally (not explicity causal coherence) and argued that coherence should not be seen as a universal domain-general benchmark of rationality independent of the evolutionary context and the goals of organisms. They concede, however, that in specific domains transitive reasoning may well be adaptive and cost-effective. Applied to causal coherence, we agree that it is not enough to presuppose transitivity theoretically, but that one needs to investigate whether transitivity assumptions in specific domains do empirically hold. Transitivity assumptions in causal reasoning should be made accessible to empirical investigation. However, our results show that at least for causal chains (for other structures cf. von Sydow, et al. 2010) people seem to assume transitivity at least as a kind of default assumption, even if one is concerned with clearly intransitive relations and if assuming transitivity is costly. This supports the view that there is more to causal induction than mere correspondence. This finding is not inconsistent with the idea that coherence assumptions of rational norms need to be scrutinized cautiously, since rational norms have often been applied in too narrow and context-insensitive ways (cf. Gigerenzer, 1992; Fiedler & von Sydow, 2016, von Sydow, 2016). However, even if one could base all effects of coherence, like the shown impact of transitivity, on domain-specific adaptations, the present results at least suggest that this would be an adaptation to a *class* of cases (e.g., causal reasoning in chains).[2]

Although we do share the view that decision making in many applications needs to be transformed into 'causal decision making' (Hagmayer & Meder, 2013, Hagmayer & Sloman, 2009; Osman, 2010), the present experimental paradigm investigates and thereby questions the correctness of representational assumptions linked to Bayes net as prominent account of causal reasoning (Pearl, 2000; Spirtes, Glymour, & Scheines, 2001; Sloman, 2005; Lagnado, Waldmann, Hagmayer, & Sloman, 2007; Waldmann, 1996). Although the present research supports the assumption of a successive mental integration of single links into causal chains (Waldmann, Cheng, Hagmayer, & Blaisdell, 2008), the present line of research dissociates correspondence- and and coherence-based predictions and thereby also shows that making coherence-based inferences based on the Markov assumption can have its downsides. The potential empirical limits of causal coherence based on the Markov assumption and its pros and cons have been discussed elsewhere in more detail (von Sydow et al., 2016).

[2] One may further discuss whether either such adaptations to classes of situations are properly to be understood to refer to correspondence only, or whether one should call such phenomena adaptations in the first place (Gould & Lewontin, 1979; von Sydow, 2014). However, these issues lie beyond the scope of the present article.

We here focused on a sequence of events *A*, *B*, *C*, and *D* that may plausibly be interpreted as a causal chain. We left open whether there are only direct or also indirect relations and whether the chain is transitive or not. Von Sydow et al. (2010) have investigated not only intransitive chains but, for instance, also common-effect structures that used the same contingencies between three events. When the temporal order and story suggested a common-effect structure and not a chain the distortion effect as predicted disappeared. Nonetheless, future research needs to corroborate effects of different causal structures in more detail. Further research should also address whether the transitivity assumption also plays a role for situations in which participants can actively intervene, e.g. suggesting a company to buy or not buy (*A* vs. ¬*A*) in order to achieve rising or falling stock prices (*D* vs. ¬*D*). Active engagement in causal systems may both ensure participants' engagement in the task and effective encoding of predictions and outcomes of their decisions, eventually overcoming the assumption of transitivity in cases where it is invalid.

Acknowledgements

We are grateful to Shirin Betzler, Julia Folz, Alina Greis, Kamala Grothe, Vera Hampel, Antonia Lange, and Alexander Wendt for their valuable work during data collection. Portions of Experiments 1 and 2 were presented at the 2014 Cognitive Science conference in Quebec, Canada (Hebbelmann & von Sydow, 2014). This research and the empirical studies were supported by the grant Sy 111/2 to Momme von Sydow from the Deutsche Forschungsgemeinschaft (DFG) as part of the priority program *New Frameworks of Rationality* (SPP 1516) as well as the DFG grant FI294/23 to Klaus Fiedler.

References

Ahn, W., & Dennis, M. (2000). Induction of causal chains. *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society* (pp. 19–24). Lawrence Erlbaum Associates, NJ: Mahwah.

Andersson, P., and Nilsson, H. (2015) Do bettors correctly perceive odds? Three studies of how bettors interpret betting odds as probabilistic information. *Journal of Behavioral Decision Making*, *28*, 331–346. doi: 10.1002/bdm.1851

Arkes, H. R., Gigerenzer, G., & Hertwig, R. (2016). How bad is incoherence?. *Decision*, *3*, 20. doi: 10.1037/dec0000043

Baetu, I., & Baker, A. G. (2009). Human judgments of positive and negative causal chains. *Journal Of Experimental Psychology: Animal Behavior Processes, 35*(2), 153-168. doi: 10.1037/a0013764

Cartwright, N. (2001). What is wrong with Bayes Nets? *The Monist, 84*, 242-264.

