

Beliefs about the Causal Structure of the Self-Concept Determine Which Changes Disrupt  
Personal Identity

Stephanie Y. Chen, Oleg Urminsky, and Daniel M. Bartels

Booth School of Business

University of Chicago

Forthcoming, *Psychological Science*

\*\*Initial unreviewed version – please contact the authors for the final version.\*\*

Please address correspondence to:  
Stephanie Y. Chen  
Booth School of Business  
University of Chicago  
5807 S. Woodlawn Ave  
Chicago, IL 60637  
[stephanie.chen@chicagobooth.edu](mailto:stephanie.chen@chicagobooth.edu)

**Abstract**

Personal identity is an important determinant of behavior, yet how people mentally represent their self-concept is not well understood. In the studies reported in this paper, we examine the age-old question of what makes us who we are. We propose a novel approach to identity which suggests that the answer lies in people's beliefs about how the features of identity (e.g., memories, moral qualities, personality traits) are causally related to each other. Features that are involved in many cause-effect relationships with other features of one's identity are perceived as more defining to a person's self-concept. In three experiments, using both measured and manipulated causal centrality, we find support for this approach. For both judgments of one's self and of others, we find that some features are perceived as more causally central than others and that changes in those more causally central features are believed to be more disruptive to identity.

Personal identity provides us with valued norms to follow, scripts for behaviors, and ways to interpret our actions (Akerlof & Kranton, 2000, 2010; Markus & Wurf, 1987; Turner, 1985), and impacts a wide range of in-lab and real-world decisions (e.g., Bertrand, Kamenica, & Pan, 2015; Bryan et al., 2011; Cohn, Fehr, & Maréchal, 2014). In particular, the sense of continuity in one's identity (e.g., connectedness, Bartels & Rips, 2010) provides motivation for making far-sighted choices (Bartels & Urminsky 2011, 2015).

This paper explores how people represent the self and what features of identity people believe they need to retain to remain the same person. Prior literature has debated the relative importance of different specific features of identity, including social categories, memories, tastes, personality, and morality (Blok, Newman, & Rips, 2005; Haslam, Bastien, & Bissett, 2004; Strohminger & Nichols 2014, 2015). We introduce and test a novel *causal self hypothesis*, which proposes that identity is more than simply a set of features, and is the result of the complex interactions between them.

Building on the concepts and categories literature in cognitive psychology, we propose that people's representations of identity incorporate causal relationships between the features of identity. In general, features of a concept that are more causally central (i.e., linked to many other features of the concept, Ahn, Kim, Lassaline, & Dennis, 2000; Rehder, 2003; Rehder & Hastie 2001; Sloman, Love, & Ahn, 1998) are more defining of a concept. So, we predict that people believe their most causally central features are most defining of their identity. For example, the importance of memories, traits, or preferences for the self-concept depends on how these features are causally related to each other and to other features of identity.

The centrality of a feature can be thought of as the number of other features it is causally linked to, whether as a cause or an effect (Rehder, 2003). Alternatively, the dependency model

(Sloman et al., 1998) suggests that causes are more central than their effects, as centrality also depends on a feature's causal depth—a measure of all the feature's direct and indirect downstream effects. While we will focus on causal connections, we also investigated the suitability of causal depth for understanding when changes in a feature will be most disruptive to identity.

In the first two experiments, we measured people's beliefs about the causal relations between features of identity and how much changes in these features would disrupt identity. If more causally central features are more defining to identity, as hypothesized, the causal centrality of features should be positively correlated with the perceived disruption to identity resulting from a change in the feature. In a third experiment, we manipulate the causal relationships between features and test the impact on how disruptive changes in those features are. Although ideas about causal centrality have been extremely influential in the study of concepts, this is the first time that this approach has been incorporated into a theory of personal identity.

## **Experiment 1**

### **Method**

The prior literature on concepts and categories (Sloman et al., 1998) found Spearman correlations between centrality and importance of approximately .6. Power tables (Bonett & Wright, 2000) suggested a sample size of 86 and we set a target of approximately 90. Ninety-two University of Chicago students completed two tasks in randomized order, a concept map task and a survey about personal identity. Twelve participants were excluded due to computer program failures, either to record data or to display randomized features, yielding 80 cases.

Each participant drew a computerized map of the causal links between 16 features of their identity (Table 1), and answered survey questions about the features. Twelve of the sixteen

features were intended to be high importance, chosen from categories of personal identity identified as important in the prior literature (memories, personality, morality, and preferences/desires; e.g., Strohminger & Nichols, 2014). The remaining four features were intended to be low importance. Two were found, in previous research, to be less important for identity (instances of semantic memories, Strohminger & Nichols, 2014) and two (fillers) were peripheral features found to be unimportant for identity in a pretest.

Table 1.

*Features Used in Experiment 1*

<b>Feature</b>	<b>Category</b>
Cherished memories of time with parents/family	Autobiographical memory
Important childhood memories	Autobiographical memory
Memories of important life milestones	Autobiographical memory
Height	Filler
Level of Hunger	Filler
Level of Wholesomeness	Morality
Level of Honesty	Morality
Level of Loyalty	Morality
Intelligence Level	Personality
Degree of Shyness	Personality
Reliability	Personality
Goals for personal life	Preferences/desires
Favorite Hobbies/Activities	Preferences/desires
Aesthetic Preferences	Preferences/desires
Knowledge of math	Semantic memory
Knowledge of music	Semantic memory

Participants used ConceptBuilder software to report beliefs about causal relations (Kim & Park, 2009), first in an unrelated practice task and then to draw the causal map of personal identity using the 16 features. The features were initially presented on the screen in random order. Participants could move the features around and draw unidirectional or bidirectional arrows between them to represent cause-effect relationships, as they saw fit (see Figure 1). For

each link specified, participants also rated the strength of the causal relationship (1=weak, 2=moderate, 3=strong).

In the survey, participants rated each feature on how much (i) a change in the feature would disrupt their identity, and (ii) they expected the feature to change in the next five years.<sup>1</sup> We found that the measures of causal centrality did not relate to the ratings of expected change (see SOM-U) so we will focus on the relationship between causal centrality and disruption to identity.

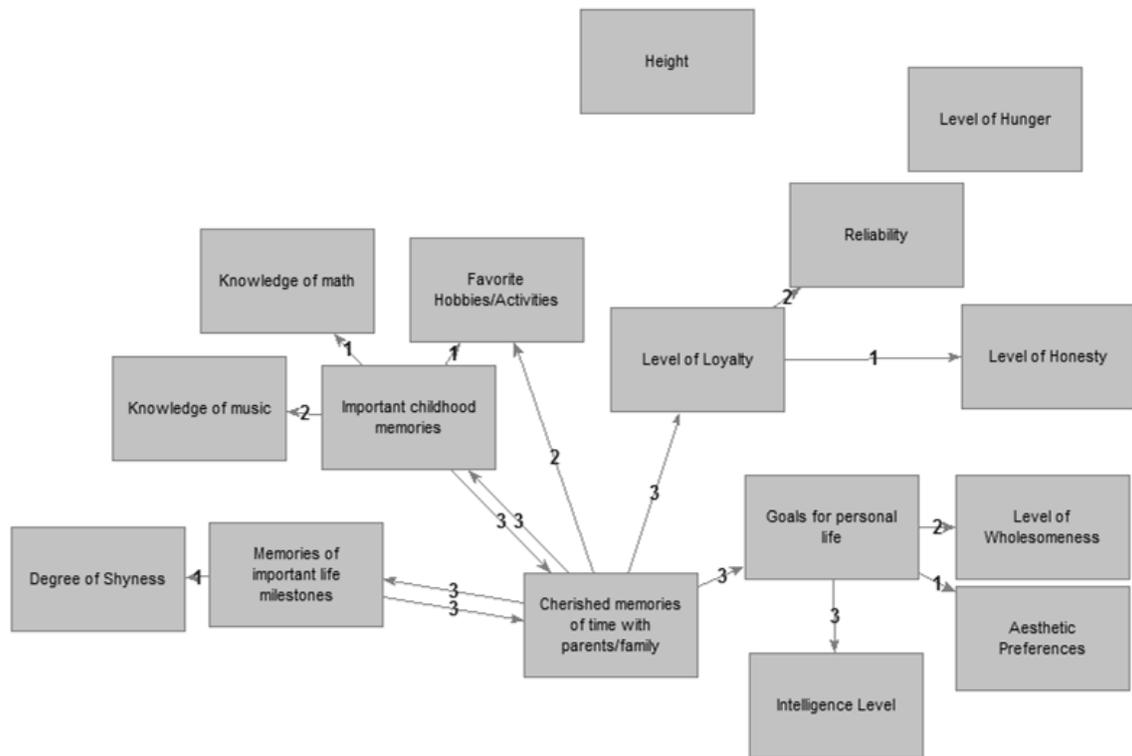


Figure 1. Example of a self-concept map. Each box contains one of the 16 features of personal identity. The arrows represent causal relationships between features. The arrow starts at the cause feature and points to the effect feature. The numbers that are on each arrow indicate the strength of the causal relationship (1=weak, 2=moderate, 3=strong).

