

The conceptual centrality of causal cycles

NANCY S. KIM

Northeastern University, Boston, Massachusetts

CHRISTIAN C. LUHMANN

State University of New York, Stony Brook, New York

AND

MARGARET L. PIERCE AND MEGAN M. RYAN

Northeastern University, Boston, Massachusetts

How do causal cycles affect judgments of conceptual centrality? Generally, a feature is central to a concept to the extent that other features in the concept depend on it, thereby rendering it immutable from the concept (Sloman, Love, & Ahn, 1998). Previous research on conceptual centrality has focused primarily on features involved in four major types of dependency structures: simple cause–effect relations, causal chains, common-cause structures, and common-effect structures. Causal cycles are a fifth type of dependency structure commonly reported in people's real-life concepts, yet to date, they have been relatively ignored in research on conceptual centrality. The results of six experiments suggest that previously held assumptions about the conceptual representation of cycles are incorrect. We discuss the implications of these findings for our understanding of theory-based concepts.

Our background knowledge significantly influences the ways in which we perceive, categorize, reason, and make decisions in the world (Murphy, 2002). Recently, a number of researchers have argued that it is the *causal* element in our background knowledge that makes it particularly useful to us (Anderson & Lindsay, 1998; Keil, 2006). Causal and explanatory background knowledge can guide our causal learning (Waldmann & Holyoak, 1992; Waldmann, Holyoak, & Fratianne, 1995), concept formation (e.g., Murphy, 2002), inductive reasoning processes (e.g., Medin, Coley, Storms, & Hayes, 2003; Medin, Lynch, Coley, & Atran, 1997), and decision making (e.g., Pennington & Hastie, 1988, 1992). In categorization, causal knowledge also influences the conceptual centralities of individual features (e.g., Ahn, Kim, Lassaline, & Dennis, 2000; Sloman, Love, & Ahn, 1998) and the coherence among features (e.g., Rehder, 2003b). Of particular relevance to the present study is that an individual feature is conceptually central, or important, to a concept to the extent that other features of the concept depend on it (Sloman et al., 1998).

Most researchers on the influence of causal knowledge on the conceptual centralities of individual features have focused primarily on the influence of acyclic causal knowledge (e.g., simple cause–effect relations, causal chains, common-cause structures, common-effect structures, or combinations thereof). However, in studies investigating people's own real-life causal theories, researchers (e.g., Haggmayer & de Kwaadsteniet, 2008; Kim & Ahn, 2002a,

2002b; Rein, Love, & Markman, 2007; Sloman et al., 1998) have found that, in addition to these kinds of causal structures, people also commonly report *causal cycles*. In previous research asking laypeople to report their causal theories of concepts, the majority of the participants spontaneously reported causal cycles (66.7% of participants considering natural kinds and artifacts, Kim, 2005; 65.0% of laypeople considering mental disorders, Kim & Ahn, 2002b).

Indeed, the ease with which we can understand causal cycles is readily apparent when considering some realistic examples. One may easily observe cycles in a range of situations—for example, that a child's poor performance in school tends to cause him to have chronic insomnia and also that having chronic insomnia tends to make him perform poorly in school. In this case, one's overall belief is that he has become caught in a causal cycle of poor schoolwork and insomnia. Similarly, one's concept of extreme prejudice might include the belief that hate group leaders create hate group followers (e.g., by means of persuasive rhetoric), as well as the belief that hate group followers (e.g., by means of their approval and support) place hate group leaders in their positions of power. In one's concept of a typical American family, one might believe that teenagers' acts of rebellion tend to cause parents to enforce strict rules and also that parental enforcement of strict rules tends to cause teenagers to act rebelliously.

In the present study, we investigated how causal cycles influence judgments of the conceptual centralities of in-

N. S. Kim, n.kim@neu.edu

dividual features. As an initial step, we focused on the simplest form of causal cycle—the two-feature feedback loop (e.g., $A \leftrightarrow B$)—and compared features involved in such structures with features involved in acyclic structures (e.g., causal chains, common-cause structures). Of the existing process models of causal theory-based categorization, the only model explicitly designed to handle cycles is the centrality model of Sloman et al. (1998).¹ The present experiments therefore provide the first direct test of its treatment of causal cycles.

Prestige Centrality

Sloman et al. (1998) defined a feature's *conceptual centrality* as the extent to which other features depend on its presence. For example, suppose a clinician believes that feelings of worthlessness cause suicidal thoughts in major depression. According to Sloman et al., the feature *feelings of worthlessness* is thereby more central to this clinician's concept of major depression than *suicidal thoughts* (if all else is held equal). Consequently, this clinician would be expected to judge a client with suicidal thoughts but not feelings of worthlessness as less likely to have depression than a client with feelings of worthlessness but not suicidal thoughts, if all else is held equal.

Such a definition of *centrality* corresponds to what are generally known as *prestige measures of centrality* (Wasserman & Faust, 1994). Prestige measures of centrality allow centrality to recursively propagate through the network, thereby providing a richer, more detailed picture of centrality. Prestige measures of centrality are based generally on the notion of *eigenvector centrality* proposed by Bonacich (e.g., 1972; Bonacich & Lloyd, 2001) and have been used to explain phenomena observed in large social networks (e.g., disease transmission, patterns of social status) and form the basis for Google's Web page ranking algorithm (Borgatti, 1995; Ipsen & Wills, 2006). Sloman et al. (1998) computed their prestige centrality measure using an iterative multiplication of the matrix describing the interfeature dependencies (see Appendix A for details). The output of this process is an ordinal ranking of the category's features based on their centrality. Centrality, computed in this way, has been shown to successfully predict not only conceptual centralities for features involved in acyclic causal structures (see also Kim & Ahn, 2002a, and Sloman & Ahn, 1999) but also patterns of property induction (Hadjichristidis, Sloman, Stevenson, & Over, 2004). Most important for our present purposes, Sloman et al.'s centrality model and prestige centrality measures in general have the ability to compute the conceptual centralities of individual features involved in causal cycles.

As will be seen, the predictions of prestige centrality measures are strongly dependent on the exact representation to which they are applied. Previous work on Sloman et al.'s (1998) centrality measure has always assumed that reasoners represent category structures exactly as presented to them, so a two-feature cycle would be represented as is shown in Figure 1A. In psychological terms, such a representation implies that the cycle features cause each other *constantly* and *continuously* (i.e., at every possible given point in time). We will refer to this representa-

tion of cycles as a *true* cycle representation. To illustrate the consequences of using a true cycle representation, we computed the predicted prestige centralities for the causal structures depicted in Figure 1A.² The small balloons depict the rank-ordered centralities, where Rank 1 corresponds to the most conceptually central feature across the three structures (true causal cycle, causal chain, common cause). As was previously described, features generally tend to be more conceptually central than their dependents. However, the critical question is: How conceptually central are cycle features relative to features in acyclic structures, holding causal dependency strengths equal?

As can be seen in Figure 1A, in a true cycle representation, the cycle features are predicted to be more conceptually central than any feature in the acyclic structures depicted below. Conceptually, this should make sense. According to prestige centrality measures, a feature's centrality depends on the number and centralities of its dependents. The more dependents a feature has, and the more central those dependents are, the more central that feature itself becomes. In a two-feature cycle, each feature has one dependent feature whose centrality is determined by the first feature. The two features in the cycle thereby mutually reinforce each other's conceptual centrality to become more and more central. Thus, according to prestige centrality measures, the features involved in the true cycle representations should be more central to the concept than any of the features in the acyclic structures, if all else is equal.

Simplified Cycles:

A Representational Hypothesis

Again, the predictions that we derived above presumed that people hold a veridical representation of the causal structures presented to them, a default assumption of the Sloman et al. (1998) model and, indeed, of all causal theory-based models of categorization. However, there are at least two reasons to believe that this assumption may not hold for causal cycles. First, we note that a true representation of causal cycles (A causes B , B causes A) is unrealistic, in that cycle features often do not cause each other constantly and simultaneously but, rather, unfold over time. For example, we previously discussed a causal cycle in which insomnia and poor school performance caused each other. It seems unlikely that people would think that a student's school performance is actually deteriorating as he or she sits up at night. Instead, it seems more realistic that the student's insomnia on Monday night leads to poor performance on Tuesday, which would then lead to a sleepless Tuesday night, and so on. This example demonstrates how causal cycles might be more realistically represented as causal chains that play out over time.³ Such a chain would allow Feature A to influence Feature B at Time 1, Feature B to influence Feature A at Time 2, and so on. Allowing this chain to extend infinitely into the future would represent the ongoing interplay of the causal cycle.

