

## The influence of causal information on judgments of treatment efficacy

JENNELLE E. YOPCHICK AND NANCY S. KIM  
Northeastern University, Boston, Massachusetts

How does the causal structure of a problem concept influence judgments of treatment efficacy? We argue that the task of evaluating treatment efficacy involves a combination of causal reasoning and categorization. After an exemplar has been categorized, a treatment task involves judging where to intervene in the causal structure to eradicate the problem, removing the exemplar from category membership. We hypothesized that the processes underlying such *category membership removal* tasks are not identical to those underlying categorization. Whereas previous experiments have shown that both the root cause (as the most generative feature) and the coherence of the exemplar heavily influence categorization, Experiments 1 and 2 showed that people base category membership removal judgments on the root cause. In Experiment 3, people spontaneously chose to remove an exemplar from category membership when asked to treat the terminal effect. We discuss how our findings are compatible with existing models of categorization. A description of pilot studies for Experiment 1 may be downloaded as supplemental materials from [mc.psychonomic-journals.org](http://mc.psychonomic-journals.org).

In the span of a lifetime, people commonly encounter some form of a problem (e.g., a mechanical problem or a disorder or disease) that requires treatment and often need to assess the efficacy of the different kinds of treatments available. For example, a person diagnosed with depression might consider whether to seek out psychotherapy or a prescription for antidepressants; another person might give an opinion about both general types of treatments to a friend. Yet despite the fact that making judgments about treatment efficacy is a ubiquitous task in the real world and, thereby, has high ecological validity, surprisingly little research in cognition has considered the question of how people make such judgments. Many kinds of information could conceivably be used in judging treatment effectiveness, including but not limited to information or misinformation about the general efficacy of broad types of treatments for a problem (e.g., psychotherapies or drug therapies), beliefs about the problem features' salience (Sloman, Love, & Ahn, 1998), rarity (McKenzie, 2006) or severity, and beliefs about the causal relationship(s) among the features of the problem (Waldmann & Holyoak, 1992).

The driving question of this article is how people's beliefs about the *causal relationships* between the features of a problem affect their judgments about the effectiveness of treatments, all else held equal. In this project, we used the problem domain of mental disorders to examine this broad question; prior work has already established that both laypeople and domain experts hold beliefs about causal relations between the symptoms, or features, of disorder concepts (Kim & Ahn, 2002a, 2002b). These rela-

tions constitute the perceived causal system underlying the concept.

At first glance, judgments about treatment effectiveness may seem to be merely a specific case of judging the likely consequences of interventions on such a causal system. Indeed, there is ample reason to believe, from the causal-reasoning literature, that adults and even children can readily predict the consequences of specific interventions on a causal system (e.g., Gopnik, Sobel, Schulz, & Glymour, 2001; Schulz, Gopnik, & Glymour, 2007; Sloman, 2005; Waldmann & Hagmayer, 2005). For example, Gopnik et al. (2001) showed children that Object A alone, but not Object B alone, activated a "blicket detector" when placed on its top. When the children were then shown an activated blicket detector with both objects on top, they reliably preferred to remove Object A to make the blicket detector stop activating. Findings by Sloman and Lagnado (2005) with adults have further shown that people told of an  $A \rightarrow B \rightarrow C$  causal chain can reliably predict what would happen to other, individual nodes in the chain when the intermediate cause (B) is intervened upon. We suggest that the case of treatment judgments extends this body of work on interventions in causal reasoning in a novel and significant way. In treatment judgments, one reasons about the consequences of interventions upon a causal system, but, importantly, that causal system also underlies a *problem* concept. As we will describe below in detail, the consequences of interventions are therefore markedly different in treatment, and treatment judgments may be guided accordingly.

---

J. E. Yopchick, [yopchick.j@neu.edu](mailto:yopchick.j@neu.edu); N. S. Kim, [n.kim@neu.edu](mailto:n.kim@neu.edu)

---

Because the causal system intervened upon in treatment also underlies a concept, predictions about people's treatment judgments can be informed jointly by the interrelated literatures on causal reasoning and on (causal) theory-based categorization. Indeed, causal reasoning cannot be divorced from causal theory-based categorization, insofar as a concept itself is represented as a causal structure. That is, at some level, the concept itself consists of the category features and the interrelations among those features (Murphy & Medin, 1985). Any interventions on the causal structure of any given category member, affecting the presence or absence of features, therefore not only will act upon the causal structure of the exemplar, but also will inescapably affect the category membership of that exemplar. We suggest that it is thereby the extra consideration of category membership that differentiates the case of treatment judgments from the case of intervening upon a causal system that does not underlie a category (e.g., Ball 1 moves Ball 2, which moves Ball 3).

From the perspective of the theory-based categorization literature, treatment judgments are particularly interesting for yet another reason. Often, in real-life treatment judgments and in the scenarios tested here, the category membership of an exemplar has previously been determined, in that the person has already been classified as having a particular disorder. Thus, treatment judgments are interesting in that the ideal goal of treatment is to eradicate the entire system, thereby removing the exemplar from category membership. We will refer to the act of intervening to move an exemplar out of a category as *category membership removal*. Although one might at first think that the category membership removal process is essentially the same as the categorization process, so that all the underlying processes involved in causal-theory-based categorization equally apply (albeit in reverse) to category membership removal, we will suggest in the remainder of this introduction that this may not be the case.

### **Influence of Causal Knowledge on (Traditional) Categorization Judgments**

To illustrate how we propose that category membership removal differs fundamentally from categorization (with respect to the theory-based approach to categorization), it is first necessary to discuss what is currently known about causal theory-based categorization. The causal relationships between the features of a concept have been found to affect categorization judgments in at least two important ways. First, Rehder and his colleagues have found robust evidence for a *coherence effect*, so that exemplars inconsistent with a category's causal laws are judged less likely to be category members (e.g., Rehder, 2003). For example, suppose that researchers have uncovered the causal relationships connecting the symptoms of Disorder X, which consist of anxiety in social situations, increased heart rate with excessive sweating, and feelings of embarrassment. Say we know that in Disorder X, anxiety in social situations *causes* an increased heart rate with excessive sweating, which, in turn, *causes* feelings of embarrassment. These two causal relationships between the features, forming a causal chain, can be described as

the causal laws underlying Disorder X. Rehder has shown that an exemplar missing the intermediate cause (e.g., increased heart rate with excessive sweating) is judged to be relatively less coherent than an exemplar missing the root cause (e.g., anxiety in social situations), since more causal laws are violated in the first case.

Second, Ahn and Rehder and their respective colleagues have also reported consistent evidence for a *root cause effect*, so that the presence of the root cause strongly affects judgments of category membership likelihood (e.g., Ahn, Kim, Lassaline, & Dennis, 2000; Rehder & Kim, 2006). This effect may be orthogonal to the coherence effect; both may occur simultaneously and contribute to the overall categorization judgment. Both Ahn and Rehder suggested that the importance people place on root causes in categorization judgments is due to their generative nature. That is, because the root cause ultimately gives rise (directly or indirectly) to all the other features of the concept, it is perceived as the most *generative* feature in the concept and, consequently, important to the concept.

In research using traditional categorization tasks, there is evidence for considerable tension between these two contributing factors in the making of categorization judgments. For example, take the case in which we contrast Patient 1, who is missing only the root cause symptom of Disorder X, and Patient 2, who is missing only the intermediate cause. Which is more likely to have Disorder X? These kinds of categorization scenarios directly pit the relative influence of the most generative feature (root cause) against the feature whose presence maximizes exemplar coherence (intermediate cause). Recent experiments have been conducted to determine the relative contributions of the two factors (readers interested in that debate should see Marsh & Ahn, 2006; Rehder & Kim, 2006, 2008). Experiments incorporating categorization tasks show, as might be expected, that people wrestle between the relative influences of the most generative cause feature (the root cause) and exemplar coherence (maximized by the presence of the intermediate cause). Indeed, previous work has suggested that the two may have a fairly comparable influence (Rehder, 2003) on categorization judgments. We highlight this case so that we can now more clearly introduce our hypothesis that the tension between exemplar coherence and feature generativity in categorization tasks should not be expected to occur in the present case of treatment judgments, or category membership removal tasks.