<sup>1</sup> Wording provided in SOM-U.

### **Pre-test**

A pre-test was conducted to check whether people viewed the concept map task as capturing how they thought about their identity. Fifty participants generated two representations of their identity: a concept map of important features of their identity and a list of the most important social categories to their identity. The majority of participants chose the concept map when asked which representation would 1) better express their self-concept (58%) and, 2) better allow someone else to understand their true self (68%). Overall, participants endorsed the concept map representation significantly more than the list of categories on a composite measure of the two items ( $M=.63$ , 95% CI=[.51, .75]),  $t(49)=2.10$ ,  $p=.041$ . This confirmed that people generated concept maps they believed reflected important information about their identity.

### **Results**

On average, participants drew 20.0 causal links between the features. Our analyses use the number of causal connections (the number of other features a target feature is directly linked to, either as a cause or as an effect) as the measure of causal centrality (more links indicating greater centrality). This operationalization is consistent with the generative model of causal categorization (Rehder, 2003).

We also calculated centrality based on the causal depth of a feature using the dependency model of causal centrality (Sloman et al., 1998, see the SOM-R for details). This measure also takes into account each feature's position in the causal chain (treating causes as more central than their effects) and the rated strength of each link. In Experiments 1 and 2, we find similar but weaker results using the causal depth measure. Accordingly, we focus instead on the causal connections measure of centrality, and discuss causal depth analyses in the SOM-R.

A repeated-measures ANOVA revealed substantial variation in the causal connections across the features,  $F(15, 789)=25.56, p<.001$ . Personal goals, intelligence level, and autobiographical memories had the most links to other features of identity (Figure 2). In contrast, the low-importance features (fillers and semantic memories) were significantly less causally central than the high-importance set of features of identity ( $M_{Low}=1.2, M_{High}=2.9, t(14)=4.40, p=.001$ ; Figure 2), as expected.

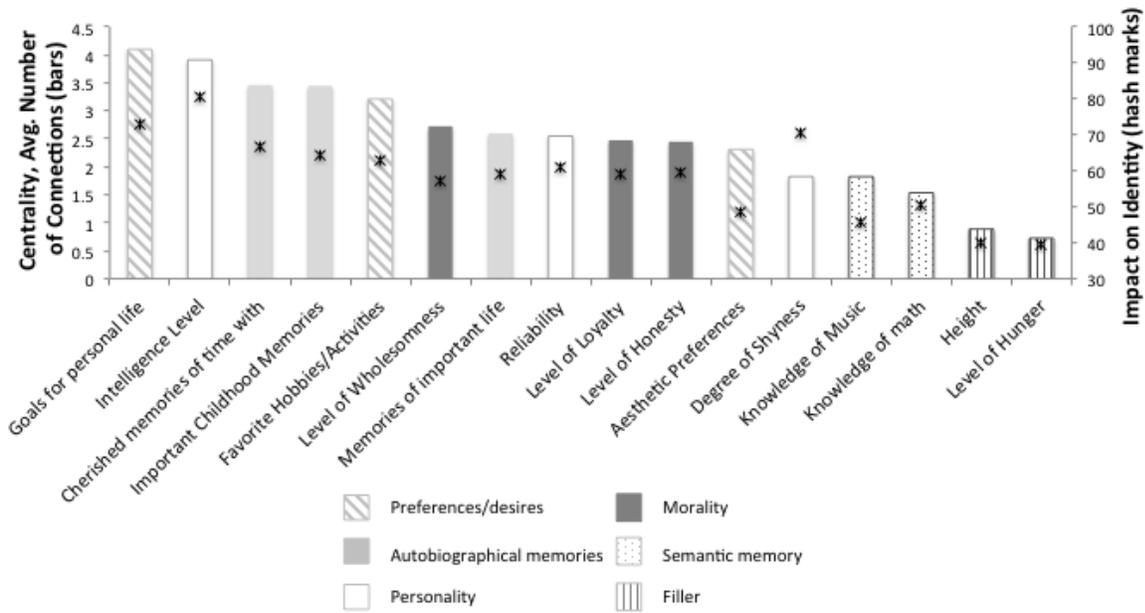


Figure 2. Results of Experiment 1. The causal connections of each feature of personal identity is shown on the left y-axis and by the bars (larger numbers indicate greater causally central). The impact a change in each feature has on identity is shown on the right y-axis and by the hash marks (larger numbers indicate greater disruption of identity).

Consistent with our hypothesis, changes in features with more causal connections were rated as more disruptive to identity. We found a significant overall Spearman correlation between the average causal connections and rated disruption to identity ( $r=.79, p<.001$ ). This positive relationship between causal connections and disruption to identity was observed at the individual

level for 80% of participants. The mean individual-level correlation between feature centrality and disruption to identity<sup>2</sup> was significantly positive ( $\bar{r}=.33$ , 95% CI = [.24, .42]),  $t(79)=7.43$ ,  $p<.001$ . Similar results were found when analyzing only the 12 high-importance features (see SOM-U). These results suggest that causal connections consistently influence identity judgments. The more causally central a feature was, the more strongly participants believed that a change to that feature would disruptive their identity.

### Experiment 2

Experiment 1 found that changes in more causally central features of identity were seen as more disruptive to the continuity of specifically one's own identity. Although our perceptions and evaluations of ourselves can be strikingly different from how we perceive others (Jones & Nisbett, 1972; Pronin, 2008), people use analogies to the self in forming judgments of even dissimilar others (Orhun & Urminsky, 2013). Experiment 2 tested whether the findings generalize to judgments of other people's identities.

### Method

Power analysis from Experiment 1 suggested sample sizes of 80 per cell. We recruited 250 U.S. Amazon Mechanical Turk respondents for a three-cell design. Eleven participants were excluded before analysis, either because of a scripting error (5), failed attention check (4), or giving all the same answers (2), yielding 239 cases. Similar results were found when including all participants who provided useable data (see SOM-U). Participants were assigned to one of three conditions (self, close-other, generic-other), and task order (causal relationships first, identity questions first) was counterbalanced. Participants in the self condition completed a causal centrality task and the identity disruption questions from Experiment 1 for features of

---

<sup>2</sup>Correlations reported are Spearman's rho. Fisher transformations were performed prior to t-tests.

their own identity. Participants in the close-other condition did the same tasks for a non-romantic close other they specified, while participants in the generic-other condition completed the tasks for a generic other person.

To measure centrality, participants completed a “listing causal relationships” task, using the same 16 features of personal identity from Experiment 1. After a practice task with feedback, participants completed 16 individually randomized trials. In each trial, participants were shown a different target feature and indicated which of the other 15 features, if any, were caused by the target feature (see Figure 3). Then, for all the features selected as direct effects of the target feature, participants rated the strength of the relationship. In a separate pre-test, we confirmed that this task yielded similar causal centrality scores as Experiment 1’s concept map task (see SOM-R).

**Think about your Aesthetic Preferences**

Which of the other features of your personal identity listed below, if any, are caused by your Aesthetic Preferences?

You may select as many or as few features as you see fit. In the below list, please select all the features that you believe are caused by the above feature.