Second, although this unfolded representation of cycles is now more realistic, it is still psychologically implausible that people will represent causal chains extending to infinity. For example, people's explanatory theories may be less

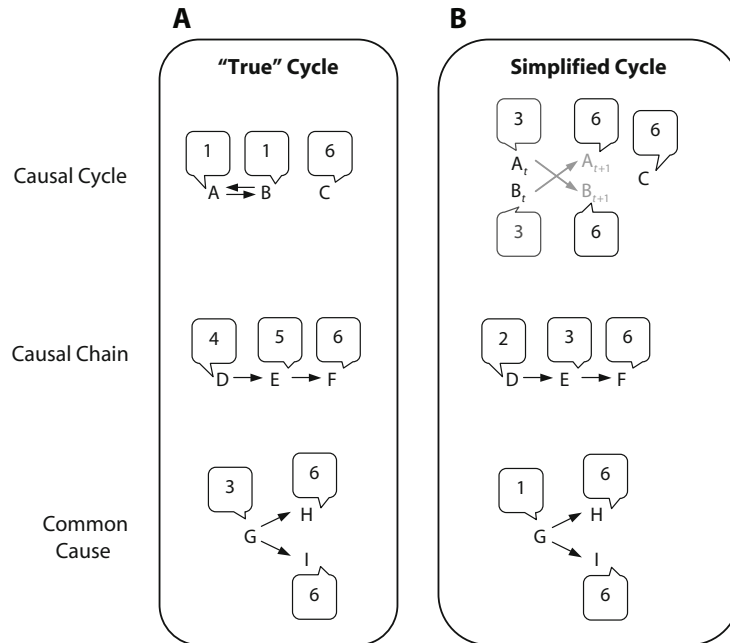


Figure 1. Schematics of the causal structures tested in Experiment 1, where letters refer to the features of a concept (see Appendix B for the actual features used). Different letters are used in this figure for ease of explanation only; in the actual materials, the items were rotated so that each set of three features took a turn playing each of the different structures. Also, the order of features and the causal arrow directions were reversed in one condition. Balloons contain prestige centrality rankings (alpha centrality, assuming $\alpha = .5$) for “true” cycle (A) versus simplified cycle (B) representations of the cyclic causal structure. Numbers indicate ranks, where Rank 1 corresponds to the most conceptually central feature. See Appendix A for more details on the model and this simulation.

complete and more fragmented than they think (diSessa, 1988, 1993; Mills & Keil, 2004; Rozenblit & Keil, 2002); people do not tend to represent or retain the full details of the explanatory, causal mechanisms underlying real-life concepts. More to the point, “people’s explanations tend to be only as specific as they need to be” (Waxman, Medin, & Ross, 2007, p. 296, n. 3). This finding (Rozenblit & Keil, 2002) is relatively restricted to explanations, in that it is strongest for explanatory theories and weaker for other knowledge types (e.g., facts, narratives). In related work, Dweck, Chiu, and Hong (1995) argued that people’s naive theories provide a general framework of beliefs that guide their attributions and behaviors but that these naive theories are not necessarily well articulated. Instead, it may be more plausible that reasoners might omit unnecessary details of the full-blown causal cycle, perhaps even without knowing that they are doing so. If the participants were reasoning over a causal structure that differed from what we assumed in our simulations, it seems quite reasonable that the resulting centrality judgments would also differ from our predictions.

Thus, as a plausible alternative representation of cycles, we present the hypothesis that people represent causal cycles as causal chains extending one step into the future (Figure 1B). This simplified representation captures the

cycle as simply as possible while still maintaining the essential nature of the cycle. Specifically, this representation captures the crucial idea that A causes B and vice versa while simultaneously simplifying the representation to the greatest extent possible (i.e., a single step). Applying the prestige centrality measures to this structure endows the cycle features with the same conceptual centrality as any other feature that has exactly one dependent (e.g., the intermediate feature of the three-feature chain). We will refer to this proposal regarding the representation of cycle features in prestige centrality calculations as the *simplified cycle hypothesis*.

The following experiments (see Table 1 for an overview) are designed to test whether reasoners represent causal cycles as true cycles per se, as was inherently assumed by Sloman et al.’s (1998) centrality theory, or whether reasoners instead represent causal cycles as the simplified representations that we have suggested. In Experiment 1, we compared judgments of conceptual centrality for features involved in cycles versus acyclic structures using blank-property, artificial disorder concepts for complete control over content and a full forward/backward counterbalancing of presentation order. In Experiments 2–6, we contrasted cycles with simple cause–effect relations (e.g., $C \rightarrow D$) within the same concept and using two different

Table 1
Overview of Experiments 1–6

| Experiment | Comparison | Issue(s) Addressed | Results |
|------------|-------------------------------------|---|----------|
| 1 | Cycle, chain, common cause, control | Initial test between hypotheses; blank properties; controlled for feature order, cycle order, temporal order | Figure 2 |
| 2 | Cycles, cause–effect | Controlled for feature content, feature order, relation content; direct comparison of structures; different dependent measure | Table 2 |
| 3 | Same | Unambiguous dependent measure | Table 2 |
| 4 | Same | Increased cycle salience | Table 2 |
| 5 | Same | Natural and artifactual kinds | Table 2 |
| 6 | Same | No diagrams, blank properties | Table 2 |

dependent measures. In Experiments 4, 5, and 6, we posed stronger tests by examining conceptual centrality judgments for cycles when the continuous, potentially infinite nature of cycles was made especially salient. In addition, in Experiment 5, we examined whether the conceptual centralities found for cycles go beyond the disorder domain, and, in Experiment 6, we ruled out any undue influence of visual cues in the causal diagrams that we used to depict cycles in Experiments 1–5. In all of these experiments, we compared the conceptual centralities of features involved in cyclic versus acyclic causal structures. As we have discussed, prestige centrality measures make very different predictions for the two possible representations of cycles—true cycles and simplified cycles—relative to acyclic structures.

EXPERIMENT 1 Initial Test Between the Alternative Cycle Representations

In keeping with previous work on the centrality model, we operationalized the conceptual centrality of a feature as that feature's *mutability* (Sloman et al., 1998), the degree to which missing the feature reduces the category membership likelihood of an exemplar. In Experiment 1, we compared the conceptual centralities of features involved in causal cycles with features involved in acyclic common-cause and causal chain structures (see Figures 1A and 1B). The comparisons of primary interest were between the cycle features and the cause features in each acyclic structure. As can be seen in the prestige centrality rankings depicted in Figure 1A, if cycles are weighted as is suggested by prestige centrality and the assumption of a true cycle representation, Features A and B at the current time step should be more conceptually central than Features D, G, and E. Alternatively, if cycles are instead represented as simplified cycles, as is depicted in Figure 1B, Features A and B should be less conceptually central than Features D and G and equivalently conceptually central to Feature E. In addition, we compared ratings for the features involved in acyclic structures to one another to ensure that past results were replicated (e.g., whether Feature G is more central than Features H and I in the common-cause structure). Note that different letters are used here for ease of illustration only; in Experiment 1, each set of blank-property features took a turn playing each type of causal structure between subjects.

We also examined whether the temporal order of features (e.g., which features were read first from left to right) influences judgments of conceptual centrality independent of the influence of causal information. Such an influence of temporal order is relatively unlikely, given that Waldmann and Holyoak (1992) analogously found that causal structure reliably and robustly guides causal learning and that temporal order unequivocally does not. Nonetheless, including a manipulation of temporal order allowed us to clearly test whether causal structure *per se*, rather than temporal order, influences the perceived conceptual centralities of features.

Method

Participants. Forty-eight Northeastern University undergraduates participated in this experiment in exchange for either partial credit in an introductory psychology course or payment at the rate of \$10/h. Compensation was the same in all six experiments.

Materials and Procedure. Four artificial concepts were used to minimize the influence of the participants' prior background knowledge. Specifically, artificial disorder concepts were used because disorders are particularly suited for creating plausible reversible causal relationships that can be readily and intuitively understood, even for blank properties. Such reversible relationships were critical in the present experiment for forming cyclical causal scenarios. Each disorder concept consisted of three features, and these features were blank properties so that we could minimize the influence of the participants' prior background knowledge (e.g., Symptom P, Symptom F, and Symptom G). The features for all of the concepts are listed in Appendix B.

Four background knowledge structures (causal cycle, causal chain, common cause, and control) for each of the four concepts were assembled. The overall number of causal relations in the first three structures was equated (two in each; see Figures 1A and 1B). All scenarios were approximately equated for length.

The participants were given a paper packet containing an instruction sheet and four of the artificial concepts, each on a separate page. Each disorder concept was accompanied by one of the four background knowledge structures, such that each participant viewed each of the four structures once across the four different concepts.⁴ The four structures were presented within subjects, because the comparative responses across structures, rather than absolute responses, were meaningful.

The order in which the concepts were presented was randomized for each participant. Which concept was embedded in which causal structure was rotated between subjects, such that each causal structure took a turn being presented with each concept across participants.

At the top of each page, the participants read the name of the disorder concept and learned that three features were characteristic of that concept. In the causal conditions, they then saw causal information about the relations between features. First, in keeping with

previous studies (Ahn et al., 2000; Rehder & Hastie, 2001; Sloman et al., 1998), the participants were shown a diagram in which the causal relationships between features were depicted with arrows. They were told that if there was no arrow between any two features, there was no causal relation between those features. Below the diagram, the causal structure was explained in a short paragraph. Any features not participating in causal relations were also explicitly described as such in this paragraph.

The temporal order in which the information was presented was also manipulated between subjects. Materials in the *forward-order condition* always depicted and described causes to the left of their effects. Causal relationship descriptions in the forward-order condition took the form "Symptom P tends to cause Symptom F." In contrast, in the *reversed-order condition*, the position of the features in the diagrams from right to left was reversed, and the causal arrows were also reversed, such that the arrows pointed left toward their effects. In this condition, the wording of the descriptions was also changed to "Symptom F tends to be caused by Symptom P," so that the abstract causal information was identical to that found in the forward-order condition, but so that the participants read about the effect before the cause. In the control scenarios, the order in which the features were listed in the diagram and description was reversed in the reversed-order condition.