### **Influence of Causal Knowledge on Treatment (Category Membership Removal) Judgments**

Although the influences of both exemplar coherence and feature generativity are highly relevant to standard categorization tasks, we hypothesize that it is feature generativity that influences category membership removal judgments. We suggest that the underlying reason for the coherence effect in traditional categorization tasks is absent in the category membership removal case (e.g., in treatment judgments). In straightforward categorization judgments, coherence matters *because* a judgment is being made about the likelihood that an exemplar was generated by the category's causal laws (Rehder, 2003). Again, in

categorization tasks, if an exemplar is maximally incoherent (i.e., in the case of the causal chain, missing the intermediate feature, so that all of the causal laws of the category are violated), people judge it to be very unlikely that the exemplar is a case of Disorder X. In contrast, consider a case of treatment judgment in which we have already been told that the patient has been diagnosed with Disorder X, and in which we know that the three symptoms of Disorder X are present. If we now intervene upon the intermediate cause, eradicating it, the fact that the patient's disorder is now inconsistent with the category's causal laws does not matter in the same way as it does in categorization judgments. In this case, we already know the diagnosis, and accordingly, the decision about whether the category's causal structure initially generated the features of the exemplar has already been made. That is, we have *already accepted* that the exemplar was generated by the category's causal laws, and the underlying reason why coherence effects occur in categorization is thereby inapplicable to category membership removal.

In contrast to the categorization case, then, the sole question underlying category membership removal (e.g., treatment judgments) is how to prevent the causal laws of Disorder X from continuing to generate its symptoms. We thereby predict that the generativity of features is the most important consideration in people's treatment (category membership removal) judgments. Namely, acting upon the root cause of the disorder should be thought to most effectively eradicate *all* the features of a disorder and prevent the future generation of the features. In contrast, because the coherence effect does not apply to category membership removal in the same way as it does to categorization, the intermediate cause's relative influence should be limited to its own generative power (i.e., its ability to generate only the terminal effect). As such, the generative power of the root cause is greater than that of the intermediate cause, and a treatment acting upon the root cause should be deemed more effective than one acting upon the intermediate cause. Note that category membership removal as we have described it may apply most clearly to the domain of *problem categories* (e.g., disorders, diseases, etc.), as we will discuss in the General Discussion section.

### Overview of the Experiments

Our intuitions regarding the processes underlying category membership removal, as compared with categorization, require a systematic test. Experiments 1 and 2 were designed to test the hypothesis that the root cause, as the most generative feature in the concept, has the strongest influence on treatment (category membership removal) judgments. Again, this prediction for category membership removal judgments is in direct contrast to the fact that in regular categorization tasks, both the presence of the root cause and the intermediate cause in a chain strongly influence judgments. In all three experiments, we used disorder concepts consisting of a three-step causal chain structure of the form *Symptom X* causes *Symptom Y*, which in turn causes *Symptom Z*.

In Experiments 1 and 2, people were presented with patient exemplars already classified as having the disorder

(i.e., the three symptoms) and were given a choice between two treatments that would intervene directly upon either Symptom X or Symptom Y. If people choose treatment in accord with the root cause (Symptom X), rather than in accord with the intermediate cause (Symptom Y), this would support our intuition that the category membership removal process is driven most strongly by the fact that Symptom X is perceived to be maximally generative. Such a result would be a novel finding in the literature, in that it would be the first evidence illuminating the processes underlying (causal theory-based) category membership removal. Moreover, such a result would still be fully compatible with theoretical constructs underlying existing models of causal-theory-based categorization (e.g., Ahn et al., 2000; Rehder, 2003). If people do not show a strong preference for a treatment that will act upon Symptom X, we may infer that both coherence and root cause effects affect category membership removal as strongly as they do categorization.

Experiment 3 was designed to build upon our findings from Experiments 1 and 2, asking whether people would choose to remove the exemplar from category membership (via intervening upon the root cause) even when the explicit task was only to treat the terminal effect (and *not* explicitly category membership removal). If people's treatment judgments are influenced by a general preference to remove the exemplar from category membership, then even in this case, people should rate as most effective the treatment that would act upon the root cause, fully eradicating the problem in the exemplar. In Experiment 3 we pitted the root cause choice against the alternatives of acting either upon a node that would more *directly* influence the terminal effect (i.e., the intermediate cause) or upon the terminal effect itself.

## EXPERIMENT 1

In Experiment 1, we manipulated the location of symptoms pre-rated as being biological or physical and not psychological or environmental (*biological*, henceforth), symptoms pre-rated as being psychological or environmental and not biological or physical (*psychological*, henceforth), and symptoms pre-rated as being equally biological and psychological (*neutral*, henceforth) within the causal structure of artificial disorders. For example, in one structure, the participants were told that the root cause of a disorder was biological, its intermediate cause was psychological, and its terminal effect symptom was neutral. We predicted that if the category membership removal process was most strongly driven by a preference to prevent the generation of symptoms, then a drug treatment would be viewed as more effective than a psychotherapy treatment. This would occur insofar as the drug treatment was thought to act most directly on the biological symptom (here, the root cause) and the psychotherapy was presumed to act most directly on a psychological symptom (the intermediate cause). In contrast, if category membership removal judgments are influenced by both feature generativity and exemplar coherence, just as in categorization, then people would not clearly prefer a drug treatment over a psychotherapy treat-

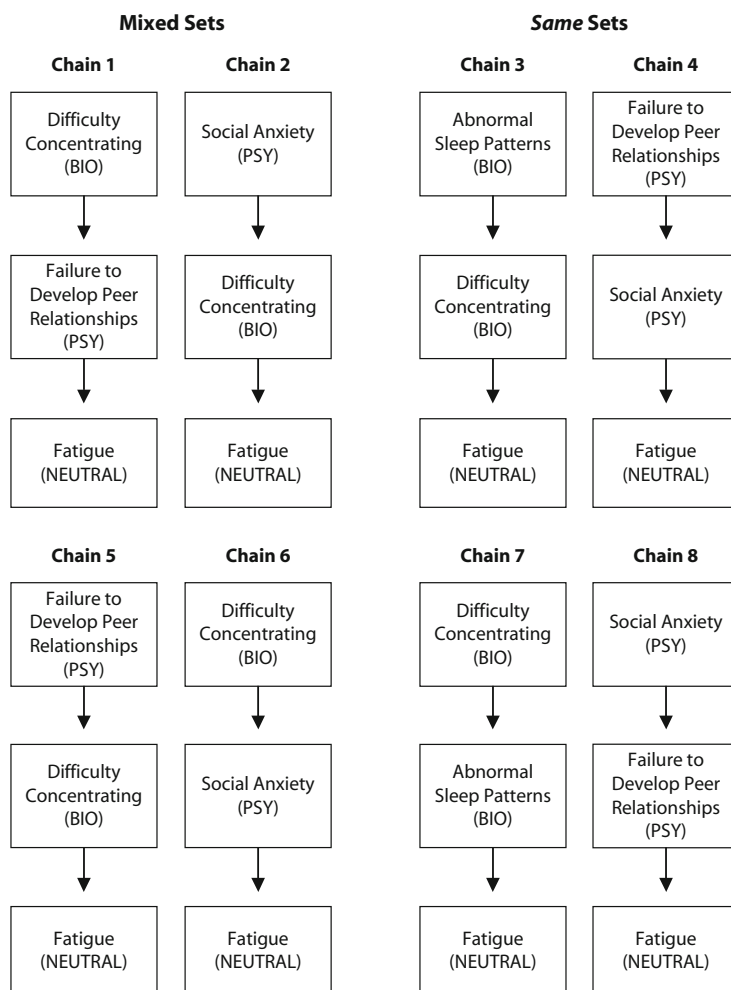
ment, because the scenario above would pit these two influences against one another.