- Level of Loyalty
- Reliability
- Cherished memories of time with parents/family
- Level of Wholesomeness
- Degree of Shyness
- Intelligence Level
- Goals for personal life
- Important childhood memories
- Level of Hunger
- Knowledge of math
- Knowledge of music
- Favorite Hobbies/Activities
- Memories of important life milestones
- Height
- Level of Honesty
- None of these features are caused by my Aesthetic Preferences

*Figure 3.* Illustration of listing causal relationships task used in Experiment 2.

## Results

On average, participants reported 75.7 causal links between the 16 features of identity. The number of links chosen did not differ by condition ( $M_{\text{self}}=71.1$ ,  $M_{\text{close-other}}=78.1$ ,  $M_{\text{generic-other}}=77.9$ ,  $F(2, 238)=.69$ ,  $p>.250$ ), suggesting participants perceived similar causal complexity in others' personal identity as in their own.

An ANOVA with condition (self vs. close-other vs. generic-other) as a between-subjects factor and features as repeated measures found an effect of feature,  $F(15, 2113)=156.34$ ,  $p<.001$ , on causal connections, as in Experiment 1. There was also a significant condition  $\times$  feature interaction,  $F(15, 2113)=1.62$ ,  $p=.049$ , suggesting that differences in centrality across the

features varied by condition (see SOM-U). As expected, low-importance features were less central than high-importance features ( $M_{Low}=2.0, M_{High}=5.6$ )  $F(1, 42)=64.48, p<.001$ , and this difference did not vary by condition,  $F(1, 42)=.04, p>.250$ .

Changes in features with more causal connections were rated as more disruptive to identity in all three conditions (self:  $r=.60, p=.015$ ; close-other:  $r=.62, p=.013$ ; generic-other:  $r=.44, p=.093$ ). The majority of participants in all conditions had a positive individual-level correlation between features' causal connections and rated disruptiveness of change (77%, 84%, and 74% in the self, close-other, and generic-other conditions). On average, the individual-level correlations were significantly positive in all conditions ( $\bar{r}_{Self}=.34, p<.001$ ;  $\bar{r}_{Close-other}=.38, p<.001$ ;  $\bar{r}_{Generic-other}=.30, p<.001$ , see Table 2). Similar results were found for the individual-level analysis when including only the 12 high-importance features (see SOM-U).

Table 2.

*Summary of Experiment 2 Results*

Condition	Causal Connections Approach	
	Aggregate Spearman Correlation	Individual Spearman Correlations
Self	$r = .60, p = .015$	mean $r = .34, t(78)=7.29, p<.001, 95\% CI = [.24, .44]$
Close-other	$r = .62, p = .013$	mean $r = .38, t(78)=9.08, p < .001, 95\% CI = [.29, .46]$
Generic-other	$r = .44, p = .093$	mean $r = .30, t(80)=6.28, p<.001, 95\% CI = [.20, .39]$

*Note.* T-tests in Individual Correlations column are one-sample t-tests of the mean Spearman rho (with Fisher transformation) against 0.

To examine whether the strength of the individual-level correlation differed between conditions, we performed a one-way ANOVA. We found no difference by condition,  $F(2,$

238) = .81,  $p > .250$ , again suggesting that the relationship between causal connections and impact on identity is similar in magnitude for self and others (see Table 2). These results suggest that more causally central features are perceived as more defining of identity not only for the self, but also when thinking about and the self-concept of others.

### Experiment 3

The first two studies found strong correlational evidence that causal centrality determines how disruptive a change in a feature would be to personal identity. Experiment 3 manipulated the causal centrality of features in a set of vignettes to test whether making a feature more causally central impacts how defining that feature is for identity.

#### Method

Prior manipulations of centrality (Ahn et al., 2000) found large effects on how influential features were to categorization judgments ( $d = .8$ ). Power analysis suggested samples of 22 per cell so we set a target of approximately 30 per vignette version. Fifty-nine participants were recruited via Amazon Mechanical Turk. One participant was removed before analysis for answering all comprehension check question incorrectly, leaving 58 participants for analysis. Similar results were found including all participants (see SOM-U).

Six vignettes described the causal relationship between three salient features of a person, with one of the features presented as the cause of the other two, which were effects (see Table 3). Since, in the vignettes, the cause feature has more causal connections than the effect feature, it is relatively more causally central. Each vignette had two versions (A and B), manipulating the causal centrality of two focal features by switching which of the two features was a cause and which was an effect. For example, in one version (A) of the vignette, Jack's lonely memories caused both his shyness and solitary preferences. In the other version (B), Jack's shyness caused

both his memories of being a lonely child and his solitary preferences (see Table 3). So, the exact same features were counterbalanced to play both the cause and effect role, to control for any idiosyncratic influences of specific features.

Each participant read six vignettes, each of which included a diagram summarizing the information (see Table 3). Participants were randomly assigned to read one of the two versions of each vignette (e.g., Version A or B). A comprehension check was instituted on a separate screen after each vignette to ensure participants understood the causal structure.

Participants then read about two people, one missing the cause feature and one missing the focal effect feature. To measure which feature was seen as more defining to identity, we then asked participants which of the two people was most likely to be the character in the vignette. Participants then reported how plausible they found the vignette on a scale of 0 (“not at all plausible”) to 100 (“extremely plausible”). To ensure participants had carefully made their selections, on a separate screen they again reported which person was more likely to be the character in the vignette. Finally, to promote close reading, participants answered a simple multiple-choice question about the details of the story.

Table 3.

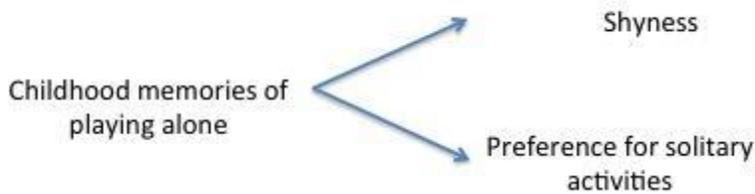
*Example Story Used in Experiment 3*

Version A

Jack has a few salient characteristics. Among them are that he is very shy, he likes solitary activities, and he has very strong memories from his childhood.

Jack didn't have many friends as a child so many of his memories from childhood are of him playing alone. As a result of these memories and experiences, he has always been shy and his favorite activities are generally solitary ones like building model airplanes. He has always thought his memories of his childhood experiences caused both his shyness and his preference for solitary activities.

This information can be summarized as follows:

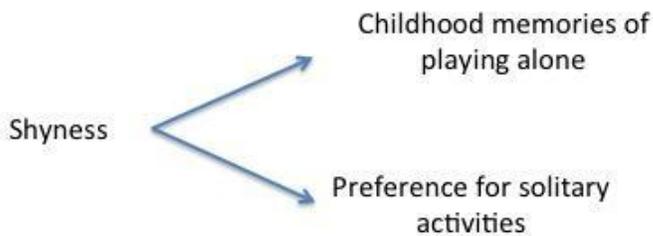


Version B

Jack has a few salient characteristics. Among them are that he is very shy, he likes solitary activities, and he has very strong memories from his childhood.

Jack has always been shy. As a result of his shyness, he didn't have many friends as a child so many of his memories from childhood are of him playing alone and his favorite activities are generally solitary ones like building model airplanes. He has always thought shyness caused both these memories of his childhood experiences and his preference for solitary activities.

This information can be summarized as follows:



## Results

We excluded trials where participants misreported the causal relationship of the features in the comprehension check (9% of trials) or provided inconsistent answers of which person was the character in the vignette when asked a second time (3% of trials). Similar results were found when no trials were excluded (see SOM-U).

The dependent measure was the average percentage of trials in which participants selected the person who was missing the effect feature. The causal self hypothesis predicts that participants will believe that the person missing the causally peripheral effect feature is the same person, rather than the person missing the causally central cause feature. This is what we found. Participants were significantly more likely to select the person missing the effect feature ( $M=68\%$ ,  $SD=23\%$ , 95% CI = [.62, .74]),  $t(57)=5.96$ ,  $p<.001$ ,  $d=.78$ , replicating the prior findings using an experimental manipulation of causal centrality. When a feature was made more causally central, changes in that feature were perceived as more inconsistent with continuity of identity.