After reading the provided information about a disorder concept, the participants then answered three category membership likelihood questions in randomized order, following the procedure of

Sloman et al. (1998). Specifically, they were asked, "if a patient has all the symptoms of [Disorder X] EXCEPT [Symptom Y], does the patient have [Disorder X]?" The participants responded on a scale of 0–100 (0, *definitely no*; 100, *definitely yes*).

Results and Discussion

Overview of analyses. In all six experiments, category membership likelihood ratings were subtracted from 100 to obtain the conceptual centralities of features. Higher numbers therefore correspond to greater conceptual centrality in all analyses (see Figure 2 and Table 2). All analyses were conducted at the $\alpha = .05$ level. All subgroups of planned comparisons were sets of Bonferroni-corrected, paired-samples *t* tests. In Experiments 2–6, the data were collapsed across the four between-subjects counterbalancing conditions and concept items for analysis. In all of the experiments except one, features of the same type did not reliably differ and were collapsed together for simplicity in the analyses (i.e., the two features of the cycle, and the two effects in the common-cause structure; all p s $\geq .16$, all η^2 s $< .09$).

A 2 (direction: forward, reversed) \times 12 (feature role) ANOVA showed that there was no effect or interaction

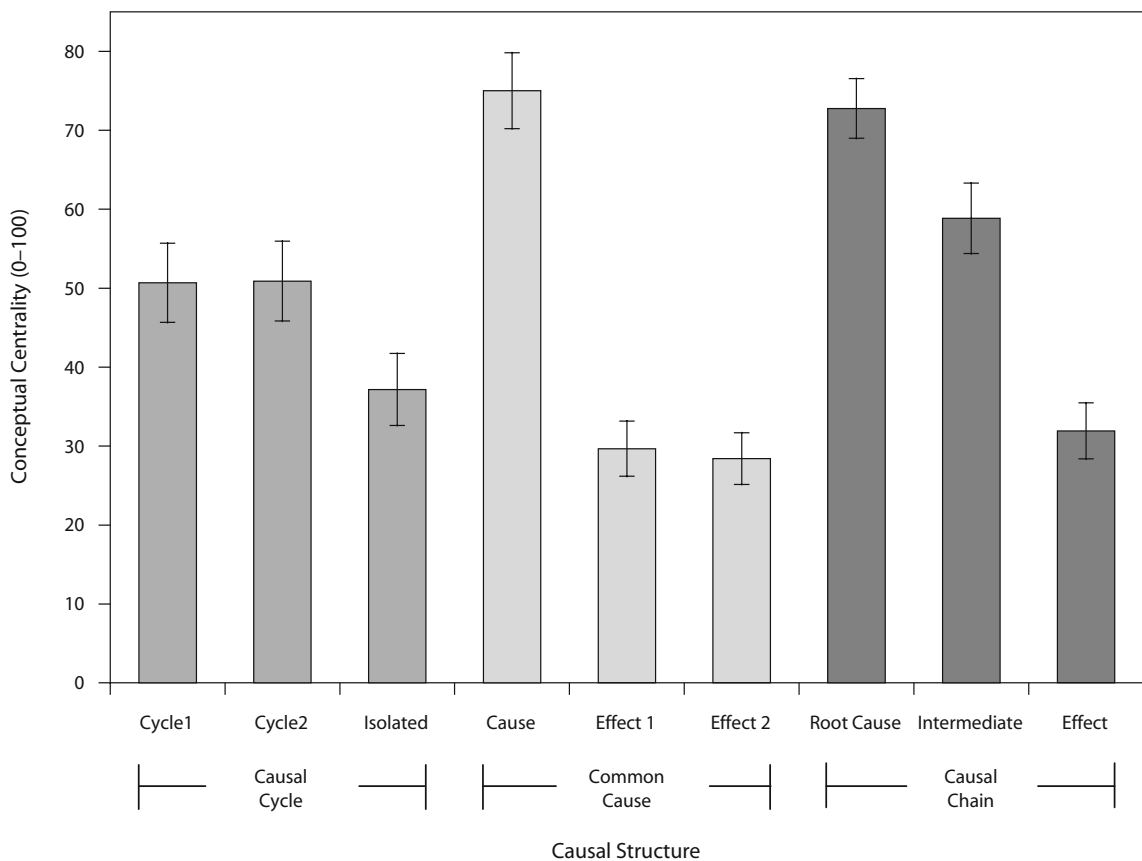


Figure 2. Results of Experiment 1. Means have been collapsed across counterbalanced items. Because no effects of temporal order were found, the means are also collapsed across the forward-order and reversed-order conditions for simplicity, and, therefore, the left-to-right feature positions depicted here are irrelevant.

Table 2
Results of Experiments 2–6

| Experiment | Cycle 1 | | Cycle 2 | | Cause | | Effect | |
|------------|----------|-----------|----------|-----------|----------|-----------|----------|-----------|
| | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> |
| 2 | 48.28 | 3.49 | 49.53 | 4.01 | 47.20 | 3.27 | 37.99 | 2.57 |
| 3 | 52.48 | 3.64 | 56.56 | 3.62 | 55.94 | 3.71 | 44.51 | 4.00 |
| 4 | 58.21 | 3.06 | 56.63 | 3.20 | 54.62 | 3.28 | 44.62 | 3.03 |
| 5 | 70.99 | 4.29 | 74.58 | 3.85 | 73.02 | 2.90 | 61.72 | 3.89 |
| 6 | 53.43 | 5.46 | 50.03 | 5.06 | 49.47 | 5.55 | 33.28 | 3.94 |

Note—Comparisons of ratings are only meaningful within studies (where ratings were made within subjects) and not between studies (where ratings were made by different participants at different points in time). Higher numbers correspond to greater conceptual centrality.

involving direction (all p s > .406, all η^2 s < .02). Thus, feature centralities were not affected by temporal order, analogous to previous findings by Waldmann and Holyoak (1992) for causal learning, and the small effect sizes suggest that increasing the sample size will not change this result. In the remainder of this article, therefore, we will not be concerned with temporal order as a potential alternate explanation of the results. (See Figure 2.)

There was, however, a main effect of feature role [$F(11,506) = 18.58$, $MS_e = 639.81$, $p < .001$, $\eta^2 = .29$]. Planned comparisons revealed that the common-cause feature ($M = 75.02$, $SE = 4.80$) was more central than the cycle features ($M = 50.79$, $SE = 5.04$) [$t(47) = 3.50$, $p = .001$, $\eta^2 = .21$], and the root cause in the causal chain ($M = 72.77$, $SE = 3.79$) was also more central than the cycle features [$t(47) = 3.26$, $p = .002$, $\eta^2 = .19$]. Furthermore, the cycle features' mean centrality did not differ from that of the intermediate-cause feature in the causal chain ($M = 58.85$, $SE = 4.47$) [$t(47) = 1.55$, $p = .127$, $\eta^2 < .05$]. These findings are uniquely consistent with a simplified-cycle representation (Figure 1B).

Ostensibly because of the blank nature of the features used, the ratings given to the three features in the control condition, collapsed across concepts, were exactly identical (all M s = 40.13, all SE s = 3.65), strongly suggesting that the findings above were driven by the use of the given causal information.

Secondary sets of planned comparisons were also conducted on the acyclic structures to confirm that our method reliably replicated that of Sloman et al. (1998) and adhered to prestige centrality predictions for acyclic structures. In the common-cause structure, the cause ($M = 75.02$, $SE = 4.80$) was more central than its dependents (collapsed together, $M = 29.04$, $SE = 3.27$) [$t(47) = 7.07$, $p < .001$, $\eta^2 = .52$]. In the causal chain, the root cause ($M = 72.77$, $SE = 3.79$) was more central than the intermediate cause ($M = 58.85$, $SE = 4.47$) [$t(47) = 2.46$, $p = .018$, $\eta^2 = .11$]. The intermediate cause, in turn, was more central than the terminal effect ($M = 31.94$, $SE = 3.55$) [$t(47) = 6.72$, $p < .001$, $\eta^2 = .49$].

To summarize, although the results were generally consistent with the predictions of the prestige centrality account (e.g., Sloman et al., 1998), the judgments of the cycle features were not. The participants deemed the cycle features less conceptually central than the root cause in both the

chain and common cause structures. In addition, the cycle features and the intermediate features in the chain structure were equally conceptually central. Changing the true cycle representation assumption to the simplified cycle representation assumption that we proposed, in contrast, does elicit predictions for cycles that match the present findings. However, some limitations of this study are that the participants made comparisons across concepts and that the blank-property stimuli were very artificial. We addressed these and other issues in Experiment 2.

EXPERIMENT 2 Cycles Versus Simple Causes

In Experiment 2, we first wanted to ensure that the participants were directly comparing causal cycles with acyclic structures. In Experiment 1, each concept contained either a cycle or an acyclic structure, and comparisons were made within subjects but across different concepts. It is not yet entirely clear that the simplified cycle hypothesis would be supported even if both types of structure were included in the same concept, allowing a direct comparison. To set this situation up with the simplest structures possible, in Experiment 2, we compared conceptual centrality judgments for cycle features with those for simple cause–effect relations.