We accounted for the possibility that factors other than causal knowledge influence people's treatment judgments (e.g., a priori beliefs about drugs vs. psychotherapy, a tendency to map treatment to the symptom read first or the symptom read most recently, etc.) by incorporating thorough counterbalancing procedures and a control condition without causal information (see below and in Figure 1).

## Method

**Participants.** Eighty Northeastern University undergraduates participated (40 in the causal condition and 40 in the control condition) in exchange for either introductory psychology course credit or \$5.

**Materials and Procedure.** We constructed four artificial disorders with three symptoms each, using real symptoms from various disorders in the *DSM-IV-TR* (American Psychiatric Association [APA], 2000). Symptoms were selected following a battery of pretests (see the online Archived Materials for more detailed descriptions and statistical results). All the symptoms were pretested and verified as, indeed, being construed by our participant population as either more biological than psychological or vice versa. Note that because all the symptoms were taken from the *DSM*, they tend to skew toward the psychological as a whole; therefore, the designations *biological* and *psychological* are relative, not absolute. Symptoms were next pretested with a separate group of participants to verify that prescription medication was indeed thought to be most helpful for the biological symptoms and counseling for the psychological symptoms; they were also pretested to equate them for seriousness/severity in the different causal structure positions (e.g., root cause



**Figure 1.** The eight different versions of causal structures in Yopiercey Disorder, each presented an equal number of times between participants. The nature of each symptom (psychological, biological, or neutral), as determined by piloting, is shown with each symptom only for illustrative purposes; these labels were not included in the causal diagrams that the participants saw in Experiment 1. Note that Chains 1–4 are identical to Chains 5–8, respectively, except that the positions of the root and intermediate causes are reversed in Chains 5–8, relative to Chains 1–4.

**Table 1**  
**Stimuli for Experiment 1**

Set Type	Item Type	Mental Disorder		
		Mikigam Disorder	Nairbick Disorder	Onyellis Disorder
Same Set 1	Psychological Cause 1	suicidal thoughts	excessive or unreasonable fear	deliberate fire starting
	Psychological Cause 2	repeated nightmares	recurrent thoughts that produce anxiety	recurrent recollections of a distressing event
Same Set 2	Biological Cause 1	hallucinations	muscle tension	motor agitation
	Biological Cause 2	insomnia	trembling and shaking	nausea and vomiting
Mixed Set 1	Psychological Cause	repeated nightmares	recurrent thoughts that produce anxiety	recurrent recollections of a distressing event
	Biological Cause	hallucinations	muscle tension	motor agitation
Mixed Set 2	Biological Cause	hallucinations	trembling and shaking	nausea and vomiting
	Psychological Cause	suicidal thoughts	recurrent thoughts that produce anxiety	recurrent recollections of a distressing event
All sets	Neutral Effect	decreased need for sleep	forgetful in daily activities	amnesia

Note—A fourth artificial mental disorder used in Experiment 1, Yopiercey Disorder, is depicted in Figure 1. These stimuli were presented to participants in diagram format as described in Experiment 1. Each cause item in each set type was shown alternately as the root cause *and* as the intermediate cause, between participants (see Figure 1).

vs. intermediate cause). Finally, all the causal chain pathways were pretested with a third pilot group to verify that our participant population deemed them to be reasonably plausible.

As was explained in the introduction to Experiment 1, the availability of causal knowledge was manipulated between participants. In the *causal condition*, the participants were told the name of the disorder, that it was a mental disorder, and that it was classified by three symptoms that were then listed (e.g., difficulty concentrating, social anxiety, and fatigue). Next, the participants read a sentence describing the causal chain for the three symptoms of the disorder, accompanied by a diagram (see examples in Figure 1). For example, they read, “Clinical researchers have found that in patients diagnosed with *Yopiercey Disorder*, *difficulty concentrating* tends to cause *social anxiety*, and *social anxiety* tends to cause *fatigue*.” Then the participants read that there were two commonly used treatments for the disorder, an artificial psychotherapy (e.g., “Wryhta therapy”) and an artificial drug therapy (e.g., “Vaxen therapy”), and that both treatments were equally readily available and cost efficient and took the same amount of time (2 weeks) to begin being effective. The participants were then prompted to rate the psychotherapy and the drug therapy for how effectively each would treat the disorder (on a scale of 0–100, where 0 = *not at all effective* and 100 = *extremely effective*). (See Figure 1 and Table 1 for the symptoms of all four artificial disorders.) The order in which the treatments were listed (drug therapy and psychotherapy) was randomized for each participant and disorder. The ratings questions for each treatment were listed in the same order as that in which the treatments were listed for that participant and disorder.

Each disorder’s three symptoms were presented in one of the following four general types of causal chains: biological → psychological → neutral (*Bio–Psy*, henceforth), psychological → biological → neutral (*Psy–Bio*, henceforth), biological → biological → neutral (*Bio–Bio*, henceforth), or psychological → psychological → neutral (*Psy–Psy*, henceforth; see Figure 1 for an example). We will refer to the *Bio–Psy* and *Psy–Bio* causal chains as the *mixed sets*, since each contained one biological symptom and one psychological symptom, and to the *Bio–Bio* and *Psy–Psy* causal chains as the *same sets*, since each contained two symptoms of the same nature (i.e., both psychological or both biological). The *same sets* were used as a manipulation check to ensure that treatment choice was, in fact, mapped onto symptom type (e.g., because the *Bio–Bio* chain contains no psychological symptoms, we would expect the drug therapy to be preferred over the psychotherapy, regardless of whether causal information was provided).

Each participant viewed all four artificial disorders once, so that each participant saw all four types of causal chains (*Psy–Bio*, *Bio–Psy*, *Bio–Bio*, and *Psy–Psy*), with each of these causal chains

presented for a different disorder. Across participants, all four artificial disorders took equal turns being presented with all four types of causal chains (see Figure 1, top row). Furthermore, we created two versions of each of the four chains for each disorder, so that all of the preceding root and intermediate causes were reversed in order in the second version. For example, compare the bottom and top rows of Figure 1, where Chain 5 has reversed causes relative to Chain 1, Chain 6 has reversed causes relative to Chain 2, and so on. This root–intermediate order reversal was also manipulated between participants, for a total of eight versions of each disorder presented across eight randomly assigned groups of participants (Figure 1). This second manipulation ensured that, across participants, each biological and psychological symptom was shown an equal number of times in the root cause position versus the intermediate position for the disorder. To recap all of the above, each participant saw all four disorders; half of the participants received the four different types of causal chains in one order, and half received them in the reverse order (i.e., the root and intermediate cause positions were reversed). Each of the eight versions of a disorder, as described above, was seen equally frequently across participants.

The *control condition* differed from the causal condition only in that participants did not learn any of the causal information and did not view the causal diagrams. However, symptoms were listed in the same order as that for the eight versions of each disorder in the causal condition.

## Results and Discussion

See Figure 2 for the results of Experiment 1. Analyses were conducted at the  $\alpha = .05$  level except as noted.

**Main analyses: Mixed sets.** A 2 (item type: *Psy–Bio* or *Bio–Psy*)  $\times$  2 (treatment type: psychotherapy or drug therapy)  $\times$  2 (condition: causal or control) ANOVA revealed that the critical three-way interaction was significant [ $F(1,78) = 8.80, p = .004, \eta^2 = .10$ ]. To further unpack the meaning of this interaction, we ran separate ANOVAs for the causal and control conditions.

A 2 (item type: *Psy–Bio* or *Bio–Psy*)  $\times$  2 (treatment type: psychotherapy or drug therapy) ANOVA on the mixed sets in the causal condition revealed that the critical interaction was significant [ $F(1,39) = 16.31, MS_e = 877.43, p < .001, \eta^2 = .30$ ; see Figure 2A]. This was the critical analysis testing our main hypothesis. There was no main effect of item type or treatment type (both  $p \geq .600$ , both  $\eta^2 \leq .01$ ).