Note that according to both approaches to causal centrality, the cause feature in this study is more causally central (both has more connections and is deeper in the causal chain). A similar experiment (reported in the SOM-U) pitted the two approaches against each other. The results of this experiment replicated the findings and were more consistent with the causal connections account of causal centrality.

We also looked at the effects of vignette plausibility. Representations of concepts, in general, are influenced by our prior knowledge (Murphy, 2002; Murphy & Medin, 1985). This suggests that prior intuitions about what causal relationships are likely to occur among features of identity may moderate identity judgments. The higher the perceived plausibility, the more likely that

participants believed that the effect feature was actually causally peripheral and the more likely that they selected the person missing the effect feature.

We found the predicted moderation by plausibility. Story plausibility was correlated with the proportion of selections of the person missing the causally peripheral effect feature ( $r = .65$ ,  $p = .023$ , 95% CI = [.12, .89]). The average proportion of selections of the person missing the effect feature was significantly higher among the six most plausible vignettes ( $M = 56.9\%$ ,  $SD = 15.8\%$ ) than the six least plausible vignettes ( $M = 80.2\%$ ,  $SD = 9.1\%$ ),  $t(10) = 2.55$ ,  $p = .029$ , 95% CI = [.03, .44]. This results suggests that participants' use of the causal information from the vignettes was moderated by how it fit with their beliefs about which causal relationships are relatively more likely to occur between features of identity.

### Discussion

People perceived more causally central features as being more necessary for continuity of identity, for both the self (Experiment 1) and others (Experiment 2). Manipulating the causal centrality of a feature changed perceptions of how defining that feature was to identity (Experiment 3).

Prior research has focused on comparing the individual importance of different types of features. These approaches seem to have missed a critical aspect of people's representations of identity, beliefs about the causal relationships between features. These relationships influence the extent to which a feature defines identity. These findings are consistent with people's general drive to explain the world using causal relations (Gopnik, 1998; Keil, 2006), and with narrative-based views of identity (McAdams, 2001, 2012).

While both the number of causal connections and causal depth related to how defining a feature was to identity, the number of causal connections was consistently more important (see

SOM-R for discussion). This suggests that features that either cause many other features or are caused by the combination of many other features (or both) will be most defining of identity. So, changes to features or the addition of new features may be less disruptive when people can causally connect these new aspects to existing features of identity. In fact, prior research has found that students whose personal narratives included more causal descriptions of experienced changes had greater emotional stability (Lodi-Smith et al., 2009).

Differences in people's beliefs about the causal structure of their identities may have important implications for understanding identity-based motivations for behavior. If people who anticipate disruptions to more causally central features are less connected to their future selves, they may make more short-sighted decisions (Bartels & Urminsky, 2011). The effectiveness of interventions that appeal to identity features (e.g., "be a voter", Bryan 2011) may depend on the causal centrality of the targeted feature.

Our representations of ourselves and others are not simply comprised of sets of features or social categories. These representations crucially incorporate beliefs about the causal relations between aspects of identity. The answer to the riddle of who we truly are lies at the nexus of causal connections between the features of our identity.

### References

- Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *Quarterly journal of Economics*, 115, 715-753.
- Akerlof, G. A., & Kranton, R. E. (2010). *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-being*. Princeton, NJ: Princeton University Press.
- Ahn, W., Kim, N. S., Lassaline, M. E., & Dennis, M. J. (2000). Causal status as a determinant of feature centrality. *Cognitive Psychology*, 41, 361-416.
- Bartels, D. M., & Rips, L. J. (2010). Psychological connectedness and intertemporal choice. *Journal of Experimental Psychology: General*, 139, 49.
- Bartels, D. M., & Urminsky, O. (2011). On intertemporal selfishness: How the perceived instability of identity underlies impatient consumption. *Journal of Consumer Research*, 38, 182-198.
- Bartels, D. M., and Urminsky, O. (2015). To know and to care: How awareness and valuation of the future jointly shape consumer savings and spending. *Journal of Consumer Research*, 41, 1469-1485.
- Bertrand, M., Kamenica, E., & Pan, J. (2015). Gender identity and relative income within households. *The Quarterly Journal of Economics*, 571, 614.
- Blok, S., Newman, G., & Rips, L. (2005). Individuals and their concepts. In W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. Wolff (Eds.), *Categorization inside and outside the laboratory: Essays in honor of Douglas L. Medin* (pp. 127–149). Washington, DC: American Psychological Association.
- Bonett, D. G., & Wright, T. A. (2000). Sample size requirements for estimating Pearson, Kendall and Spearman correlations. *Psychometrika*, 65, 23-28.

- Bryan, C.J., Walton, G.M., Rogers, T., & Dweck, C.S. (2011). Motivating voter turnout by invoking the self. *Proceedings of the National Academy of Sciences*, 108(31), 12653-12656.
- Cohn, A., Fehr, E., & Maréchal, M.A. (2014). Business culture and dishonesty in the banking industry. *Nature*, 516, 86-89.
- Haslam, N., Bastian, B., & Bissett, M. (2004). Essentialist beliefs about personality and their implications. *Personality and Social Psychology Bulletin*, 30, 1661-1673.
- Gopnik, A. (1998). Explanation as orgasm. *Minds and Machines*, 8, 101–118.
- Jones, E. E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones et al. (Eds.), *Attribution: Perceiving the causes of behavior*. Morristown, NJ: General Learning Press.
- Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology*, 57, 227-254.
- Kim, N. S., & Park, E. Y. (2009). ConceptBuilder: An open-source software tool for measuring, depicting, and quantifying causal models. *Behavior research methods*, 41(1), 128-136.
- Lodi-Smith, J., Geise, A., Robins, R. W., & Roberts, B. W. (2009). Narrating personality change. *Journal of Personality and Social Psychology*, 96, 679–689.
- Markus, H., & Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology*, 38, 299-337.
- McAdams, D. P. (2001). The psychology of life stories. *Review of general psychology*, 5, 100.
- McAdams, D. P. (2013). The psychological self as actor, agent, and author. *Perspectives on Psychological Science*, 8, 272-295.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological review*, 92(3), 289.
- Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

- Orhun, A. Y., & Urminsky, O. (2013). Conditional Projection: How Own Evaluations Influence Beliefs About Others Whose Choices Are Known. *Journal of Marketing Research*, 50(1), 111-124.
- Pronin, E. (2008). How we see ourselves and how we see others. *Science*, 320, 1177-1180.
- Rehder, B. (2003). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 1141-59.
- 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.
- Sloman, S., Love, B., & Ahn, W. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, 22, 189-228.
- Strohinger, N., & Nichols, S. (2014). The essential moral self. *Cognition*, 131(1), 159-171.
- Strohinger, N., & Nichols, S. (2015). Neurodegeneration and identity. *Psychological Science*, 26, 1468-1479.
- Turner, J. C. (1985). Social categorization and the self-concept: A social cognitive theory of group behavior. *Advances in group processes: Theory and research*, 2, 77-122.

### Appendix 1: Causal Depth Analyses

To calculate causal centrality based on the causal depth of a feature we used the dependency model of causal centrality (Sloman et al., 1998). According to this iterative model,  $C_i$ , the centrality of feature  $i$ , is determined (at each time step) by summing across the centrality of the concept's other features (at time,  $t$ ),  $c_{j,t}$ , multiplied by how dependent each feature,  $j$ , is on feature  $i$ ,  $d_{ij}$ :

$$C_{i,t+1} = \sum_j d_{ij} c_{j,t} \quad (1)$$

The implementation of the model is a repeated matrix multiplication that comes to a stable ranking within a small number of iterations (Sloman et al., 1998; Kim & Park, 2009). The ConceptBuilder software performs 15 iterations and the initial centrality of all features (at time 0) is set to 0.5.

As  $d_{ij}$  is a positive value when feature  $i$  causes feature  $j$ , according to this model, the deeper a feature is in the causal chain, the more defining it is to the concept. In the concept map task, the dependence,  $d_{ij}$ , is the value (1=weak, 2=moderate, 3=strong) that participants assigned to the strength of each causal relationship they drew.