Again, the question was how people deal with causal cycles. For example, see Figures 3A–3C, in which F1–F4 correspond to the four features of a single concept. The comparison of primary interest was between the cycle features (e.g., F1 and F2 in Figures 3B and 3C) and the single-cause feature (F3). Prestige centrality, when assuming the true cycle representation, predicts that the cycle features should be significantly more conceptually central than the cause feature in a simple cause–effect relationship (see Figure 3B). In contrast, if causal cycles are represented as simplified cycles extending just one time step into the future, as was suggested by the results of Experiment 1, the conceptual centralities of cycle features and the cause feature in the simple cause–effect relation should not differ (see Figure 3C).

In addition, we used more realistic stimuli in Experiment 2 and fully counterbalanced for content effects, for the order in which the type of relation appeared (cycle or simple cause–effect relation on the left versus right),

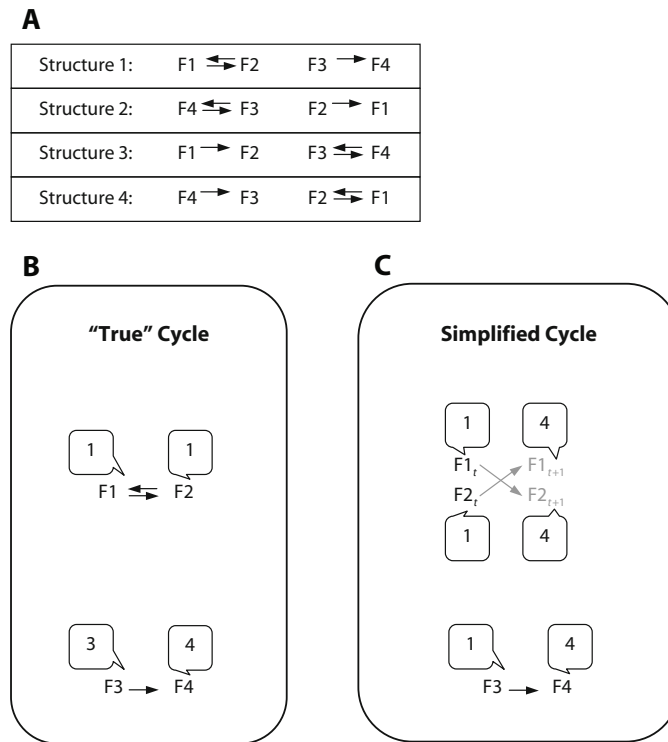


Figure 3. Schematics of the four within-subjects conditions used in Experiments 2–6, where F1–F4 refer to the features (A). Balloons contain prestige centrality rankings (alpha centrality) for “true” cycle (B) versus simplified cycle (C) representations of the cyclic causal structure.

and for overall left-to-right and right-to-left feature positioning.

Method

Participants. Another group of 24 Northeastern University undergraduates participated.

Materials and Procedure. Four artificial disorder concepts previously developed by Kim and Ahn (2002b) were modified for use in the present experiment. Each concept comprised four features from actual Diagnostic and Statistical Manual of Mental Disorders (4th ed.; American Psychiatric Association, 2000) disorders. The features for each concept were taken from four different classes of disorders to minimize the participants’ prior knowledge about any relationships between the features within any one artificial concept (see Appendix B). Causal mechanism information for each relationship, as in Kim and Ahn (2002b), was also provided (see Appendix C for an example). For each concept, four between-subjects causal scenarios were created, such that presentation order and the use of causal relationships could be completely counterbalanced. Specifically, in two of these scenarios, the features were listed from left to right in the order F1, F2, F3, F4; in the other two, the features were listed in the reverse order from left to right (F4, F3, F2, F1). In one of each of these presentation orders, the two leftmost features were involved in a causal cycle and the two rightmost features were involved in a simple cause–effect relation. In the remaining two conditions, the two rightmost features were involved in a causal cycle and the two leftmost features were involved in the simple cause–effect relation. (See Figure 3A for a schematic illustrating the different conditions.) Thus, each feature–feature causal relationship was used both in cycles and in simple cause–effect relationships across the four between-subjects scenarios.

Furthermore, this manipulation ensured that every feature took a turn acting as the cause, effect, left-hand cycle feature, and right-hand cycle feature between subjects. This eliminated any potential extraneous effects of feature content. The participants viewed each scenario in a different concept, for a total of four concepts. In other words, each of the 24 participants made a total of 16 category membership likelihood ratings, rating each of the four different types of features (cause, effect, left-hand cycle, right-hand cycle) four separate times across the four different artificial categories.

The dependent measure question that the participants answered for each feature was the following: “Imagine to yourself a patient who has all the symptoms of [Disorder Y]. If a patient is in all ways like a typical person with [Disorder Y] EXCEPT that he or she does NOT have the symptom of [Symptom Z], does the patient have [Disorder Y]?” Ratings were again made on a scale of 0–100 (0, *definitely no*; 100, *definitely yes*). This question, adapted from Sloman et al. (1998), was used to expand the generality of the findings to a different dependent measure. In all other respects, the procedure was the same as that in Experiment 1.

Results and Discussion

A repeated measures ANOVA (feature type: cycle, cause, effect) revealed a main effect of feature type [$F(2,46) = 4.93, MS_e = 167.75, p = .010, \eta^2 = .18$]. As is uniquely consistent with a simplified cycle representation, cycle features ($M = 48.91, SE = 3.58$) were not considered more conceptually central than the single-cause feature ($M = 47.20, SE = 3.27$) [$t(23) = 0.40, p = .692, \eta^2 < .01$]. The effect size of $\eta^2 < .01$ strongly suggests

that the absence of a statistically significant effect cannot be due to a lack of power. Cycle features were also deemed more conceptually central than the single-effect feature ($M = 37.99$, $SE = 2.57$) [$t(23) = 3.37$, $p = .003$, $\eta^2 = .33$]. Finally, the single-cause feature was more conceptually central than its effect [$t(23) = 2.53$, $p = .019$, $\eta^2 = .22$]. These results once again exclusively support the simplified cycle hypothesis, this time when contrasting features were involved in cycles versus simple acyclic cause-effect relations. We obtained these results using realistic stimuli that were fully controlled for content, both with respect to individual features and to the relations between the features, and that were fully counterbalanced for presentation order. Moreover, these results were obtained for concepts in which both cycles and acyclic structures were included for direct comparison. (See Table 2.)

EXPERIMENT 3

Ruling Out Misinterpretation of the Dependent Measure for Realistic Features

In Experiment 1, the dependent measure question clearly stated that all of the features were present in each transfer item except for one. However, the disorder concepts with realistic features used in Experiment 2 were presented with the potentially ambiguous dependent measure question, "Imagine to yourself a patient who has all the symptoms of [Disorder Y]. If a patient is in all ways like a typical person with [Disorder Y] EXCEPT that he or she does NOT have the symptom of [Symptom Z], does the patient have [Disorder Y]?" It seems unlikely that people interpreted the phrase "in all ways like a typical person with [Disorder Y]" as indicating that symptoms other than Z could conceivably be missing in this patient, because again, the question in Experiment 1 was clearly unambiguous on this point, and the results of both experiments supported the same hypothesis. However, it might be more likely that the scenario could be interpreted incorrectly for disorders with realistic features, as in Experiment 2, than for disorders with blank properties, as in Experiment 1. That is, inducing the presence of other features given the presence of a realistic feature might be more likely than inducing the presence of other features given the presence of a blank feature. Therefore, in the present experiment, we examined whether the results of Experiment 2 would be replicated when using disorders with realistic features and the clearly unambiguous dependent measure question from Experiment 1.

Method

Participants. The participants were another group of 40 Northeastern University undergraduate students.

Materials and Procedure. The materials and procedure of Experiment 2 were used in the present experiment, except that the participants were asked, "if a patient has all the symptoms of [Disorder X] EXCEPT [Symptom Y], does the patient have [Disorder X]?" The participants responded on a scale of 0–100 (0, *definitely no*; 100, *definitely yes*).

Results and Discussion

A repeated measures ANOVA (feature type: Cycle 1, Cycle 2, cause, effect) revealed a main effect of feature

type [$F(3,117) = 3.70$, $MS_e = 331.69$, $p = .014$, $\eta^2 = .09$].⁵ Cycle features (Cycle 1, $M = 56.56$, $SE = 3.64$; Cycle 2, $M = 52.48$, $SE = 3.62$) were not rated as more conceptually central than the single-cause ($M = 55.94$, $SE = 3.71$) feature [Cycle 1 to cause, $t(39) = 0.14$, $p = .886$, $\eta^2 < .01$; Cycle 2 to cause, $t(39) = -0.81$, $p = .422$, $\eta^2 < .02$]. Overall, cycle features were deemed more conceptually central than the single effect [Cycle 1 to effect, $M = 44.51$, $SE = 4.00$, $t(39) = 2.76$, $p = .009$, $\eta^2 = .16$; Cycle 2 to effect, $t(39) = 1.96$, $p = .057$, $\eta^2 = .09$], and the single-cause feature was likewise judged more conceptually central than its effect [$t(39) = 2.34$, $p = .025$, $\eta^2 = .12$]. The results thereby once again exclusively support the simplified cycle hypothesis. In this experiment, the controlled, counterbalanced design of Experiment 2 was paired with an unambiguous dependent measure and yielded the same results. (See Table 2.)

