

HOW PERSONAL THEORIES OF THE SELF
SHAPE BELIEFS ABOUT IDENTITY CONTINUITY

Sarah Molouki, Stephanie Y. Chen, Oleg Urminsky, & Daniel M. Bartels
University of Chicago

Abstract

Although previous research has suggested that one's personal identity is maintained via the stability of certain essential features (such as personality traits, memories, or moral qualities), the current framework proposes that people hold a dynamic representation of their identity that includes theories about their own development over time. In this chapter, we discuss two types of theories people hold about their identity, and we suggest that changes that are inconsistent with these theories are perceived as disruptive to identity. First, people have causal theories about how features of identity are interrelated—beliefs about which features were caused by or the cause of other features. Changes in more causally central features—those involved in a greater number of cause-effect relationships—are seen as more disruptive to the overall causal theory and thus, are more disruptive to continuity of identity. Second, people also hold expectations and desires about how their personal qualities will change in the future. Because people generally expect improvements, positive changes are seen as identity-consistent, whereas negative changes (especially to central features) are seen as disruptive to identity. Overall, we conclude that since the self-concept does not seem to be defined by a static list of personal features, identity continuity is most preserved when changes in features are consistent with (rather than conflicting with) one's own causal and developmental theories of the self.

Introduction

Work on transformative experiences, including some work in the current volume, has tended to focus on normative questions. For example, given the difficulty of grasping the subjective value and experience of living as the transformed individual prior to experiencing a transformation, how can a person make rational decisions related to such experiences? In this chapter, we take a step back and explore people's beliefs about what constitutes a transformative experience. Theoretical characterizations of what generally constitutes transformative experience have been proposed (“...you can't know what it is going to be like to be you after the experience.

It...changes your core preferences about what matters”; Paul, 2014, p. 17), as well as several canonical illustrative examples (e.g., having a child, becoming a vampire, undergoing a religious conversion, seeing color for the first time; Paul, 2014). As a complement to this theoretical question, we aim to explore the corresponding empirical question of what types of changes people *believe* will transform them into a different individual.

Throughout life, each of us goes through myriad personal changes. However, we do not believe that all personal changes will transform us into a different individual to the same extent. For example, imagine that Jane is a student with a lifelong passion of becoming a musician. She will likely undergo significant improvements in knowledge and experience while going through music school, however, she would generally view this personal change to be consistent with, rather than disruptive of, who she is as an individual. In contrast, she would likely consider a sudden loss of her musical abilities due to brain damage to be quite disruptive to who she is as an individual. Why does a person expect some personal changes to be transformative and not others, and what determines this distinction?

When considering people’s subjective beliefs about their own personal transformation, it may be useful to think in terms of the perceived *degree* of transformation (e.g., the degree of disruption to self-continuity), consistent with Parfit’s idea of varying degrees of psychological connectedness between present and future selves (Parfit, 1971; 1984). Given the frequency of personal change that people experience throughout their lives (both positive and negative, as well as expected and unexpected), people may rarely feel as if their self-concept is completely static. By the same token, even dramatic transformative change seldom erases all traces of the pre-existing self. Thus, people’s subjective beliefs about disruption to self-continuity largely occur on a continuum.

In this chapter, we propose that people’s self-concepts, the set of beliefs about what makes them who they are as individuals, include theories that guide which changes and experiences they believe will be more or less disruptive to who they are. We discuss two different classes of theories that people’s self-concepts include. First, people have theories about how aspects of the self causally relate to each other, and thereby which personal aspects hold a causally central role. Second, people have theories about the ways in which their specific personal qualities will change in the future—how they will develop into the person they expect to become. To illustrate this distinction, Jane would have distinct theories about (1) how other

aspects of her identity are causes or consequences of her involvement with music and (2) how she expects her musical ability to change in the future. We propose that personal changes and experiences that are inconsistent with either of these two types of theories are perceived as disruptive to self-continuity, in that people believe undergoing these changes will transform them, to some degree, into a different person.

Regardless of the normative question of whether people can think rationally about potentially transformative experiences (Paul, 2014; 2015), the fact remains that people must and do make decisions about such experiences. Furthermore, recent evidence suggests that understanding the extent to which a person views a particular experience as transformative can be useful for understanding what decisions that person will make. For example, the belief that one will be a different person in the future has an important influence on intertemporal choices. Across various types of decisions, people are more likely to favor their short-term needs over long-term needs when they feel that their future self will be a significantly different person than who they are now (see Urminsky, 2017 for a review). Thus, understanding what changes people believe will be transformative may be broadly useful, as these beliefs can affect not only decisions related to the change itself (e.g., whether or not to undergo it), but also broader decisions related to temporal tradeoffs.