Cartwright, N. (2006). From metaphysics to method: comments on manipulability and the causal Markov condition. *British Journal for the Philosophy of Science, 57*, 197-218. doi: 10.1093/bjps/axi156

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London B: Biological Sciences*, 205, 581-598. doi: 10.1098/rspb.1979.0086

Hagmayer, Y., & Meder, B. (2013). Repeated causal decision making. *Journal Of Experimental Psychology: Learning, Memory, And Cognition*, *39*, 33-50. doi: 10.1037/a0028643

Hagmayer, Y., Meder, B., von Sydow, M., & Waldmann, M. R. (2011). Category transfer in sequential causal learning: The unbroken mechanism hypothesis. Cognitive Science, 35, 842–873. doi:10.1111/j.1551-6709.2011.01179.x

Hagmayer, Y. A., & Sloman, S. A. (2009). Decision makers conceive of their choices as intervention. *Journal of Experimental Psychology: General*, *138*, 22-38. doi: 10.1037/a0014585

Hausman, D., & Woodward, J. (1999) Independence, invariance, and the causal Markov condition. *The British Journal for the Philosophy of Science, 50,* 521-583. doi: 10.1093/bjps/50.4.521

Hebbelmann, D. & von Sydow, M. (2014). Betting on transitivity in an economic setting. *Proceedings of the Thirty-Sixth Annual Conference of the Cognitive Science Society* (pp. 2339-2344). Austin, TX: Cognitive Science Society.

Herrnstein, R. J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior, 13*, 243-266. doi: 10.1901/jeab.1970.13-243

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs: General and Applied, 79*, 1-17. doi: 10.1037/h0093874

Johnson, S. G. B. & Ahn, W-k. (2015). Causal networks or causal islands? Represenation of mechanisms and the transitivity of causal judgment. *Cognitive Science*, 1-36. doi: 10.1111/cogs.12213

Lagnado, D. A. & Sloman, S. A. (2006). Time as a guide to cause. *Journal of Experimental Psychology, Learning, Memory, and Cognition, 32*, 451-460. doi: 10.1037/0278-7393.32.3.451

Lagnado, D. A., Waldmann, M. R., Hagmayer, Y., & Sloman, S. A. (2007). Beyond covariation: cues to causal structure. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 86–100). Oxford, England: Oxford University Press.

Lien, Y., & Cheng, P. W. (2000). Distinguishing genuine from spurious causes: A coherence hypothesis. *Cognitive Psychology*, *40*, 87–137. doi: 10.1006/cogp.1999.0724

Mayrhofer, R., & Waldmann, M. R. (2015). Agents and causes: Dispositional intuitions as a guide to causal structure. *Cognitive Science, 39,* 65–95. doi: 10.1111/cogs.12132

Osman, M. (2010) Controlling Uncertainty: A review of human behavior in complex dynamic environments. *Psychological Bulletin, 136(1),* 65-86. doi: 10.1037/a0017815

Pearl, J. (2000). *Causality: Models, reasoning, and inference.* Cambridge, MA: Cambridge University Press.

Rehder, B., & Burnett, R. C. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology, 50*, 264 –314. http://dx.doi.org/10.1016/j.cogpsych.2004.09.002

Rottman, B. M., & Hastie, R. (2014). Reasoning about causal relationships: Inferences on causal networks. *Psychological Bulletin*, *140*(1), 109. doi: 10.1037/a0031903

Sloman, S. (2005). *Causal Models: How people think about the world and its alternatives*. Cambridge, MA: Oxford University Press.

Sober, E. (1988). The principle of the common cause. In J. Fetzer (ed.), *Probability and causality* (pp. 211-228). Dordrecht: Reidel.

Sober, E. & Steel, M. (2012). Screening-off and causal incompleteness: a no-go theorem. *The British Journal for the Philosophy of Science, 64,* 1-38. doi: 10.1093/bjps/axs021

Spirtes, P., Glymour, C., & Scheines, R. (2001), *Causation, Prediction, and Search* (2nd edition). New York, NY: Springer.

Spohn, W. (2001), Bayesian Nets are all there is to causal dependence. In: M.C. Galavotti, P. Suppes, and D. Costantini (eds.), *Stochastic Dependence and Causality* (pp. 157-172). Stanford: CSLI Publications.

von Sydow, M. (2014). *‘Survival of the fittest’ in Darwinian metaphysics - tautology or testable theory?* (pp. 199-222) In E. Voigts, B. Schaff & M. Pietrzak-Franger (Eds.). *Reflecting on Darwin*. Farnham, London: Ashgate. ISBN 978-1-4724-1407-6,

von Sydow, M., Hagmayer, Y. & Meder, B., (2016). Transitive reasoning distorts induction in causal chains. *Memory & Cognition, 44*(3), 469-487. . doi:10.3758/s13421-015-0568-5.

von Sydow, M., Meder, B., & Hagmayer, Y. (2009). *A transitivity heuristic of probabilistic causal reasoning.* In *Proceedings of the Thirty-First Annual Conference of the Cognitive Science Society* (pp. 803-808). Austin, TX: Cognitive Science Society.

von Sydow, M., Meder, B., Hagmayer, Y. & Waldmann, M. R. (2010). How causal reasoning can bias empirical evidence. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2087-2092). Austin, TX: Cognitive Science Society.

Vulkan, N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys*, *14*, 101-118. doi: 10.1111/1467-6419.00106

Waldmann, M. R. (1996). Knowledge-based causal induction. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), *The psychology of learning and motivation, Vol. 34: Causal learning* (pp. 47-88). San Diego: Academic Press.

Waldmann, M. R., Cheng, P. W., Hagmayer, Y., & Blaisdell, A. P. (2008). Causal learning in rats and humans: a minimal rational model. In N. Chater, & M. Oaksford (Eds.), *The probabilistic mind. Prospects for Bayesian Cognitive Science* (pp. 453-484). Oxford: University Press.

Waldmann, M. R., Meder, B., von Sydow, M., & Hagmayer, Y. (2010). The tight coupling between category and causal learning. *Cognitive Processing*, *11*, 143–158. doi:10.1007/s10339-009-0267-x