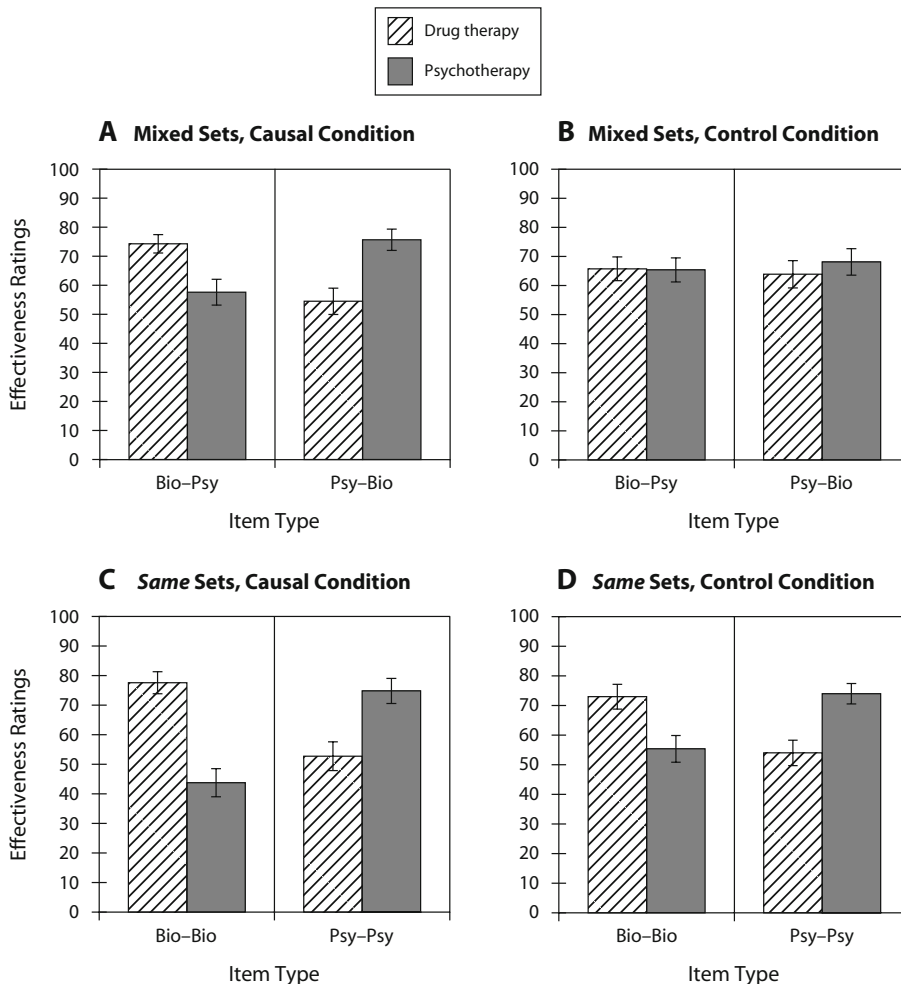


Figure 2. Results of Experiment 1.

A set of four Bonferroni-corrected planned comparisons was conducted at the  $\alpha = .0125$  level. In the Bio-Psy chains, the participants rated the drug therapy ( $M = 74.30$ ,  $SE = 3.29$ ) as more effective than the psychotherapy ( $M = 57.63$ ,  $SE = 4.72$ ) [ $t(39) = 2.79$ ,  $p = .008$ ,  $\eta^2 = .17$ ]. In the Psy-Bio chains, the participants rated the psychotherapy ( $M = 75.65$ ,  $SE = 3.89$ ) as more effective than the drug therapy ( $M = 54.50$ ,  $SE = 4.82$ ) [ $t(39) = 3.18$ ,  $p = .003$ ,  $\eta^2 = .21$ ]. The drug therapy was rated as more effective for the Bio-Psy chains than for the Psy-Bio chains [ $t(39) = 3.75$ ,  $p = .001$ ,  $\eta^2 = .26$ ], and the psychotherapy was rated as more effective for the Psy-Bio chains than for the Bio-Psy chains [ $t(39) = 3.60$ ,  $p = .001$ ,  $\eta^2 = .25$ ]. The results therefore support our hypothesis that root cause effects have the strongest influence on treatment choices for category membership removal.

We next examined the corresponding treatment ratings for the control condition to ensure that the same pattern did not occur in the absence of causal chains (see Figure 2B.) A 2 (item type: Psy-Bio or Bio-Psy)  $\times$  2 (treat-

ment type: psychotherapy or drug therapy) ANOVA revealed no main effects or interactions (all  $ps \geq .455$ , all  $\eta^2s \leq .01$ ). These results demonstrate that there was no treatment preference without causal knowledge, suggesting that the preference to map treatment onto the root cause in the causal condition was indeed based on the causal structure of the disorder and not on other factors (e.g., the fact that the symptom in the root cause position was listed first).

**Manipulation check: Same sets.** We ran identical analyses on the *same* sets to verify that the participants were indeed matching treatment type to symptom type (see Figures 2C and 2D). A 2 (item type: Psy-Psy or Bio-Bio)  $\times$  2 (treatment type: psychotherapy or drug therapy)  $\times$  2 (condition: causal or control) ANOVA revealed only a significant two-way interaction of item type and treatment type [ $F(1,78) = 57.96$ ,  $MS_e = 754.62$ ,  $p < .001$ ,  $\eta^2 = .43$ ]. Four Bonferroni-corrected planned pairwise comparisons consistently verified that treatments matched to symptom type were rated as more effective than treat-

ments mismatched with symptom type [all  $t_s(79) \geq 4.54$ , all  $p_s < .001$ , all  $\eta^2_s \geq .21$ ]. No other effects or interactions in the ANOVA neared significance (all  $p_s \geq .140$ , all  $\eta^2_s < .03$ ).

**Comparison of the mixed and same sets in the causal condition.** Finally, we conducted a 2 (item type: Bio-root [i.e., Bio-Bio or Bio-Psy] or Psy-root [i.e., Psy-Psy or Psy-Bio])  $\times$  2 (therapy type: drug therapy or psychotherapy)  $\times$  2 (set type: mixed sets or *same* sets) ANOVA on the causal condition data to compare the strength of the critical item type  $\times$  therapy type interaction in the mixed sets (Figure 2A) versus the *same* sets (Figure 2C). Although the item type  $\times$  therapy type interaction reached significance [ $F(1,39) = 42.63$ ,  $MS_e = 1,030.87$ ,  $p < .001$ ,  $\eta^2 = .52$ ], the three-way item type  $\times$  therapy type  $\times$  set type interaction did not [ $F(1,39) = 2.44$ ,  $MS_e = 672.24$ ,  $p = .127$ ,  $\eta^2 < .06$ ]. These results suggest that the mapping of therapy type to root cause did not occur less strongly for the mixed sets than for the *same* sets. Since each of the *same* sets contained two of the same types of symptom, whereas the mixed sets contained only one of each, these results testify to the relative power of the root cause in influencing judgments.

In summary, the results are strongly supportive of our hypothesis that the root cause effect is most influential in treatment (category membership removal) tasks. However, some issues of symptom content needed to be resolved, as will be discussed below.

## EXPERIMENT 2

Experiment 1 incorporated fictitious disorders and treatments, but the disorders were composed of actual symptoms from various disorders included in the *DSM-IV-TR* (APA, 2000). It could be that the findings of Experiment 1 were exaggerated because the specific content of the symptoms we chose happened to make the symptoms more compelling when in the root cause position. That is, it is possible that the content of our stimuli somehow interacted with being in the root cause position, in a way that enhanced the effect of causal structure above and beyond the effect of the structure itself. Therefore, to minimize potential confounding effects of content, we used blank property symptoms in Experiment 2 in order to more cleanly investigate the effect of causal knowledge when a treatment is chosen.

### Method

**Participants.** Ninety-six Northeastern undergraduates participated (48 causal and 48 control) for either introductory psychology course credit or \$5.