EXPERIMENT 4

Potentially Infinite Cycles Versus Simple Causes

Experiments 1, 2, and 3 demonstrate that reasoners treat cycle features as less central than prestige centrality measures predict that they should be for true-cycle representations. However, it is not entirely clear whether the participants in these experiments really understood that causal influence in a cycle can potentially go on infinitely—that is, not just from A to B and back to A, but also back to B again, and so on. Thus, we replicated Experiment 2 with a critical change in the materials that was designed to increase the salience of the potentially infinite nature of causal cycles. Whenever a causal cycle was explained, we made the continuously cycling nature of the cycle explicit and clear. This change allowed a more stringent test of the simplified cycle hypothesis.

Method

Participants. Another group of 32 undergraduate students at Northeastern University participated.

Materials and Procedure. The materials and procedure of Experiment 2 were used in the present experiment, with one modification. In each causal cycle scenario (e.g., of the abstract form "A causes B, and B causes A"), a sentence was added to the end of the explanation explicitly clarifying that "then A, in turn, causes B again, and so on." This made clear the continuously cycling nature of the causal cycle. This is also a stronger test, because the cycle features were mentioned more frequently and discussed at more length than the simple cause-effect features.

Results and Discussion

A repeated measures ANOVA (feature type: cycle, cause, effect) revealed a main effect of feature type [$F(2,62) = 7.21$, $MS_e = 201.04$, $p = .002$, $\eta^2 = .19$]. Cycle features ($M = 57.42$, $SE = 2.82$) were not deemed more conceptually central than the single-cause feature ($M = 54.62$, $SE = 3.28$) [$t(31) = 0.67$, $p = .508$, $\eta^2 = .01$]. A power analysis indicated that doubling the number of participants would not have changed this result ($p = .34$). Cycle features were rated as more conceptually central than the single-effect feature ($M = 44.62$, $SE = 3.03$) [$t(31) = 3.38$, $p = .002$, $\eta^2 = .27$]. Finally, the single-

cause feature was judged to be more conceptually central than its effect [$t(31) = 4.15, p < .001, \eta^2 = .36$]. The results again support the simplified cycle hypothesis, even when the continuously cycling nature of cyclical relations is made explicit. (See Table 2.)

EXPERIMENT 5 Generalizing Findings to Natural and Artifactual Kinds

The experiments reported here so far have focused on disorder concepts, but are the present findings limited to the disorder domain? It is possible that disorder concepts differ from natural and artifactual kinds, in that when reasoning about disorder concepts, people might be more likely to infer the presence of additional unspecified causes of the symptoms. For example, when learning that inflated self-esteem is a symptom of a particular disorder, people might not only learn the causal relations being taught, but might also assume that some events in this person's past might have also contributed to causing this symptom. Such assumptions might generally be made in the domain of disorders/diseases, such that, even in the case of disorders composed of blank properties in Experiment 1, people could potentially have exhibited this tendency. To test systematically, and under fully controlled conditions, whether or not prestige centrality would accurately describe reasoners' treatment of causal cycles outside of the disorder domain, we replicated Experiment 4 using two previously created living natural kinds, one nonliving natural kind, and one artifactual kind. It may also be particularly important to include this experiment because the Sloman et al. (1998) prestige centrality model was originally developed using natural kinds and artifacts.

Method

Participants. The participants were another group of 24 Northeastern University undergraduates.

Materials and Procedure. The procedure of Experiment 4 and the unambiguous dependent measure question from Experiments 1 and 3 were used. We used artificial concepts previously developed by Rehder (2003a), because he had already provided plausible causal relations between the features of these concepts; they were therefore readily adapted to create cycles in the present experiment. The four natural and artifactual kind categories that we used had four features each (see Appendix B).

Results and Discussion

A repeated measures ANOVA (feature type: cycle, cause, effect) revealed a main effect of feature type [$F(2,46) = 7.25, MS_e = 138.06, p = .002, \eta^2 = .24$]. Cycle features ($M = 72.79, SE = 3.89$) were not judged to be more conceptually central than the single-cause feature ($M = 73.02, SE = 2.90$) [$t(23) = 0.07, p = .941, \eta^2 < .01$], and cycle features were judged more conceptually central than the single-effect feature ($M = 61.72, SE = 3.89$) [$t(23) = 3.18, p = .004, \eta^2 = .31$]. The single-cause feature was deemed more conceptually central than its effect [$t(23) = 3.20, p = .004, \eta^2 = .31$]. Overall, we found support for the simplified cycle hypothesis using manufactured natu-

ral and artifactual kinds, suggesting that this result is not specific to the domain of disorders. (See Table 2.)

EXPERIMENT 6 Ruling Out Effects of Cues From the Visual Diagram

The results of Experiments 4 and 5 supported the simplified cycle hypothesis, even when it was made explicit and clear to the participants that the cycle in each case was meant to go on potentially infinitely—not just from A to B and back to A, but also back to B again, and so on. However, perhaps the participants were instead taking cues from the visual diagram itself. It is possible that because the cycle in the diagram depicted only two arrows, the participants ignored the information about continuous cycling given to them in the paragraph. Thus, in Experiment 6, we tested whether people were unduly influenced by the visual diagram in perceiving cycle features to be no more conceptually central than simple cause features.

Method

Participants. The participants were another group of 24 Northeastern University undergraduates.

Materials and Procedure. As in Experiment 4, the continuously cycling nature of the causal cycle was made clear. Importantly, however, the visual diagrams used in Experiment 4 were omitted. We also used the blank property concepts from Experiment 1, as well as the unambiguous dependent measure question from Experiments 1, 3, and 5: "If a patient has all the symptoms of [Disorder X] EXCEPT [Symptom Y], does the patient have [Disorder X]?" Again, ratings were made on a 0–100 scale (0, *definitely no*; 100, *definitely yes*). Finally, as in Experiments 2–5, we used the counterbalancing conditions depicted in Figure 3A.

Results and Discussion

A repeated measures ANOVA (feature type: cycle, cause, effect) revealed a main effect of feature type [$F(2,46) = 4.80, MS_e = 506.02, p = .013, \eta^2 = .17$]. Cycle features ($M = 51.73, SE = 5.05$) were not thought to be more conceptually central than the single-cause feature ($M = 49.47, SE = 5.55$) [$t(23) = 0.26, p = .798, \eta^2 < .01$], and cycle features were deemed more conceptually central than the single-effect feature ($M = 33.28, SE = 3.94$) [$t(23) = 3.59, p = .002, \eta^2 = .36$]. The single-cause feature was judged to be more conceptually central than its effect [$t(23) = 3.30, p = .003, \eta^2 = .32$]. In summary, the results remain unchanged even when there are no visual cues present to suggest that the cycle should be abbreviated and the potentially continuously cycling nature of cyclical relations is made explicit. (See Table 2.)

GENERAL DISCUSSION

In the present experiments, we sought to investigate how category judgments are made when the features of a category are connected in a causal cycle. Despite theoretical and empirical work about how causal structure influences categorization, causal cycles remain outside the scope of the majority of current theories. An exception is the prestige centrality model put forth by Sloman et al.

(1998). Prestige centrality measures easily handle causal cycles and thus provide a testable set of predictions for the present experiments. The prestige centrality account, together with the assumption about the true or veridical representation of cycles, predicts that participants would rank the features involved in our causal cycles as more central to the concept than features in the acyclic structures.

In contrast, across six experiments, the participants' judgments indicated that they did not consider cycle features to be the most central to the underlying concept. Specifically, cycle features were judged to be less central than the root cause of a three-feature chain and the cause in a common-cause structure. Instead, cycle features were judged to be as central as a feature that caused just one other feature (e.g., the intermediate feature in a three-feature chain and the cause in the simple cause-effect structure). These findings held true across comparisons with different types of acyclic structures, comparisons within and between concepts, different dependent measures, blank-property concepts, and realistic artificial concepts, with and without instructions to consider the cycle's potentially infinite nature, with and without accompanying visual diagrams, and inside and outside the disorder domain. Our proposal that the assumption of veridical or "true" representations of cycles be replaced by simplified cycle representations when using prestige centrality measures can account for the full set of data that we obtained.

The idea of a simplified representation accounting for the apparent deviations of our results seems attractively parsimonious, but exact details of the simplified structure remain unknown. For example, the prestige centrality of a given feature is influenced by the number of features that depend on that feature but is not influenced by the number of features on which it depends. Thus, although our results seem to uniquely suggest a simplified representation that extends only a single step into the future, we are unable to determine how far into the past our reasoners' representations extend. Simplified cycle representations extending even two steps into the future would not yield centrality predictions that match the results of our studies (the cycles would be given more weight). Yet simplified cycle representations extending one, two, or three steps into the past would all generate equivalent centrality predictions and would thus be equally successful in accounting for the present results. Thus, our participants' simplified cycle representations could look exactly like that in Figure 1B, but they could just as well include an elaborate description of all of the past causes of A_t and B_t (e.g., A_{t-1} , B_{t-1} , A_{t-2} , B_{t-2} , etc.). Future work will be necessary to determine the full extent of reasoners' simplifications.

Implications for Models of Theory-Based Categorization

The present results are interesting in that our reasoners appeared to be representing the provided causal information in simplified form. It remains possible that prestige centrality models (e.g., Sloman et al., 1998) instead inaccurately describe the process by which reasoners compute conceptual centralities. However, the failure of prestige centrality with regard to causal cycles stands in clear con-

trast to its success in predicting the judgments of features involved in acyclic structures (e.g., Kim & Ahn, 2002a; Sloman et al., 1998). Thus, it is most parsimonious to suggest instead that "true cycle" representations do not accurately depict reasoners' representations of cycles. Once a simplified cycle representation is instead presumed, prestige centrality predicts the conceptual centralities of features involved in cycles and acyclic structures with equal ease and accuracy.