#### Causal Depth Analyses – Experiment 1

We found evidence that this measure of causal centrality influenced identity judgments. The Spearman rank correlation between causal depth and disruption to identity ratings was significant,  $r = .65$ ,  $p = .008$ . The mean individual-level correlation between feature centrality and disruption to identity<sup>3</sup> was significantly positive ( $M = .23$ , 95% CI=[.14, .31]),  $t(79) = 5.27$ ,  $p < .001$ . This positive relationship between causal depth and disruption to identity was observed for 78% of subjects.

---

<sup>3</sup>All correlations reported are Spearman's rho. Fisher transformations were performed prior to t-tests.

### **Causal Depth vs. Causal Connections Approaches – Experiment 1**

To examine how each measure of centrality relates to disruption of identity, we regressed identity disruption ratings on both measures (*z*-scored within-subject) at the individual level. The mean coefficient for the causal connections term ( $\bar{\beta}_{Connections}=.28, SD=.30, 95\% CI=[.21, .35]$ ) was significantly positive,  $t(79)=8.36, p<.001$ , and greater than the mean coefficient for the causal depth term ( $\bar{\beta}_{Depth}=.02, SD=.29, 95\% CI=[-.05, .08]$ ),  $t(79)=5.02, p<.001$ . The coefficient for the causal connections term was larger than the coefficient for causal depth for 70% of participants. The mean coefficient for the causal depth term was not greater than zero,  $t(79)=0.52, p>.250$ . These results confirm that the causal depth measure does not significantly add to the predictive power of a model that includes the number of causal connections.

### **Pre-test Comparing Concept Map and Listing Causal Relationships Tasks**

In a separate pre-test, we confirmed that the “listing causal relationships” task used in Experiment 2 and the concept map task from Experiment 1 yielded similar causal centrality scores. Thirty subjects performed both tasks with the same 16 features of identity used in Experiments 1 and 2. For each subject, we calculated the Spearman correlation between the causal centrality of the 16 features generated by the two tasks. The average Spearman correlation was significantly greater than 0 for both the causal connections approach ( $M = .67, SD = .45, 95\% CI=[.51, .84]$ ),  $t(29) = 8.27, p < .001$  and the causal depth approach ( $M = .44, SD = .38, 95\% CI=[.30, .59]$ ),  $t(29) = 6.36, p < .001$ . The Spearman correlation was positive for the vast majority of subjects—29 out of 30 for the causal connections approach, and 26 out of 30 for the causal depth approach. So, the two tasks appear to measure causal centrality in a similar way.

### **Causal Depth Analysis – Experiment 2**

The results of the causal depth correlational analyses revealed a similar pattern to the causal connections analyses. The overall correlation between causal depth and disruption to identity was significant in the close-other condition, and marginally significant in the self and generic-other conditions. The results of the individual-level analysis revealed that the mean Spearman correlation for causal depth was significantly positive for all conditions ( $M_{\text{self}} = .26$ ,  $M_{\text{close-other}} = .32$ ,  $M_{\text{generic-other}} = .22$ ),  $t(78) = 5.14$ ,  $t(78) = 6.61$ ,  $t(80) = 4.78$ ,  $ps < .001$ . The majority of participants in all conditions had a positive individual-level correlation between features' causal depth and rated disruptiveness of change (72%, 80%, and 72% in the self, close-other, and generic-other conditions, respectively). A one-way ANOVA revealed that the mean Spearman correlation did not differ by condition,  $F(2, 238) = 1.14$ ,  $p > .250$ , suggesting that the relationship between causal depth and disruption to identity was similar across conditions (see Table S1).

Table S1. Summary of Experiment 2 Results

Condition	Causal Depth Approach	
	Aggregate Spearman Correlation	Individual Spearman Correlations
Self	$r = .49, p = .05$	mean $r = .26$ , $t(78) = 5.15$ , $p < .001$ , 95% CI = [.16, .36]
Close-other	$r = .65, p = .01$	mean $r = .32$ , $t(78) = 6.61$ , $p < .001$ , 95% CI = [.22, .42]
Generic-other	$r = .42, p = .11$	mean $r = .22$ , $t(80) = 4.79$ , $p < .001$ , 95% CI = [.13, .31]

*Note.* T-tests in Individual Correlations column are one-sample t-tests of the mean Spearman rho (with Fisher transformation) against 0.

### Causal Depth vs. Causal Connections Approaches – Experiment 2

To examine the relative impact of the two forms of causal centrality on disruption to identity scores, for each subject, we regressed identity disruption ratings on both measures of causal centrality (all measures z-scored within subject). We performed a 3 (condition: self, close-other,

generic-other) × 2 (causal centrality approach: causal connections vs causal depth) ANOVA on the resulting betas. The main effect of condition was not significant,  $F(2, 236) = 1.08, p > .250$ , nor was the condition × causal centrality approach interaction,  $F(2, 236) = .71, p > .250$ , suggesting that condition did not influence the predictive value of these two causal centrality measures.

We found a main effect of causal centrality approach,  $F(1, 236) = 36.13, p < .001$ . For all conditions the mean regression coefficient for the causal connections term ( $M_{self} = .26, M_{close-other} = .32, M_{generic-other} = .31$ ) was significantly positive,  $ts > 5.45, ps < .001$ , and greater than the mean coefficient for the causal depth term ( $M_{self} = .04, M_{close-other} = .03, M_{generic-other} = -.05$ ),  $ts > 2.8, ps < .01$  (see Table S2). The mean coefficient for the causal depth term was not significantly positive for any condition,  $ts < 1.03, ps > .250$ . Replicating the results of Experiment 1, the regression analysis suggests that the causal depth approach does not significantly add to the predictive power of a model that includes the number of causal connections.

Table S2. Summary of Experiment 2 Regression Results

Condition	Mean coefficient		t-tests Comparing Terms
	Causal Connections Term, $M (SD)$	Causal Depth Term, $M (SD)$	
Self	.26 (.42) 95% CI=[.16, .35]	.04 (.37) 95% CI=[-.05, .11]	$t(78) = 2.89, p = .005$
Close-other	.32 (.42) 95% CI=[.23, .42]	.03 (.45) 95% CI=[-.07, .13]	$t(78) = 3.43, p = .001$
Generic-other	.31 (.45) 95% CI=[.21, .41]	-.05 (.46) 95% CI=[-.16, .05]	$t(80) = 4.03, p < .001$

**Discussion**

In both Experiments 1 and 2 (and in an additional experiment presented in the SOM-U where causal centrality of features was manipulated), we found that the causal centrality of features of identity was more determined by the number of causal connections a feature was involved in, rather than the causal depth of the feature. As causes generally occur before their

effects, this means that the features that develop early in our lives (e.g., childhood memories) will not necessarily remain the most defining to identity, contrary to what a causal depth approach would suggest. Rather how defining an early-developing feature is to identity will depend on how many other features it causes. Late-developing features will be important if they are caused by the combination of other features and, in turn, cause new features.

## Appendix 2: Supplemental Materials

### Experiment 1

#### Wording of Survey Questions

Disruption to identity question: *We would now like to understand how a change in each of the features below would change your identity. That is, for each of the features below, imagine that you are completely different on that dimension (e.g., for height, if you are tall, imagine that something changed so that you are now short). Do you think that you would still be the same person you are now, or would you be a different person? Please indicate your answer with each of the sliders below where 0 means, "I would be the exact same person" and 100 means, "I would be a completely different person."*

Expected change question: *We would now like to understand how much you think each of the features below may change in the future. Please indicate how much you think each feature will change in the **next 5 years** with each of the sliders below where 0 means, "Will not change at all" and 100 means, "Will change completely."*

#### Relationship between Causal Centrality and Expected Change

We found that the two measures of causal centrality, casual connections and causal depth, did not relate to the ratings of expected change in each feature ( $\bar{r}_{Connections}=.07, p = .061, 95\% \text{ CI} = [0.0, .15]$ ;  $\bar{r}_{Depth}=-.03, p = .437, 95\% \text{ CI} = [-.11, .05]$ ). In fact, some of the most central features (childhood memories) and least central features (height) were expected to remain the most stable. This suggests that the causal centrality of a feature, although strongly related to the importance of the feature's stability for identity, is distinct from the anticipated likelihood of change in that feature.