**Materials and Procedure.** The materials and procedure in Experiment 2 were nearly identical to those used in Experiment 1; the only exception was that the symptoms used were blank properties (e.g., Symptom Q) rather than *DSM-IV-TR* (APA, 2000) symptoms. Alongside each blank property, we also explicitly stated whether the symptom was biological or psychological. For example, participants might be told about a mental disorder classified by the following three symptoms: "Symptom Q, which is biological in nature," "Symptom R, which is psychological in nature," and "Symptom S, which is equally biological and psychological in nature." Thus, the

symptoms in Experiment 2 contained only enough content to indicate that they were symptoms of a mental disorder, that they were different from one another (e.g., as denoted by the Q, R, and S), and that they were of a particular nature (i.e., biological, psychological, or both equally).

### Results and Discussion

See Figure 3 for the results of Experiment 2. Analyses were conducted at the  $\alpha = .05$  level, except as noted.

**Main analyses: Mixed sets.** As in Experiment 1, there was a significant three-way interaction of item type (Psy-Bio or Bio-Psy), treatment type (psychotherapy or drug therapy), and condition (causal or control) in the mixed sets [ $F(1,94) = 8.01$ ,  $p = .006$ ,  $\eta^2 = .08$ ]. Thus, we ran separate analyses for the causal and control conditions.

In the causal condition only (Figure 3A), there were no main effects (all  $p_s \geq .334$ , all  $\eta^2_s \leq .02$ ); however, again, the critical interaction reached significance [ $F(1,47) = 16.86$ ,  $MS_e = 679.31$ ,  $p < .001$ ,  $\eta^2 = .26$ ]. All four Bonferroni-corrected paired comparisons were significant at the  $\alpha = .0125$  level [all  $t_s(47) \geq 2.66$ , all  $p_s \leq .011$ , all  $\eta^2_s \geq .13$ ], supporting our main hypothesis. The participants reliably preferred drug therapy for biological root causes and psychotherapy for psychological root causes.

In the control condition only (Figure 3B), there was no significant main effect of item type ( $p = .166$ ,  $\eta^2 = .04$ ). This time, however, there was a marginally significant main effect of treatment type [ $F(1,47) = 1.98$ ,  $MS_e = 334.83$ ,  $p = .059$ ,  $\eta^2 = .07$ ], such that people showed a marginal preference for psychotherapy ( $M = 59.66$ ,  $SE = 2.14$ ) over drug therapy ( $M = 54.55$ ,  $SE = 2.39$ ). There was also a marginally significant interaction of item type and treatment type [ $F(1,47) = 3.87$ ,  $MS_e = 161.23$ ,  $p = .055$ ,  $\eta^2 = .08$ ]. In the Psy-Bio chains, there was a reliable preference for psychotherapy ( $M = 60.35$ ,  $SE = 2.16$ ) over drug therapy ( $M = 51.65$ ,  $SE = 2.52$ ) [ $t(47) = 3.44$ ,  $p = .001$ ,  $\eta^2 = .20$ ], but this tendency was likely driven by the marginal main effect of treatment type. Although the drug therapy was also rated as marginally more effective at the  $\alpha = .0125$  level for the Bio-Psy chains than for the Psy-Bio chains [ $t(47) = 2.33$ ,  $p = .024$ ,  $\eta^2 = .10$ ], this difference was significantly larger in the causal condition ( $M = 16.42$ ,  $SE = 3.58$ ) than in the control condition ( $M = 5.81$ ,  $SE = 2.49$ ) [ $t(47) = 2.53$ ,  $p = .013$ ,  $\eta^2 = .06$ ]. Thus, overall, the marginal findings in the control condition do not diminish the support found for our hypothesis in the causal condition. Neither of the remaining two contrasts approached significance (both  $p_s \geq .552$ , both  $\eta^2_s \leq .01$ ).

**Manipulation check: Same sets.** See Figures 3C and 3D. In the *same* sets, there was a significant two-way item type  $\times$  treatment type interaction [ $F(1,94) = 112.07$ ,  $p < .001$ ,  $\eta^2 = .54$ ]. Four Bonferroni-corrected planned comparisons ( $\alpha = .0125$ ) verified that treatments matched to symptom type were rated as more effective than treatments mismatched with symptom type [all  $t_s(47) \geq 8.69$ , all  $p_s < .001$ , all  $\eta^2_s \geq .46$ ]. No other effects or interactions reached significance (all  $p_s \geq .091$ , all  $\eta^2_s \leq .03$ ).

**Comparison of the mixed and same sets in the causal condition.** Finally, we conducted a 2 (item type)  $\times$  2 (therapy type)  $\times$  2 (set type: mixed sets or *same* sets)

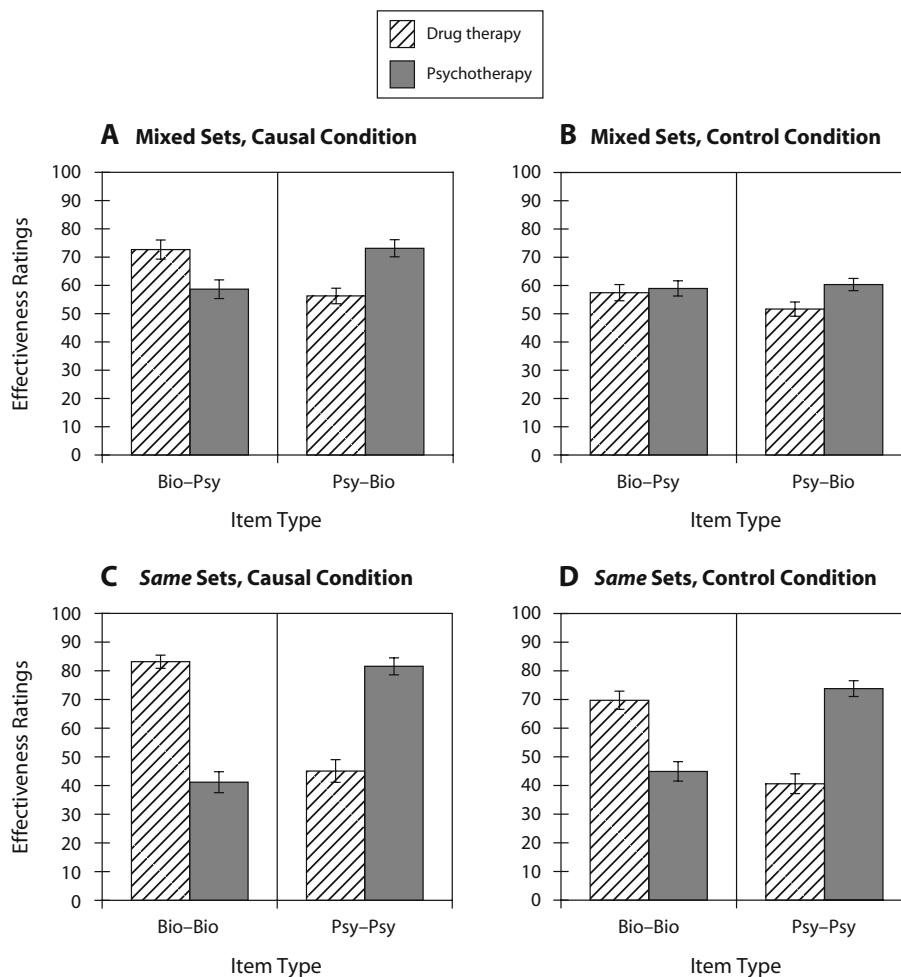


Figure 3. Results of Experiment 2.

ANOVA on the causal condition data. This time, both the item type  $\times$  therapy type interaction and the three-way item type  $\times$  therapy type  $\times$  set type interaction were reliable (both  $ps < .001$ , both  $\eta^2s \geq .34$ ). The direction of means suggests a more pronounced interaction of item type and therapy type in the *same* sets (Figure 3C) than in the mixed sets (Figure 3A). This result, which was not found in Experiment 1, likely indicates that people perceived the intermediate cause to be generative (of the terminal effect), although significantly less generative in nature, overall, than the root cause.