Furthermore, our results' support for the simplified cycle hypothesis may make it possible for models that do not explicitly deal with causal cycles to now make predictions about categories whose features participate in causal cycles. For example, although neither the causal status hypothesis (Ahn, 1998) nor causal-model theory (Rehder & Hastie, 2001) have hitherto been able to make predictions about how cycle features influence categorization, these models may be able to generate predictions for a simplified cycle representation. We therefore expect that future work on these models may be able to accommodate simplified cycles more easily than "true" causal cycles. Such work, in turn, may also allow future experiments to better inform the debate between the centrality model, causal-model theory, and the causal status hypothesis as competing process models of theory-based categorization.

As we did in the present study, many past investigators of causal theory-based categorization have relied on the use of artificial, simplified categories to systematically control for prior knowledge (e.g., Ahn, 1998; Ahn et al., 2000; Rehder, 2003a, 2003b). The possibility therefore remains that people reason differently using causal models of these artificial categories than when they use causal models of real-world categories (e.g., depression, influenza, cows, eggplants); for example, one might argue that the participants in our study did not have a deep understanding of the causal relations between symptoms. For at least the following five reasons, however, we believe that our results are nonetheless informative with respect to real-world categorization processes. First, realistic features were used in four of the six studies, and in these studies, we provided mechanism information to facilitate understanding of the relations (see Appendix C). Second, our participants were informed of the broader category to which each concept belonged. For example, in Experiments 1-4 and 6, they were aware that the categories were disorders; thus, the participants' existing framework theories of how people's circumstances and behaviors may generally influence one another were available to fill in any major gaps in knowledge. In Experiment 5, the participants were also told of the broader categories to which the item belonged (e.g., animals at the superordinate level, shrimp at the basic level); therefore, again, none of these categories were completely foreign to the participants.

Third, numerous studies have shown that people spontaneously look for causal mechanism information (Ahn, Kalish, Medin, & Gelman, 1995), even if they are not provided with the details and, furthermore, that they spontaneously create detailed causal mechanisms to explain why any two features appear together in a concept (Hastie,

Schroeder, & Weber, 1990; Kunda, Miller, & Clare, 1990). Fourth, note that the features in some of our studies had more content (e.g., a persistent lack of self-control), whereas others had less (e.g., Symptom X), but the results of all of the experiments were exactly the same.

Fifth, as was noted earlier, people's explanatory theories may be less complete and more fragmented than they think (diSessa, 1988, 1993; Mills & Keil, 2004; Rozenblit & Keil, 2002). Accordingly, it is likely that people do not necessarily hold highly detailed causal theories of concepts in real life either; they may often have only skeletal frameworks of explanations, and therefore our stimulus materials may not be quite as dramatically different from real-life concepts as one might suppose.

Even so, we acknowledge that future work will nonetheless be needed to examine the conceptual centralities of cycle features contained in people's spontaneously generated real-life theories of concepts (e.g., Sloman et al., 1998). In particular, one might further speculate that the tasks that we used may have involved more analytical processes, whereas it is possible that conceptual judgment tasks with greater ecological validity may elicit a more automatic mode of processing. Whether conceptual centrality judgment tasks for cycles change depending on whether the process is analytical or automatic in nature remains an important question for future work.

Broader Implications and Further Questions Regarding Cycles

The present experiments also raise the broader question of whether and how we simplify our representations of causal background knowledge in general. An intriguing question concerns the threshold of complexity; that is, how much complexity in a causal structure will we process as being relevant to judgments such as conceptual centrality, property induction, and the like? Among the basic types of acyclic structures, this issue might be most relevant to long causal chains. In such structures, people might, for example, simplify the chain by representing more global differences between deep causes, intermediate causes, and terminal effects rather than representing the various incremental differences between the numerous features in the middle of the chain. This possibility remains to be tested systematically.

Furthermore, it is likely that causal cycles differ conceptually from acyclic causal structures in potentially important ways. For example, it can be useful to distinguish between cycles of a maintaining nature as opposed to an escalating nature. In maintaining cycles, each causal relation maintains the state or presence of the other feature rather than precipitously increasing or decreasing it. For instance, suppose that Jane is able to run laps because she is in decent cardiovascular health and also that she is able to maintain her decent cardiovascular health because she runs laps. In a maintaining version of this cycle, she might run the same number of laps each time she exercises and might maintain her level of cardiovascular health at about the same level. In contrast, some causal cycles in the world are of an escalating nature, as in $A_1 \rightarrow B_1 \rightarrow A_2 \rightarrow B_2$, and

so on, wherein A_2 and B_2 are more extreme versions of A_1 and B_1 . In an escalating version of the cycle between cardiovascular health and running laps, Jane might try to run just one more lap each time she exercises, such that, eventually, her decent cardiovascular health becomes excellent cardiovascular health. In such cases, will people represent cycles farther than a single time step into the future? Future work will be needed to examine judgments of conceptual centrality while manipulating whether the cycle is explicitly described as a maintaining cycle, as opposed to an escalating cycle.

Another way in which cycles may differ from acyclic structures involves the frequency with which their features are assumed to reoccur over time. That is, in the world, features involved in a cycle may typically be assumed to reoccur more frequently over a fixed period of time than features involved in acyclic structures. In a related study, we are currently examining the possibility that rapid cycling may increase the importance of cycle features relative to slow cycling. For example, if the second feature in a rapidly cycling cycle occurs at time $t+1$, at that same time, the second feature in a slowly cycling cycle has not yet occurred. Thus, in accord with the notion that people only track cycles forward for a single time step, we should expect to see conceptual centrality differences between cycle features in rapid versus slow cycles.

In the present article, we focused on causal cycles in their simplest form: two features that cause each other in a back-and-forth loop. The question of how people might judge the conceptual centralities of features involved in larger cycles, even of the only slightly enlarged form $A \rightarrow B \rightarrow C \rightarrow A$, raises additional questions. In the context of making such a judgment about Feature A, will A be credited for explaining both B and C, or for only B? Larger cycles have been examined in real-world measurements of causal theories (e.g., Kim & Ahn, 2002b; Sloman et al., 1998) but remain to be tested by controlled manipulation of theories. This case becomes even more interesting when considering conceptual centrality judgments for longer causal chains that eventually wrap around to become a cycle; for example, $A \rightarrow B \rightarrow C \rightarrow D \rightarrow E \rightarrow A$. For such cycles, the immediate context in which the judgment takes place is probably of increased importance. That is, one may typically think only about a certain segment of the cycle at a time, thinking of the cycle in its entirety when primed to do so by context.

Furthermore, knowing which cause in a larger (e.g., $A \rightarrow B \rightarrow C \rightarrow A$) cycle was thought to occur first may factor into conceptual centrality judgments. Building again from a previous example, suppose that we think a child's poor performance in school and chronic insomnia cause each other in a causal cycle, but, at the very beginning of this problem, it was in fact the chronic insomnia that started it all. In our experiments, we found that in short ($A \leftrightarrow B$) cycles, people did not differentiate between the cycle feature listed first or second. However, it is also possible that the larger the cycle, the more people may differentiate the features; the consequences of being the initiating cause become more significant as more features are

affected by that initial cause. Similarly, cycle features may not be judged as equally conceptually central to each other in the case of asymmetric cycles (e.g., cycles in which the strength of the causal relation $A \rightarrow B$ is greater than the strength of $B \rightarrow A$).

Implications of the Simplified Cycle Hypothesis Beyond Conceptual Centrality

An additional important consequence of causal knowledge is that it affects the inductive inferences that are made about features of a concept (e.g., Hadjichristidis et al., 2004; Kim & Keil, 2003; Kim, Yopchick, & de Kwaadsteniet, 2008; Proctor & Ahn, 2007; Rehder & Burnett, 2005). If our representational proposal for cycles holds true across reasoning tasks, one might predict that cycles should be treated like acyclic structures in inductive inference.

It also remains to be seen which kinds of concepts might be more or less likely to foster beliefs about cyclical relationships. As was suggested in the introduction, there is clearly a diverse array of domains in which it is possible to imagine cyclical relationships, and again, previous work has shown that cycles in real-life theories are reported with equal frequency for disorders (Kim & Ahn, 2002b) and for natural/artifactual kinds (Kim, 2005). We therefore speculate that beliefs about cycles may appear across many different concepts, from mind-body feedback loops to symbiotic relations in ecology. It is also an open question as to what types of features might be more likely to participate in cyclical relationships. One might begin by supposing that deep, essence-like features, such as genes, will be unlikely to participate in cycles, but when examining broad framework theories, it seems that even they may take part in cyclical relationships. For example, in the case of diseases and disorders, genetic predispositions may be perceived to interact so closely with environmental factors that our perception of which factor constitutes the root cause may, again, depend almost entirely on context. Additional research will be needed to determine whether this is the case.

AUTHOR NOTE

We thank Woo-Kyoung Ahn, John Coley, Brad Love, Edward Park, Andrea Patalano, and Jennelle Yopchick for helpful discussions and suggestions. We also thank Daniel Paulus for proofreading the manuscript. Parts of Experiments 2 and 4 were presented at the 2006 Cognitive Science Society meeting, Vancouver, British Columbia. Correspondence concerning this article should be directed to N. S. Kim, Department of Psychology, 125 Nightingale Hall, Northeastern University, 360 Huntington Avenue, Boston, MA 02115-5000 (e-mail: n.kim@neu.edu).