#### Causal Centrality Analysis with only High Importance Features

**Causal Connections.** Changes to features with more causal connections on average were rated as moderately more disruptive to identity,  $r=.56$ ,  $p=.063$ . The majority of participants in all conditions had a positive correlation between features' causal connections and rated disruption (67% of participants). The mean individual-level correlation between feature centrality and disruption<sup>4</sup> was significantly positive,  $\bar{r}=.18$ , 95% CI = [.10, .27]),  $t(78)=4.26$ ,  $p<.001$ . One participant gave the same answer to the disruption to identity question for all high importance features and had to be excluded from this analysis.

**Causal Depth.** Changes to features with higher causal depth scores (see SOM-R for details on how this score was calculated) on average were not rated as more disruptive to identity,  $r=.25$ ,  $p>.250$ . However, the majority of participants in all conditions had a positive correlation between features' causal depth and rated disruption (61% of participants). The mean individual-level correlation between feature centrality and disruption was significantly positive,  $\bar{r}=.12$ , 95% CI = [.02, .22]),  $t(78)=2.44$ ,  $p=.017$ . One participant gave the same answer to the disruption to identity question for all high importance features and had to be excluded from this analysis.

---

<sup>4</sup>All correlations reported are Spearman's rho. Fisher transformations were performed prior to t-tests to better satisfy the normality assumption.

## Experiment 2

### Correlations Between Causal Centrality and Disruption to Identity with All Features

We performed both the aggregate and individual level Spearman correlational analysis including all participants (i.e. even those who failed the attention checks)<sup>5</sup>. The results are similar to those presented in the main manuscript and are summarized in Table SU1.

Table SU1. Summary of Experiment 2 results including all participants who failed attention checks

Condition	Causal Connections Approach	
	Aggregate Spearman Correlation	Individual Spearman Correlations
Self	$r = .60, p = .014$	mean $r = .34, t(79) = 7.20, p < .001, 95\% \text{ CI} = [.25, .43]$
Close-other	$r = .62, p = .013$	mean $r = .37, t(80) = 8.88, p < .001, 95\% \text{ CI} = [.28, .45]$
Generic-other	$r = .42, p = .110$	mean $r = .29, t(81) = 6.19, p < .001, 95\% \text{ CI} = [.20, .38]$

*Note.* T-tests in Individual Correlations column are one-sample t-tests of the mean Spearman rho (with Fisher transformation) against 0.

### Correlations Between Causal Centrality and Disruption to Identity with only High Importance Features

**Causal Connections.** Changes to features with more causal connections on average were not rated as more disruptive to identity, in all three conditions (self:  $r = .17, p = .588$ ; close-other:  $r = .18, p = .573$ ; generic-other:  $r = -.17, p = .604$ ). However, the majority of participants in all conditions had a positive correlation between features' causal connections and rated disruption

<sup>5</sup> All participants who provided usable data are included in this analysis. Correlations for two participants who gave all the same answers could not be calculated so they are excluded from this analysis. A technical error prevented five participants from viewing any features, so they are also excluded from this analysis.

(65% of participants in each condition). On average, the individual-level correlations were significantly positive in all conditions ( $\bar{r}_{Self}=.15, p=.003, 95\% \text{ CI} = [.05, .25]$ ;  $\bar{r}_{Close-other}=.17, p<.001, 95\% \text{ CI} [.09, .25]$ ;  $\bar{r}_{Generic-other}=.10, p=.056, 95\% \text{ CI} = [.00, .20]$ ).

**Causal Depth.** Changes to features with higher causal depth scores (see SOM-R for details on how this was calculated) on average were not rated more disruptive to identity in all three conditions (self:  $r=-.17, p=.588$ ; close-other:  $r=.22, p=.485$ ; generic-other:  $r=-.23, p=.471$ ). The majority of participants in all conditions had a positive correlation between features’ causal depth and rated disruption (55%, 63%, 59% of participants in the self, close-other, and generic-other conditions). However, the average individual-level correlations was significantly positive in only the close-other condition ( $\bar{r}_{Self}=.08, p=.127, 95\% \text{ CI} = [-.02, .19]$ ;  $\bar{r}_{Close-other}=.12, p=.015, 95\% \text{ CI} [.03, .22]$ ;  $\bar{r}_{Generic-other}=.07, p=.144, 95\% \text{ CI} = [-.03, .17]$ ).

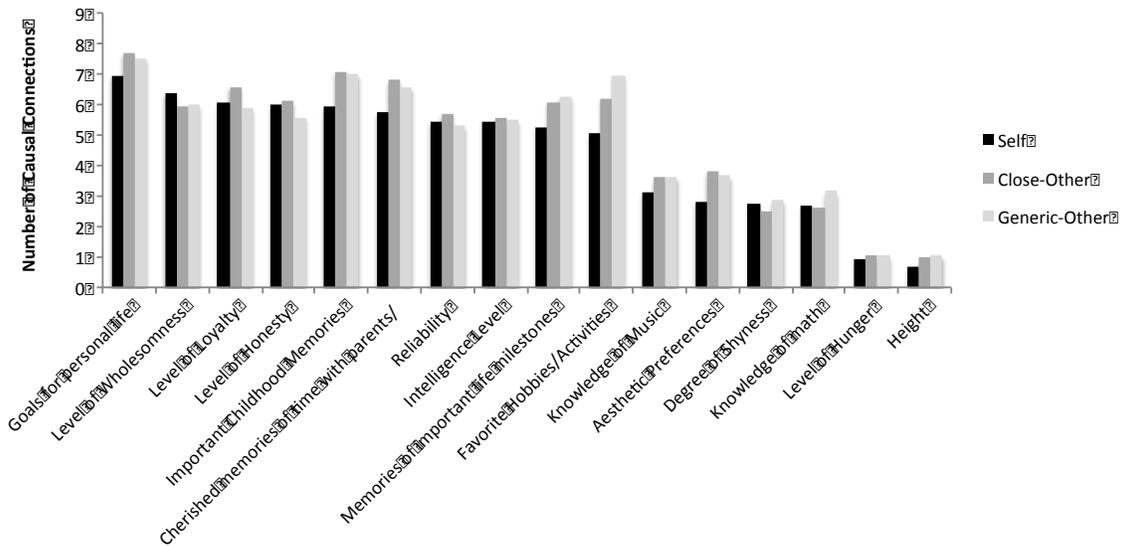


Figure SUI. Results of Experiment 2. Number of causal connections for each feature by condition.

**Experiment 3**

Table SU2. Vignettes used in Experiment3 and Supplemental Experiment 4\*

Version A	Version B
Jack has a few salient characteristics. Among them are that he is very shy, he likes solitary activities, and he has very strong memories from his childhood.	Jack has a few salient characteristics. Among them are that he is very shy, he likes solitary activities, and he has very strong memories from his childhood.
Jack didn't have many friends as a child so many of his memories from childhood are of him playing alone. As a result of these memories and experiences, he has always been shy and his favorite activities are generally solitary ones like building model airplanes. He has always thought his memories of his childhood experiences caused both his shyness and his preference for solitary activities.	Jack he has always been shy. As a result of his shyness, he didn't have many friends as a child so many of his memories from childhood are of him playing alone and his favorite activities are generally solitary ones like building model airplanes. He has always thought shyness caused both these memories of his childhood experiences and his preference for solitary activities.
Mary has a few salient characteristics. Among them are that she is very ambitious, her professional goal is to become a neurosurgeon, and many of her important memories are of her various academic accomplishments.	Mary has a few salient characteristics. Among them are that she is very ambitious, her professional goal is to become a neurosurgeon, and many of her important memories are of her various academic accomplishments.
As long as anyone can remember, Mary wanted to be a neurosurgeon. As a result of this goal, Mary became a very good student so many of her important memories are of her various academic accomplishments. Her goal also led her to develop a very high level of ambition. She has always thought that her desire to become a neurosurgeon caused both these important memories of her accomplishments and her ambitious personality.	As long as anyone can remember, Mary has been very ambitious. As a result of her ambition, Mary was a very good student so many of her important memories are of her various academic accomplishments. Her ambition also led her to develop the professional goal of being a neurosurgeon. She has always thought that her ambitious personality caused both her desire to be a neurosurgeon and her memories of her academic experiences.
Henry has a few salient characteristics. Among them are that he is extremely honest, he has many memories of the lessons his parents taught him as a child, and he has a lot of close friends.	Henry has a few salient characteristics. Among them are that he is extremely honest, he has many memories of the lessons his parents taught him as a child, and he has a lot of close friends.
When he was a child, Henry's parents emphasized the importance of honesty and he has many memories of them praising him for being honest. As a result of these lessons and memories, Henry has always been able to have many close friends and he developed into an extremely honest person; no one can remember an instance of Henry being dishonest. Henry has always thought that his memories of his parents emphasis on and praise of honesty caused his ability to have many close friends and his extreme honesty.	Henry has always been an extremely honest person; no one can remember an instance of Henry being dishonest. As a result of his honesty, Henry has always been able to have many close friends and has many memories of his parents praising him for being honest and emphasizing the importance of honesty. Henry has always thought that his honesty caused both his parents to emphasize these life lessons and his ability to have so many close friends.