Overall, Experiment 2 replicated the results of Experiment 1, using blank property symptoms and thereby ruling out any undue influence of content effects that the real symptoms used in Experiment 1 might have had.

### EXPERIMENT 3

Thus far, we have presented treatment judgments as a particularly interesting case of causal reasoning, in that

they simultaneously involve intervening on a causal system and removing an exemplar from category membership. We proposed that the process underlying category membership removal can be described as selecting the intervention most likely to prevent the generation of all known features of the concept (i.e., preferring to intervene upon the root cause). Experiments 1 and 2 supported this notion, bringing us to our secondary question.

In treatment judgments, is reasoning influenced by the fact that a causal intervention will simultaneously change the exemplar's category membership? Experiments 1 and 2 were not explicitly designed to show clearly that treatment judgments differ from how intervention judgments might be made for a causal system that does *not* underlie a category. In Experiment 3, we hypothesized that the extra consideration of category membership does influence treatment judgments, so that people will prefer to fully eradicate the exemplar from *problem category* membership, even if they are not explicitly asked to do so.



For example, imagine a situation in which the explicit goal of the task is not category membership removal but, rather, the treatment of a single terminal effect feature. Will people prefer to act on the cause feature that will most directly affect the terminal effect (i.e., the intermediate node)? Will they prefer to treat the terminal effect by intervening on the root cause, fully eradicating the exemplar's category membership even when the explicit task is merely to treat the single feature? Or will they prefer to treat the terminal effect itself? If participants do not factor the category membership of the exemplar into their intervention decision, they should opt to choose the treatment that they believe will act most directly on the terminal effect symptom itself. Since the terminal symptom is always neutral (i.e., equally biological and psychological), if this is indeed how people choose treatment in this scenario, they should give equal ratings for psychotherapy and drug therapy. On the other hand, if people do prefer to fully remove the exemplar from membership in the disorder category, they should once again pick the treatment acting on the root cause.

## Method

**Participants.** Ninety-six Northeastern undergraduates participated (48 causal and 48 control) for introductory psychology course credit.

**Materials and Procedure.** The materials and procedure in Experiment 3 were identical to those in Experiment 2, except for a change in the clinical task presented to the participants. In Experiment 3, the participants read a scenario for each disorder that asked them to imagine that a particular client enters a clinician's office complaining, for example, of Symptom Q, stating that Q is an unbearable problem in his or her daily functioning. Participants then read that the client was found to also have Symptoms R and S after a clinical interview and was, therefore, diagnosed with Yopiercey Disorder. Then, as in Experiment 2, the participants read (for example) that the disorder was classified by "Symptom S, which is psychological in nature," "Symptom R, which is biological in nature," and "Symptom Q, which is equally psychological and biological in nature." In the causal condition only, the participants then also read about a causal chain connecting the three symptoms and saw an illustrating diagram, as in Experiment 2. Finally, all the participants read that the goal of treatment was to effectively alleviate the terminal effect symptom (e.g., Symptom Q) for the client, since this symptom was an unbearable problem in the client's daily functioning. All the participants rated two treatment options (drug therapy and psychotherapy) separately with respect to how effectively each would treat Symptom Q (on a scale of 0–100, where 0 = *not at all effective* and 100 = *extremely effective*).

## Results and Discussion

See Figure 4 for the results of Experiment 3. Analyses were conducted at the  $\alpha = .05$  level, except as noted.

**Main analyses: *Mixed sets.*** A 2 (item type: Psy–Bio or Bio–Psy)  $\times$  2 (treatment type: psychotherapy or drug therapy)  $\times$  2 (condition: causal or control) ANOVA revealed that the three-way interaction was significant [ $F(1,94) = 10.28, p = .002, \eta^2 = .10$ ].

In the causal condition only (see Figure 4A), the critical two-way interaction of item type and treatment type was significant [ $F(1, 47) = 11.67, MS_e = 4,172.01, p = .001, \eta^2 = .20$ ]. There were no main effects (both  $ps > .238$ ). Three of four Bonferroni-corrected planned comparisons

reached significance at the  $\alpha = .0125$  level [all  $ts(47) > 2.62$ , all  $ps \leq .012$ , all  $\eta^2s \geq .13$ ], and the fourth neared significance [ $t(47) = 1.92, p = .061, \eta^2 = .07$ ], all indicating a preference to map treatment to the root cause. Even when the immediate goal is to specifically treat the terminal symptom, it appears that complete category membership removal is perceived as the most effective means of doing so.

In the control condition only (see Figure 4B), there were no significant main effects or interactions (all  $ps \geq .739$ , all  $\eta^2s < .01$ ), indicating that in the absence of causal knowledge, people had no preference for one treatment over the other.

**Manipulation check: *Same sets.*** See Figures 4C and 4D. The expected two-way interaction of item type and treatment type was obtained [ $F(1,94) = 42.18, MS_e = 794.80, p < .001, \eta^2 = .31$ ]. Four planned comparisons verified that treatment type was matched to symptom type [all  $ts(95) > 4.35$ , all  $ps < .001$ , all  $\eta^2s \geq .17$ ].

**Comparison of the *mixed* and *same sets* in the causal condition.** A 2 (item type: Bio–root or Psy–root)  $\times$  2 (therapy type)  $\times$  2 (set type: *mixed* sets or *same* sets) ANOVA on the causal condition data showed that, as in Experiment 2, both the item type  $\times$  therapy type interaction and the three-way item type  $\times$  therapy type  $\times$  set type interaction were reliable (both  $ps \leq .005$ , both  $\eta^2s \geq .16$ ). Once again, the direction of means suggests a more pronounced interaction of item type and therapy type in the *same* sets (Figure 4C) than in the *mixed* sets (Figure 4A).

In summary, Experiment 3 showed that people preferred to map treatment to the root cause even when asked to alleviate a single symptom not directly connected to it.

## GENERAL DISCUSSION

In three experiments systematically controlling for prior knowledge, we found that people base judgments of treatment effectiveness on the causal structure of a disorder concept. Specifically, if the root cause is psychological, psychotherapy is seen as most effective; if it is biological, drug therapy is seen as most effective. Artificial disorders and treatments were used in all three experiments to minimize confounds of previous knowledge. We also removed any preceding familiarity with individual symptoms by using blank properties in Experiments 2 and 3. The fact that the same results were obtained for causal structures containing minimal content suggests, interestingly, that people need not know detailed mechanisms to make judgments about treatment in accord with the causal structure (see also Rozenblit & Keil, 2002). In all the experiments, the participants learned that both treatments took equal amounts of time to begin being effective and that they were equally cost effective. Thus, people's use of causal information was the reason for the difference in treatment effectiveness ratings between the causal and the control conditions in all three experiments.