REFERENCES

- AHN, W.-K. (1998). Why are different features central for natural kinds and artifacts? The role of causal status in determining feature centrality. *Cognition*, **69**, 135-178.
- AHN, W.-K., KALISH, C. W., MEDIN, D. L., & GELMAN, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition*, **54**, 299-352.
- AHN, W.-K., & KIM, N. S. (2001). The causal status effect in categorization: An overview. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 40, pp. 23-65). San Diego: Academic Press.
- AHN, W.-K., KIM, N. S., LASSALINE, M. E., & DENNIS, M. J. (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 rev.). Washington, DC: Author.
- ANDERSON, C. A., & LINDSAY, J. J. (1998). The development, perseverance, and change of naive theories. *Social Cognition*, **16**, 8-30.
- BONACICH, P. (1972). Factoring and weighing approaches to status scores and clique identification. *Journal of Mathematical Sociology*, **2**, 113-120.
- BONACICH, P., & LLOYD, P. (2001). Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, **23**, 191-201.
- BORGATTI, S. P. (1995). Centrality and AIDS. *Connections*, **18**, 112-115.
- DEAN, T., & KANAZAWA, K. (1989). A model for reasoning about persistence and causation. *Computational Intelligence Journal*, **5**, 142-150.
- DI SESSA, A. A. (1988). Knowledge in pieces. In G. Forman & P. B. Puffall (Eds.), *Constructivism in the computer age* (pp. 49-70). Hillsdale, NJ: Erlbaum.
- DI SESSA, A. A. (1993). Toward an epistemology of physics. *Cognition & Instruction*, **10**, 105-225.
- DWECK, C. S., CHIU, C., & HONG, Y. (1995). Implicit theories and their role in judgments and reactions: A world from two perspectives. *Psychological Inquiry*, **6**, 267-285.
- FRIEDMAN, N., MURPHY, K., & RUSSELL, S. (1998). Learning the structure of dynamic probabilistic networks. In *Proceedings of the Fourteenth International Conference on Uncertainty in Artificial Intelligence* (pp. 139-147). Madison, WI: Morgan Kaufmann.
- HADJICHRISTIDIS, C., SLOMAN, S., STEVENSON, R., & OVER, D. (2004). Feature centrality and property induction. *Cognitive Science*, **28**, 45-74.
- HAGMAYER, Y., & DE KWAADSTENIET, L. (2008). Creating causal models: How therapists analyze clients' problems. In *Proceedings and abstracts of the 79th Annual Meeting of the Eastern Psychological Association* (p. 25). Piscataway, NJ: EPA.
- HASTIE, R., SCHROEDER, C., & WEBER, R. (1990). Creating complex social conjunction categories from simple categories. *Bulletin of the Psychonomic Society*, **28**, 242-247.
- IPSEN, I. C. F., & WILLS, R. S. (2006). Mathematical properties and analysis of Google's PageRank. *Boletín de la Sociedad Española de Matemática Aplicada*, **34**, 191-196.
- KEIL, F. C. (2006). Explanation and understanding. *Annual Review of Psychology*, **57**, 227-254.
- KIM, N. S. (2005). Stability and instability over time in explanatory theories of concepts. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society* (p. 2500). Mahwah, NJ: Erlbaum.
- KIM, N. S., & AHN, W.-K. (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.-K. (2002b). The influence of naive causal theories on lay concepts of mental illness. *American Journal of Psychology*, **115**, 33-65.
- KIM, N. S., & KEIL, F. C. (2003). From symptoms to causes: Diversity effects in diagnostic reasoning. *Memory & Cognition*, **31**, 155-165.
- KIM, N. S., & PARK, E. Y. (2009). ConceptBuilder: An open-source software tool for measuring, depicting and quantifying causal models. *Behavior Research Methods*, **41**, 128-136.
- KIM, N. S., YOPCHICK, J. E., & DE KWAADSTENIET, L. (2008). Causal diversity effects in information seeking. *Psychonomic Bulletin & Review*, **15**, 81-88.
- KUNDA, Z., MILLER, D. T., & CLARE, T. (1990). Combining social concepts: The role of causal reasoning. *Cognitive Science*, **14**, 551-578.
- LANGVILLE, A. N., & MEYER, C. D. (2003). Deeper inside PageRank. *Internet Mathematics*, **1**, 335-380.
- MEDIN, D. L., COLEY, J. D., STORMS, G., & HAYES, B. K. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review*, **10**, 517-532.
- MEDIN, D. L., LYNCH, E. B., COLEY, J. D., & ATRAN, S. (1997). Categorization and reasoning among tree experts: Do all roads lead to Rome? *Cognitive Psychology*, **32**, 49-96.
- MILLS, C. M., & KEIL, F. C. (2004). Knowing the limits of one's understanding: The development of an awareness of an illusion of explanatory depth. *Journal of Experimental Child Psychology*, **87**, 1-32.

MURPHY, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

NODELMAN, U., SHELTON, C. R., & KOLLER, D. (2002). Continuous time Bayesian networks. In *Proceedings of the Eighteenth International Conference on Uncertainty in Artificial Intelligence* (pp. 378-387). San Francisco: Morgan Kaufmann.

PAGE, L., BRIN, S., MOTWANI, R., & WINOGRAD, T. (1999). *The PageRank citation ranking: Bringing order to the Web* (Tech. Rep.). Palo Alto, CA: Stanford University, Computer Science Department.

PENNINGTON, N., & HASTIE, R. (1988). Explanation-based decision making: The effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 521-533.

PENNINGTON, N., & HASTIE, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality & Social Psychology*, **62**, 189-206.

PROCTOR, C., & AHN, W.-K. (2007). The effect of causal knowledge on judgments of the likelihood of unknown features. *Psychonomic Bulletin & Review*, **14**, 635-639.

REHDER, B. (2003a). Categorization as causal reasoning. *Cognitive Science*, **27**, 709-748.

REHDER, B. (2003b). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 1141-1159.

REHDER, B., & BURNETT, R. C. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology*, **50**, 264-314.

REHDER, B., & HASTIE, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General*, **130**, 323-360.

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.

REIN, J. R., LOVE, B. C., & MARKMAN, A. B. (2007). Feature relations and feature salience in natural categories. In *Proceedings of the Twenty-Ninth Annual Conference of the Cognitive Science Society* (pp. 593-598). Mahwah, NJ: Erlbaum.

ROZENBLIT, L. R., & KEL, F. C. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, **26**, 521-562.

SLOMAN, S. A., & AHN, W.-K. (1999). Feature centrality: Naming versus imagining. *Memory & Cognition*, **27**, 526-537.

SLOMAN, S. A., LOVE, B. C., & AHN, W.-K. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, **22**, 189-228.

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.

WALDMANN, M. R., HOLYOAK, K. J., & FRATIANNI, A. (1995). Causal models and the acquisition of category structure. *Journal of Experimental Psychology: General*, **124**, 181-206.

WASSERMAN, S., & FAUST, K. (1994). *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.

WAXMAN, S., MEDIN, D. [L.], & ROSS, N. (2007). Folkbiological reasoning from a cross-cultural developmental perspective: Early essentialist notions are shaped by cultural beliefs. *Developmental Psychology*, **43**, 294-308.

WILLS, R. S. (2006). Google's PageRank: The math behind the search engine. *Mathematical Intelligencer*, **28**, 6-11.

NOTES

1. Other accounts of the influence of causal knowledge in categorization (e.g., Ahn, 1998; Ahn & Kim, 2001; Ahn et al., 2000; Rehder, 2003a, 2003b; Rehder & Hastie, 2001; Rehder & Kim, 2006) have not been designed to deal with causal cycles. In the General Discussion section, we consider possible implications of our data for these accounts.
2. The Sloman et al. (1998) measure itself cannot be meaningfully applied to the structures used in the present study for reasons that have been outlined elsewhere (Bonacich & Lloyd, 2001). Solutions to these problems have led to the development of alternative prestige centrality measures (Bonacich & Lloyd, 2001; Ipsen & Wills, 2006; Page, Brin, Motwani, & Winograd, 1999; Wills, 2006), which we used to generate predictions instead. See Appendix A for details.
3. There is ample precedent for implementing this approach computationally in the machine learning literature (Dean & Kanazawa, 1989; Friedman, Murphy, & Russell, 1998; Nodelman, Shelton, & Koller, 2002).
4. Comparing causal structures across concepts could be problematic in that the participants may not have directly compared them with one another; however, all of the participants gave ratings for all structure types. More important, in Experiments 2-6, we required the participants to compare causal structures within concepts, which yielded results supporting the same hypothesis.
5. In Experiment 3 alone, the cycle features differed [$t(39) = 2.22$, $p = .033$, $\eta^2 = .11$]; however, the assignment of the labels given to the cycles is essentially arbitrary (see Figure 3), and this is therefore not particularly problematic.