\*Vignettes modified in Supplemental Experiment 4 to have four features and to accommodate a common effect structure.

Version A	Version B
<p>Jane has a few salient characteristics. Among them are that she is very intelligent, she has many cherished childhood memories of time spent with her parents, and her favorite activities generally involve learning new things.</p> <p>As long as anyone can remember, Jane has always been very intelligent. Her intelligence caused her parents to teach her about everything from science to music and she has many memories of her parents teaching her all sorts of things about the world. Her intelligence also cause her develop a love for activities that involve learning—one of her favorite activities is going to museums. Jane has always thought that her intelligence caused both her memories of her parents teaching her about the world and her love of learning-related activities.</p>	<p>Jane has a few salient characteristics. Among them are that she is very intelligent, she has many cherished childhood memories of time spent with her parents, and her favorite activities generally involve learning new things.</p> <p>As long as she can remember, her parents always wanted to teach her about the world. From an early age they taught her about everything from science to music. As a result of these memories, Jane developed into a very intelligent person and her favorite activities generally involve learning new things—one of her favorite activities is going to museums. Jane has always thought that her memories of her parents teaching her about the world caused both her intelligence and her love of learning-related activities.</p>
<p>Anne has a few salient characteristics. Among them are that she is very open-minded, she has always preferred to be friends with quirky people, and she loves adventurous activities.</p> <p>As long as anyone can remember, Anne has always been very open-minded. As a result, Anne developed a love for adventurous activities and a preference to be friends with people who had quirky personalities. She has always thought her open-mindedness caused both her preferences in friends and her love of adventurous activities.</p>	<p>Anne has a few salient characteristics. Among them are that she is very open-minded, she has always preferred to be friends with quirky people, and she loves adventurous activities.</p> <p>As long as anyone can remember, Anne has always preferred to be friends with people who have quirky personalities. As a result, Anne developed a love for adventurous activities and an open-minded personality. She has always thought her preference in friends caused both her open-mindedness and her love of adventurous activities.</p>
<p>Robert has a few salient characteristics. Among them are that he is very loyal, he has many fond memories of childhood, and his favorite places are the same as they were when he was a child.</p> <p>As long as anyone can remember, Robert has always been very loyal. He is extremely loyal to those who are close to him. As a result of his loyalty, he has many fond memories of his childhood friends who have remained his best friends for his entire life. This is also true of his favorite places. His favorite place to visit is his hometown and his favorite restaurant has been the same since he was five. Robert has always thought his loyalty caused both his fond memories of his childhood friends and his preferences for places to visit and eat.</p>	<p>Robert has a few salient characteristics. Among them are that he is very loyal, he has many fond memories of childhood, and his favorite places are the same as they were when he was a child.</p> <p>Robert has many fond memories of his childhood spent with his childhood friends who have remained his best friends for his entire life. As a result of these friendships, he developed a very loyal personality; he extremely loyal to those who are close to him. This is also true of his favorite places. His favorite place to visit is his hometown and his favorite restaurant has been the same since he was five. Robert has always thought his memories of his childhood friendships caused him develop both his loyal personality and his preferences for places to visit and eat.</p>

### Results (No Trials Excluded)

According to our hypothesis, removing a causally central feature should be more disruptive to identity than removing a causally peripheral feature. So, we predicted that participants should

be more likely than chance (50%) to pick the individual missing the causally peripheral (effect) feature as being the same person as the character in the vignette. Consistent with the hypothesis, participants were significantly more likely to select the person missing the causally peripheral feature than the person missing the causally central feature ( $M = 66\%$ ,  $SD = 23\%$ , 95% CI=[.60, .72]),  $t(58) = 5.6$ ,  $p < .001$ .

### Supplemental Experiment 4

The aim of Experiment 4 was to understand more precisely how causal beliefs influence identity judgments, distinguishing between causal connections and causal depth. Experiment 4 was similar to Experiment 3 except that we contrasted a common cause structure, as in Experiment 3, with a common effect structure, which allows us to distinguish between the two approaches to defining causal centrality. The common effect structure had one effect feature which was jointly caused by three other features (see Figure SU2). As in Experiment 3, participants read six vignettes that described the causal relationships between the salient features of individuals. Half of these vignettes described a common cause structure and the other half described a common effect structure. The focal task was again to select which of two individuals, one missing the effect feature and one missing a cause feature, was more likely to be the character in the story.

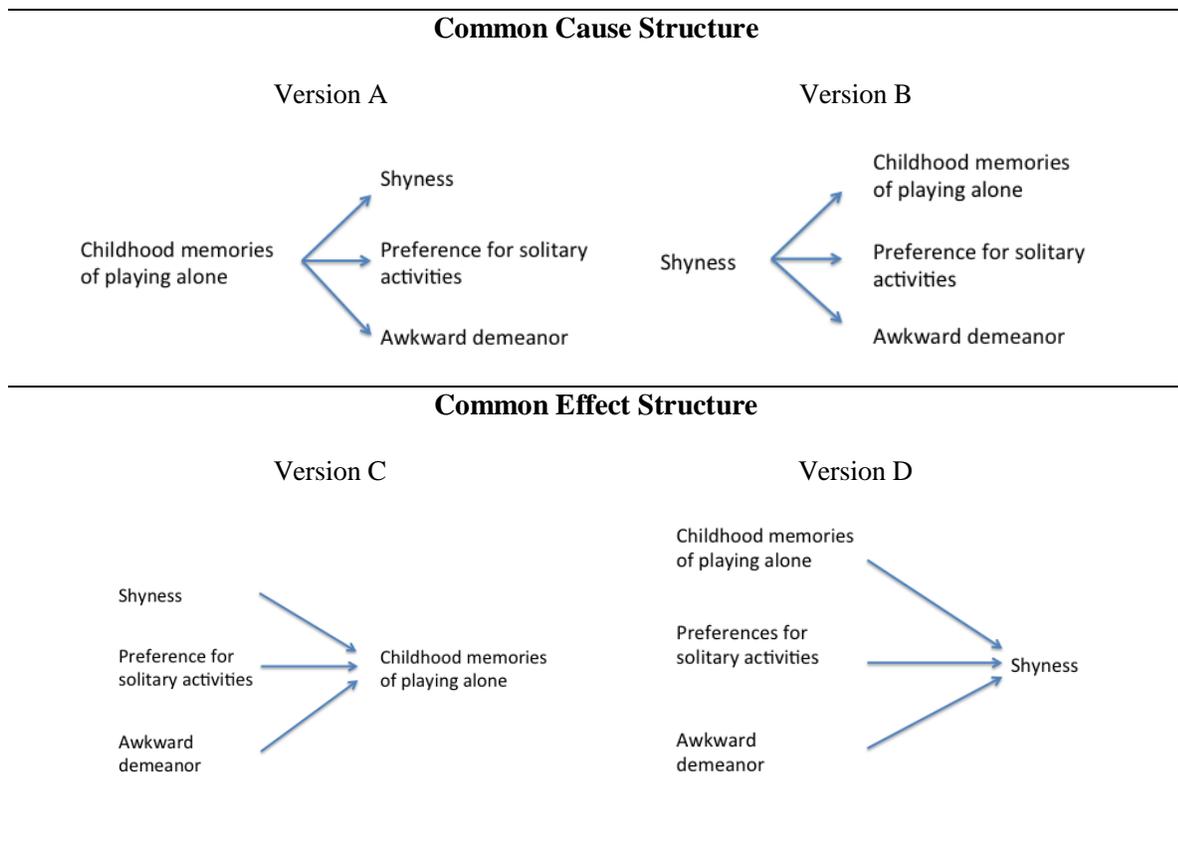
The two accounts of causal centrality make different predictions about which individual should be selected in the common effect vignettes. The causal depth approach predicts that the individual missing the effect is more likely to be judged as the character in the story. This is because features deeper in the causal chain (the cause features) are more important to categorization (in this case, to identity judgments) and retaining that feature is therefore more important to identity continuity. In contrast, the number of causal relationships approach predicts that participants will tend to pick the person missing the cause feature. This is because, according to this approach to causal centrality, retention of the causal links is important and removing the effect (while retaining all three causes) disrupts three causal links while removing a cause disrupts only one link.