The results of Experiments 1 and 2 further suggest that the process underlying treatment judgments involves intervening upon an exemplar in a way that will best prevent the causal generation of all the category features in

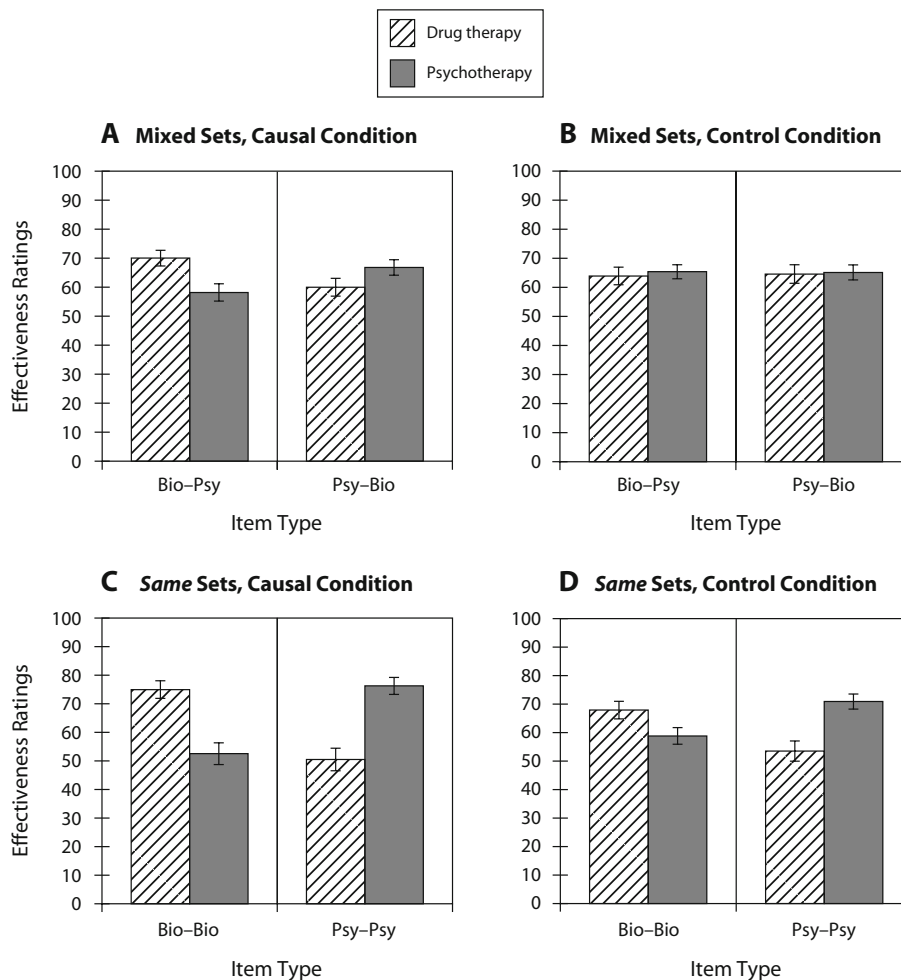


Figure 4. Results of Experiment 3.

that exemplar. We suggested earlier that, simultaneously, such an intervention removes the exemplar from category membership. Experiments 1 and 2 showed that people choose to intervene upon the root cause feature, which gives rise to all of the category's features. This overriding root cause preference in the category membership removal process can be contrasted with the case of categorization. Again, in categorization, it has previously been established that both feature generativity (maximized in the root cause) and exemplar coherence (maximized by the presence of the intermediate cause) play important roles. Coherence is important in categorization *because* it helps people judge the likelihood that an exemplar was generated by the causal structure underlying the category (Rehder, 2003). In contrast, in category membership removal, the exemplar has already been diagnosed into the disorder category. Thus, the decision about which category's causal structure initially generated the features of the exemplar has already been made. The causal relations in

the structure are thereby presumed to remain operative, and the disorder can be fully eradicated only by intervening on the root cause to stop the generation of symptoms. In category membership removal, the intermediate cause's relative influence on treatment judgments is limited to its own generative power (i.e., its ability to generate only one feature, the terminal effect).

The results of Experiment 3 suggest that even when people are not told to eradicate all the disorder features in the exemplar, but only to eradicate a terminal effect feature, they nonetheless elect to fully remove the exemplar from the disorder category by treating the root cause. This occurs even when people have the reasonable alternatives of acting upon a feature (the intermediate cause) that would more directly affect the terminal effect and of acting upon the terminal effect itself. We take these results to indicate that consideration of the category's membership in the disorder category does influence treatment judgments. Namely, people could easily

choose to act directly upon the terminal effect in order to meet the requirements of the task, but instead they show a preference for completely removing the exemplar from category membership by treating the most generative symptom in the concept (the root cause).

We speculate that this general root cause preference is dependent on the nature of the root cause itself—that is, insofar as the root cause is thought to be generative of all its effects. For example, even if a virus launches an incredibly long causal chain of consequences in the body, it is likely that people will nonetheless believe in the desirability of an antiviral treatment as long as the virus is still actively giving rise to those consequences or can be expected to become generative again in the future. If the root cause that already launched its effects is no longer expected to be active, people may then be content to intervene upon a more peripheral node. Similarly, even if the root cause is hidden, unobservable, and not itself distressing, we would predict that people would still choose to treat it if it will continue to be generative of the other features of the problem. Even in the case of a root cause that is virtually untreatable, as in the case of permanent brain damage, we suspect that people would still believe that treating the brain damage would be *ideal* (if only it could be done), as long as it could still cause its effects.

There is an alternative interpretation of our results that has not yet been considered. It is possible that people first categorized the exemplar into a broader, more general category (e.g., *psychological disorder* or *biological disorder*) on the basis of the root cause and then simply picked a treatment that matched this broader category. Although our present data cannot completely rule out this possibility, we note that in Experiments 2 and 3, the root cause was *not* the sole influence on treatment judgments, as evidenced by a significantly stronger interaction in the *same* sets causal condition than in the *mixed* sets causal condition (see Figure 3C vs. 3A and Figure 4C vs. 4A). The idea of a more general categorization also seems relatively less plausible overall as an explanation for the results of Experiment 3. In Experiment 3, people were explicitly instructed to treat the terminal effect symptom. It is therefore unclear why any broader categorization would be seen as necessary in this case. Furthermore, because the terminal effect was described as equally psychological and biological in nature, it is unclear why people would choose to map treatment to the root cause, *unless* they were concerned about the regeneration of symptoms from the root cause, as we have argued. Nonetheless, future work will be needed to more closely examine this alternative interpretation, particularly with respect to Experiment 1.

### Implications for Theory-Based Categorization

We suggest that the task of judging treatments is related to, but distinct from, the well-known *transformation* paradigm (see Keil, 1989; Rips, 1989). In such tasks, the general intuition held by adults and older children is that category membership can change for artifacts, but not for natural kinds, when surface features are changed. For example, adults and older children believe that melting a

coin and remolding it to be used as a key *does* turn the coin into a key but that painting a raccoon black and white and putting smelly stuff on its tail does *not* make it a skunk. The transformation task is similar to our treatment tasks in that both involve reasoning about what would happen to the causal system underlying a concept if a feature were intervened upon and about how these changes to the causal system would affect category membership. However, the transformation task involves reasoning about what would happen to the causal system of the concept if a feature in the causal system is changed to a different feature, rather than if it were eradicated or eliminated. Here, changing the root cause feature to a completely different feature may be perceived to lead to the generation of a completely different category's causal system. For example, changing the genetic composition of a skunk might change it to a raccoon, insofar as the newly created raccoon genes are expected to generate all the other features of a raccoon. In contrast, in treatment tasks, eliminating the root cause feature will lead to the elimination of the causal system and removal from category membership. For example, in treatment tasks, people are not judging what would happen if they *changed* the influenza virus (i.e., the root cause) into HIV in an exemplar; they are judging what would happen to the overall problem if they *eliminated* the influenza virus in an exemplar. Of course, the possibility remains that people think of treatment as category membership removal *and* recategorization into a category of *normal*. This possibility is outside the scope of the present experiments.

Some outstanding questions also remain to be examined with respect to how our understanding of people's reasoning about interventions on causal systems can inform our understanding of causal-theory-based categorization processes. One such question is how people use interventions, and feedback from those interventions, to assist in making accurate categorization decisions. For example, when a physician is unsure of the correct medical diagnosis for a patient, the physician might try a treatment and see whether or not the symptoms disappear, only retroactively making a diagnosis. This, too, is a common diagnostic situation with high ecological validity. Such a scenario fuses causal theory-based categorization with the use of interventions to discern an underlying causal structure (e.g., Gopnik et al., 2004).