APPENDIX A

Sloman et al. (1998) argued that a feature's centrality should depend on the centralities of its dependent features (i.e., effects) and the strength of the dependency. To formalize this concept of centrality, they proposed the following formula.

$$c_{i,t+1} = \sum_j d_{ij} c_{j,t} \tag{1}$$

This equation computes the centrality of a feature i by multiplying $c_{j,t}$, which represents the conceptual centrality of feature j , and d_{ij} , which represents how strongly feature j depends on feature i (see the sample set of dependency matrices, d , in Figure A1). This quantity is then summed across all of the features in the category (Sloman et al., 1998). Because the centrality values are interdependent, the model iterates, computing new centrality values based on those computed in the previous step. The model iterates until the centralities settle into a stable pattern. This is more generally known as the power method of computing eigenvector centrality (Wills, 2006), a type of prestige centrality.

| | A | B | C | D | E | F | G | H | I | A _t | B _t | A _{t+1} | B _{t+1} | C |
|---|---|---|---|---|---|---|---|---|---|----------------|----------------|------------------|------------------|---|
| A | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 |
| B | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

True Cycle
Causal Chain
Common Cause
Simplified Cycle

Figure A1. Dependency matrices (d) representing the causal structures tested in Experiment 1.

APPENDIX A (Continued)

Unfortunately, eigenvector centrality cannot be meaningfully computed for the simple structures utilized in the present study (Bonacich & Lloyd, 2001). One reason for this is that features with no dependents (i.e., dangling nodes), such as the terminal effects of a chain, acquire no centrality and thus contribute no centrality to their parent features, ultimately endowing all features in the chain with a centrality of zero. This and other problems led to the development of more sophisticated prestige centrality measures.

Below, we apply two such measures to the causal structures used in the present study. We note that these simulations assume that the strengths of the various dependencies are equivalent but that this is not an assumption of the computations themselves, which can handle various dependency strengths. We chose to hold strengths constant to maximize structural influence on the centrality predictions. This allows us to better discriminate between the different possible cycle representations.

We first computed a variant of eigenvector centrality referred to as *alpha centrality* (Bonacich & Lloyd, 2001). Alpha centrality proposes that centrality should be a factor of the number and centrality of a feature's dependents (just as in Sloman et al., 1998) plus an additional, fixed quantity. This latter quantity can be conceptualized as centrality derived from exogenous (not structural) sources and prevents the model from endowing nodes (i.e., features) with centralities of zero. The parameter α is allowed to vary and represents the degree to which overall centrality is determined by the structure of the graph. Computing alpha centrality for the structures illustrated in Figure 1A (represented by the dependency matrices in Figure A1) produces a solution that is consistent with the results of our study, but with one exception: It additionally suggests that the cycle features (A and B) will be more central than all other features for all values of α between 0 and 1. There is no value of α that will allow the cycle features' centrality to fall between that of the initial and terminal features in the causal chain structure (D and F, respectively), as was observed in Experiment 1. Computing alpha centrality is even more straightforward for the structures used in Experiments 2–6, since these yield solutions that do not depend on the value of α . For these structures (see Figures 3A–3C), the alpha centrality of the cycle features is 2.0, whereas the features involved in the chain acquire alpha centralities of 1.5 and 1.0 (for cause and effect, respectively). Here, again, alpha centrality is consistent in its prediction that the cycle features should be more central than features involved in acyclic causal structures, contrary to our findings.

We next applied the well-known PageRank algorithm (Ipsen & Wills, 2006; Page et al., 1999; Wills, 2006), used by Google to produce their centrality (or importance) rankings of individual Web pages. This algorithm includes adaptations intended to avoid conferring maximal centrality on nodes involved in small, isolated cycles (referred to as *rank sinks*), which tend to crop up on large-scale, real-world networks (Langville & Meyer, 2003). The PageRank algorithm uses a personalization vector that corresponds to the probability that an individual will eventually choose to visit any given Web page by means other than following the network of Web links (e.g., typing a Web address by hand). We used the uniform vector to avoid biasing the ranking toward any one feature. Additionally, the algorithm has a damping factor, which is a free parameter typically set to a value of .85. This damping factor acts much like α in the alpha centrality model: The higher the damping factor is, the more the resulting output will reflect the dependency structure.

Applying PageRank to the structures used in Experiment 1, we used a range of different values for this damping factor. The results of these simulations illustrate that the cycle features acquire greater centrality than any other feature for all damping factors greater than .5. For damping factors below this value, cycle features are only less central than the common cause feature; there are no values of the damping factor that allow the cycle features to be less central than the initial feature in the three-feature causal chain (Figure 1). For the structures used in Experiments 2–6, the algorithm always endows the cycle features with greater centrality than any other features; this is a pattern that does not depend on the value of the damping factor. Thus, PageRank mirrors the alpha centrality measures in conferring greater centrality on the cycle features than on the features involved in the acyclic structures.

Importantly, applying either alpha centrality or PageRank to the simplified cycle representation produces centrality values that match the results of the present study. The dependency matrix in Figure A1 illustrates how we represented cycle features as steps in a causal chain extending one step into the future. Conceptually, this representation implies that Feature A, at the present time, determines the value of Feature B at time $t+1$. Similarly, Feature B, at the present time, determines the value of Feature A at time $t+1$. When applying the prestige centrality models to this representation, the cycle features (A_t and B_t) acquire a centrality value equivalent to that of other features with exactly one dependent. Taken together, these simulations illustrate two points. First, it appears that the high levels of centrality given to veridically represented cycle features are not due to some idiosyncrasy of any one particular variant of eigenvector centrality but, rather, are essential to the prestige centrality notion itself. Second, it appears that a simple, psychologically plausible modification to the underlying causal structure can produce centrality values that match the results of the present study, regardless of the specific variant of centrality computed.

The open-source software package ConceptBuilder/ConceptAnalysis (Kim & Park, 2009) can compute centrality predictions of the Sloman et al. (1998) model for real-world causal models. Currently, we are extending the scope of ConceptAnalysis to enable researchers to calculate alpha centrality and PageRank centrality predictions for causal model data collected using ConceptBuilder.

APPENDIX B
Artificial Concept Names and Features for Experiments 1–6

| Concept Name | Feature 1 | Feature 2 | Feature 3 | Feature 4 |
|---|--------------------------------------|---|--------------------------------------|---------------------------------|
| Experiments 1 and 6, Artificial Disorder Concepts With Blank Properties | | | | |
| Dysnoresis | Symptom G | Symptom Z | Symptom Q | *Symptom T |
| Methinismus | Symptom P | Symptom F | Symptom V | *Symptom C |
| Philips' syndrome | Symptom R | Symptom W | Symptom H | *Symptom M |
| Hadronuria | Symptom N | Symptom X | Symptom D | *Symptom L |
| Experiments 2–4, Artificial Disorder Concepts (Modified to Create Cycles From Kim & Ahn, 2002b) | | | | |
| Dysnoresis | Overly dependent on others | Inflated self-esteem | Preoccupation with feelings of envy | Inability to feel pleasure |
| Methinismus | Intense fear of gaining weight | Recurrent unjustified suspicions of spouses' infidelity | Unreasonably scorning authority | Exaggerated startle response |
| Philips' syndrome | Persistent lack of self-control | Flamboyant, dramatic speech | Fear of going to sleep | Unkempt appearance |
| Hadronuria | Is easily fatigued | Lack of empathy | Depersonalization (feeling detached) | Feels compelled to keep talking |
| Experiment 5, Natural and Artifactual Kinds (Modified to Create Cycles From Rehder, 2003a) | | | | |
| Neptune Military Personal Computers | Magnetic power supply | Fast clock | Bright screen image | Hot temperature |
| Meteoric sodium carbonate | Radioactivity | Atoms arranged in eight-bond pyramids | Very reactive | A positive charge |
| Kehoe ants | Blood high in iron sulfate | Hyperactive immune system | Fast nest building | Thick blood |
| Lake Victoria shrimp | High quantity of acetylcholine (Ach) | Long-lasting flight response | High body weight | Accelerated sleep cycle |

*Experiment 6 only.

APPENDIX C
Sample Page From the Experimental Packets (Experiment 3)

The following diagram indicates the four symptoms of Hadronuria.



In the diagram above, an arrow between two symptoms indicates the direction in which one causes the other.

The following facts are true of the symptoms of Hadronuria. **Being easily fatigued** tends to cause these patients to be irritable and self-pitying, the combination of which renders them **incapable of empathizing with others' suffering. Their lack of empathy**, in turn, tends to cause these patients to feel **easily fatigued** when the suffering of others pushes them to become agitated. **Depersonalization** (feeling detached from oneself) tends to compel these patients to **keep talking**, in order to assure themselves that they are really there.

Now, imagine to yourself a patient who has all the symptoms of **Hadronuria**. When you have done so, please answer each of the following questions on a scale of 0 to 100, where 0 = definitely no and 100 = definitely yes.

If a patient has all the symptoms of Hadronuria EXCEPT that he or she does NOT have the symptom of being easily fatigued, does the patient have Hadronuria? Your answer (0–100): _____

If a patient has all the symptoms of Hadronuria EXCEPT that he or she does NOT have the symptom of feeling compelled to keep talking, does the patient have Hadronuria? Your answer (0–100): _____

If a patient has all the symptoms of Hadronuria EXCEPT that he or she does NOT have the symptom of lack of empathy, does the patient have Hadronuria? Your answer (0–100): _____

If a patient has all the symptoms of Hadronuria EXCEPT that he or she does NOT have the symptom of depersonalization, does the patient have Hadronuria? Your answer (0–100): _____