## Method

Sixty participants were recruited via Amazon Mechanical Turk, and randomly assigned to read one of two sets of vignettes. Four participants were removed before analysis, either for answering the comprehension check question wrong or for failing an attention check, leaving 56 participants for analysis. Similar results were found including all participants.

Participants were instructed that they would read six short passages about how a person believed he became who he is and answer related questions. These vignettes would be accompanied by diagrams that summarized the information and could be used to answer the questions.

Each vignette described the causal relationship between four salient features of a person. Half the vignettes described a common cause structure, where one feature caused the other three. The other half described a common effect structure, where three features jointly caused the other one. For example, a participant might read that Jack's memories of being a lonely child caused his shyness, his preference for solitary activities and his awkward demeanor (common cause: Figure SU2, Version A). Alternatively, they could have instead read that his shyness, solitary preferences and demeanor caused his childhood memories (common effect: Figure SU2, Version C).



*Figure SU2.* Structure of vignettes used in Experiment 4. There were four versions of each vignette: two versions for each causal structure. Versions A/B and C/D counterbalanced the placement of the target features as cause or effect.

Each vignette had two versions, manipulating the causal centrality of two focal features by switching which of the two features was a cause and which was an effect. For example, one common cause vignette presented Jack’s childhood memories as causing the other three features, including shyness (Version A). The other presented Jack’s shyness as causing the other three features, including his memories (Version B). Likewise, one version of the common effect vignette presented Jack’s childhood memories as an effect of the other three features, including

shyness (Version C). The other presented Jack's shyness as an effect of the other three features, including his memories (Version D).

Based on the number of causal connections measure, shyness would be more causally central than childhood memories in Versions B and D, whereas childhood memories would be more causally central than shyness in Versions A and C. This difference between the common cause and common effect versions allows us to contrast the causal connections and causal depth approaches to causal centrality. For example, although shyness has more causal connections than memories in both Versions B and D, shyness has more causal depth (i.e. occurs earlier in the causal chain) in Version B, but memories has more causal depth than shyness in Version D.

Before reading the vignettes, participants saw an example diagram and were instructed on its interpretation (i.e., arrows started at the cause features and pointed to the effect features). Participants then performed four practice trials, with feedback, where they saw a diagram and selected the correct interpretation from three options.

Each participant then read six vignettes, including three displaying a common cause and three displaying a common effect. Which vignettes displayed a common cause vs. common effect were counterbalanced between-participants. The order of common cause vs. common effect vignettes was randomized within-participants, and participants were randomly assigned to read one of the two versions of each vignette (e.g., Version A or Version B in the common cause version). The exact same features played both the cause and effect role in both causal structures, to control for any idiosyncratic influences of specific features.

A comprehension check was instituted on a separate screen after each vignette to ensure participants understood the causal structure. Participants then read about two people, one missing the focal cause feature and one missing the focal effect feature. The vignette text and diagram

were also presented on this screen so that the participant could reference them. The importance of the features to continuity of identity was measured by asking participants which of the two people was most likely to be the character in the vignette. Participants then reported how plausible they felt the vignette was on a scale of 0 to 100 (from “not at all plausible” to “extremely plausible”). To ensure participants had carefully made their selections, on a separate screen they again reported which person was more likely to be the character in the vignette. Finally, to promote close reading of vignettes, participants answered a simple multiple-choice question about the details of the story.

## Results

We excluded trials where participants misreported the causal relationship of the features in the comprehension check (12% of trials) or provided inconsistent answers of which person was the character in the vignette (5% of trials). However, results were similar when no trials (or participants) were excluded.

The dependent measure was the average percentage of trials where participants selected the person who was missing the effect feature. For the common cause trials, the causal self hypothesis predicts participants should pick the person missing the effect feature as the same person rather than pick the person missing the cause feature. This is what we found. Participants were significantly more likely to select the person missing the effect feature ( $M=72\%$ ,  $SD=32\%$ , 95% CI=[.64, .81],  $t(55)=5.27$ ,  $p<.001$ , replicating the result of Experiment 3. Results were similar when no trials (or subjects) were excluded ( $M = 67\%$ ,  $SD = 31\%$ , 95% CI=[.59, .75]),  $t(59) = 4.21$ ,  $p < .001$ .

For the common effect trials, however, the two approaches to causal centrality yield different predictions. The number of causal connections approach suggests a missing effect should disrupt

identity more than a missing cause, because the effect feature participates in more causal relationships than the cause features do. Based on this definition of causal centrality, the results should be the reverse of the common cause trials, with participants picking the person missing the effect *less* than the person missing the cause. In contrast, the causal depth approach implies that order in the causal chain is what matters, and a missing cause should therefore disrupt identity more than a missing effect would. This approach predicts that participants will pick the person missing the effect more than the person missing the cause, just as they did for common cause trials.

Our results are again more consistent with the number of causal connections approach. In the common effect condition, participants were *less* likely to select the person missing the effect feature—the feature that had more connections but was less deep in the causal chain ( $M=39\%$ ,  $SD=32\%$ ,  $95\% CI=[.31, .47]$ ),  $t(55)=2.60$ ,  $p=.012$ . Results were similar when no trials (or subjects) were excluded ( $M = 38\%$ ,  $SD = 28\%$ ,  $95\% CI=[.30, .45]$ ),  $t(59) = 3.33$ ,  $p = .002$ . The average percentage of missing effect selections was significantly different between the common cause and common effect conditions, ( $Ms=72\%$  vs.  $39\%$ ,  $t(55)=5.20$ ,  $p<.001$ ). Again, results were similar when no trials (or subjects) were excluded ( $Ms = 67\%$  vs.  $38\%$ ),  $t(59) = 5.18$ ,  $p < .001$ .

Because different causal structures may also differ in how natural they appear to be (Ahn, 1999), we also examined the plausibility of the two different types of causal structures. The common cause and common effect vignettes were rated as equally plausible ( $M_{commoncause}=74.2$ ,  $M_{commoneffect}=72.5$ ,  $t(11)=.47$ ,  $p>.250$ ). So, the observed difference in selections between the two conditions cannot be explained by a difference in the believability of the two causal structures. These results experimentally corroborate our correlational findings. Change in more causally

central features, as defined by the number of causal connections, is more disruptive to perceived continuity of identity.

Lastly, we looked at the effects of vignette plausibility. As in Experiment 3, we found a correlation between the plausibility of the story and the proportion of selections of the person missing the causally peripheral feature with fewer connections ( $r = .64, p < .001, 95\% \text{ CI} = [.32, .83]$ ). The average proportion of selections of the person missing the causally peripheral feature was significantly higher among the twelve most plausible vignettes ( $M = 74.2\%, SD = 15.8\%$ ) than the twelve least plausible vignettes ( $M = 55.7\%, SD = 20.0\%$ ),  $t(22) = 2.51, p = .020, 95\% \text{ CI} = [.03, .34]$ .

The correlation between plausibility judgments and the proportion of selections of the person missing the causally peripheral feature replicates the results of Experiment 3 and provides evidence that participants incorporated their own beliefs about causal structure in making identity judgments. Participants' use of the causal information from the vignette was moderated by how it fit with their beliefs about the plausibility of this information.