Another question is how our findings may potentially build upon the *category use effect* (e.g., Ross, 1997, 1999), in which features associated with a particular use influence categorization decisions more strongly than do other features. In one study by Ross (1997), participants learned about two novel disease concepts, each of which consisted of four equally predictive symptoms; two of the symptoms for each disease additionally predicted which of two treatments to use. The results showed that features relevant to both categorization *and* category use (e.g., treatment) were categorized more accurately than features relevant only to categorization (Ross, 1997). Such findings, in conjunction with those of the present experiments, suggest that people may also be more *accurate* at classifying exemplars containing root cause features than at classifying those that do

not: As we have argued, people may judge root cause features to be maximally relevant to how the category is used, insofar as intervening upon root cause features is useful in eradicating or changing category membership. This possibility awaits a systematic investigation.

### The Domain of Mental Disorders

Finally, there are at least four interesting ramifications of having focused on the disorder domain in the present experiments. First, it leads us to consider the possibility that the ideal of eradicating category membership may be somewhat specific to categories of *problems* (e.g., disorders, diseases, computer problems, car problems). For example, it is highly unlikely that anyone would ever attempt to make a dog no longer a dog or a tulip no longer a tulip; neither does it seem that such goals would be considered desirable or ideal. We speculate broadly that the explicit goal of *changing* category membership may be particularly relevant for artifacts, the goal of *eradicating* category membership may be relatively specific to problem categories, and neither changing nor eradicating category membership may often be applicable to natural kinds.

Second, we have suggested that the generative aspect of the root cause is the reason for people's preference for treating it. It is worth considering whether assumptions about the generative nature of cause features might be relatively domain dependent. In disorders and diseases, people may have the general, default expectation that a cause feature of the disorder or disease will continue to generate other features downstream unless something intervenes. Even many diseases that "run their course" come to an end only because of intervention via the body's immune system response (note that many typically short-lived diseases, such as the common cold, can persist for extremely long periods of time if the body has an immunodeficiency, such as AIDS). In contrast, the cause features of many natural kind categories may not as frequently be expected to continuously generate features downstream. For example, once a raccoon's genes have guided it through the normal path of development, the resulting adult raccoon is not expected to continuously regenerate its features (e.g., the generation of its arms and legs is expected to occur only that one time during development). Obviously, one can think of exceptions to these general patterns, but it is nonetheless interesting to speculate whether people might have broad default assumptions about the generativity of features based on domain.

In addition, the domain of mental disorders is particularly interesting with respect to what the present findings might suggest about lay theories of the relation between psychological and biological constructs (e.g., mind and body). For example, people might hold broad framework theories stipulating that either the biological generally gives rise to the psychological or vice versa. In our experiments, participants in the causal condition always saw causal chains in which biological symptoms caused psychological symptoms, and also chains in which psychological symptoms caused biological symptoms. Prior research has shown that people discount causal information when they consider it to be incompatible with their framework

theory of that domain (Ahn et al., 2000). Thus, if people had a broader framework theory specifying that one type of symptom should generally cause the other in a single direction (e.g., that the biological causes the psychological), we should have found causal theory effects only in corresponding items. In contrast, people accepted the plausibility of both directions of causality (see Figures 2–4).

Finally, we believe that the present experiments, centered around an ecologically valid task, may be useful from an applied perspective. Even people who are not clinicians may often play a role in the choice of their own treatment. For example, they may decide which type of health professional to visit (e.g., a physician for a prescription or a therapist for psychotherapy), which referrals to follow on and which to ignore, which treatment recommendations to follow or not (e.g., which prescriptions to fill or, once filled, to take), or whether to seek out a second opinion on treatment. In many cases, these kinds of decisions are likely to be influenced by people's judgments about the probable effectiveness of given treatments. Cognitive experiments such as those reported here may, consequently, have potential implications for developing future studies of treatment compliance (Delaney, 1998; National Institute of Mental Health, 2006).

### AUTHOR NOTE

We thank John Coley, Leontien de Kwaadsteniet, Joanne Miller, Andreea Patalano, Brian Ross, Nancy Snyder, and three anonymous reviewers for very helpful comments on earlier drafts of the manuscript, and Daniel Paulus for proofreading the manuscript. Experiments 1 and 2 were presented at the 2007 Annual Meeting of the Psychonomic Society, Long Beach, CA. Correspondence concerning this article should be directed to either J. E. Yopchick or N. S. Kim, Department of Psychology, Northeastern University, 125 Nightingale Hall, 360 Huntington Avenue, Boston, MA 02115-5000 (e-mail: yopchick.j@neu.edu or n.kim@neu.edu).

### REFERENCES

- AHN, W., KIM, N. S., LASSALINE, M. E., & DENNIS, M. (2000). Causal status as a determinant of feature centrality. *Cognitive Psychology*, *41*, 361-416.
- AMERICAN PSYCHIATRIC ASSOCIATION (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Washington, DC: Author.
- DELANEY, C. (1998). Reducing recidivism: Medication versus psychosocial rehabilitation. *Journal of Psychosocial Nursing & Mental Health Services*, *36*, 28-34.
- GOPNIK, A., GLYMOUR, C., SOBEL, D. M., SCHULZ, L. E., KUSHNIR, T., & DANKS, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, *111*, 3-32.
- GOPNIK, A., SOBEL, D. M., SCHULZ, L. E., & GLYMOUR, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, *37*, 620-629.
- KEIL, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.
- KIM, N. S., & AHN, W. (2002a). Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *Journal of Experimental Psychology: General*, *131*, 451-476.
- KIM, N. S., & AHN, W. (2002b). The influence of naive causal theories on lay concepts of mental illness. *American Journal of Psychology*, *115*, 33-65.
- MARSH, J. K., & AHN, W. (2006). The role of causal status versus interference links in feature weighting. In R. Sun (Ed.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 561-566). Mahwah, NJ: Erlbaum.

- MCKENZIE, C. R. M. (2006). Increased sensitivity to differentially diagnostic answers using familiar materials: Implications for confirmation bias. *Memory & Cognition*, **34**, 577-588.
- MURPHY, G. L., & MEDIN, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, **92**, 289-316.
- NATIONAL INSTITUTE OF MENTAL HEALTH (2006). *Research on adherence to interventions for mental disorders*. Retrieved April 28, 2008, from grants.nih.gov/grants/guide/pa-files/PA06-324.html.
- REHDER, B. (2003). Categorization and causal reasoning. *Cognitive Science*, **27**, 709-748.
- REHDER, B., & KIM, S. (2006). How causal knowledge affects classification: A generative theory of categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **32**, 659-683.
- REHDER, B., & KIM, S. (2008). The role of coherence in causal-based categorization. In V. Sloutsky, B. Love, & K. McRae (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 285-290). Mahwah, NJ: Erlbaum.
- RIPS, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21-59). Cambridge: Cambridge University Press.
- ROSS, B. H. (1997). The use of categories affects classification. *Journal of Memory & Language*, **37**, 240-267.
- ROSS, B. H. (1999). Postclassification category use: The effects of learning to use categories after learning to classify. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 743-757.
- ROZENBLIT, L. R., & KEIL, F. C. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, **26**, 521-562.
- SCHULZ, L. E., GOPNIK, A., & GLYMOUR, C. (2007). Preschool children learn about causal structure from conditional interventions. *Developmental Science*, **10**, 322-332.
- SLOMAN, S. A. (2005). *Causal models: How people think about the world and its alternatives*. Oxford: Oxford University Press.
- SLOMAN, S. A., & LAGNADO, D. A. (2005). Do we "do"? *Cognitive Science*, **29**, 5-39.
- SLOMAN, S. A., LOVE, B. C., & AHN, W. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, **22**, 189-228.
- WALDMANN, M. R., & HAGMAYER, Y. (2005). Seeing versus doing: Two modes of accessing causal knowledge. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **31**, 216-227.
- WALDMANN, M. R., & HOLYOAK, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General*, **121**, 222-236.

#### SUPPLEMENTAL MATERIALS

Descriptions of the pilot studies may be downloaded as supplemental materials for this article from [mc.psychonomic-journals.org/content/supplemental](http://mc.psychonomic-journals.org/content/supplemental).

(Manuscript received May 23, 2008;  
revision accepted for publication September 5, 2008.)