1. The Self-Concept and Personal Change

People's self-concept, their set of beliefs about what makes them who they are as individuals, plays a central role in guiding cognition and behaviors in both personal and social domains. In personal decision-making, perceptions of one's own abilities shape goal-setting and motivation (Markus & Nurius, 1986; Markus & Wurf, 1987), salient aspects of the self can influence current preferences and consumption choices (LeBoeuf, Shafir, & Bayuk, 2010; Reed, 2004; Reed, Forehand, Puntoni, & Warlop, 2012), and a sense of self-continuity over time can provide the impetus for choosing behaviors with positive long-term consequences when they conflict with short-term desires (Bartels & Rips, 2010; Bartels & Urminsky, 2011; 2015). In social contexts, the self-concept can serve as a starting point for evaluations of others (Markus & Wurf, 1987; Ross, Greene, & House, 1977) and guide norms and strategies of interpersonal interaction (Akerlof & Kranton, 2000; Markus & Wurf, 1987; Swann, 1983).

A central question in the study of beliefs about personal transformation is what people believe defines their self-concept. That is, what features, if changed, does an individual believe

would make her into a different person (disrupt her self-continuity)? While the focus of this chapter and our empirical research is the self-concept and self-continuity, much previous work on these topics has examined beliefs about other people with the implicit assumption that people think about their self-concept and the concept of other individuals in a similar way. As such, we review literature on disruption to an individual's continuity that both examines how people think about themselves (self-continuity) and how they think about another person (individual-continuity). However, as most previous research on beliefs about individual-continuity has not directly compared beliefs about self-continuity and the continuity of another person, whether and how the self-concept differs from our concepts of other people are important questions for future research.

Many approaches to studying self- and individual-continuity assume that there is a type of feature that tends to consistently be most defining of the concepts of the self and of other people. For example, some research suggests that mental (rather than physical) changes are the most disruptive to perceptions of the continuity of a specific individual (Nichols & Bruno, 2010). Different researchers have proposed different types of features as particularly defining of the self, including autobiographical memories (Blok, Newman, & Rips, 2005; Nichols & Bruno, 2010), moral characteristics (Heiphetz, Strohminger, & Young, 2017; Strohminger & Nichols, 2014; 2015), social categories (Akerlof & Kranton, 2010; Hogg, Terry, & White, 1995) and personality traits and preferences (Haslam, Bastian, & Bissett, 2004; Gelman, Heyman, & Legare, 2007).

According to such feature-based approaches, the class of features that is most defining of an individual must remain stable to preserve perceived continuity of that individual. For example, Strohminger and Nichols (2014) suggest that the persistence of one's moral characteristics is most relevant to maintaining individual-continuity, whereas changes in personal characteristics unrelated to morality will cause less perceived disruption.

An alternative view (Parfit, 1984) defines self-continuity in terms of the similarity or degree of overlap in the total set of a person's psychological properties across a particular time period, rather than the degree of change in a particular type of feature. From this perspective, the greater the magnitude of change (i.e., the greater the overall change in one's total set of psychological features), the greater the disruption to self-continuity. In this view, because there is no core "self" that exists beyond a particular set of psychological features, a larger magnitude of

change to these features will disrupt the total constellation of psychological characteristics that defined one's self-concept. Thus, changes to different features may cause different levels of disruption among different individuals based on the degree to which a given feature initially served the purpose of distinguishing that individual from other people.

However, there is reason to believe that the type of feature changed and the total magnitude of the change are not the only determinants of what makes an experience disruptive to an individual's continuity. For example, one body of research has found that the valence of change affects perceived continuity of an individual into the future, with negative changes seen as more disruptive to continuity than positive ones (Newman, Bloom, & Knobe, 2014; Newman, De Freitas, & Knobe, 2015; Tobia, 2015). If a person goes from being extremely cruel to extremely kind, that is viewed as less disruptive to who she is than the same change in the opposite direction (extremely kind to extremely cruel; Tobia, 2015). Indeed, people tend to endorse the idea that an individual whose personality and behavior changed from cruel to kind has revealed an aspect of the "true self." In other words, the positive qualities had always formed a fundamental, potentially hidden, part of her nature (Newman et al., 2014; Tobia, 2015) and thus, this type of change does not pose any disruption to who she truly is. These findings challenge the idea that feature type or magnitude of change are the only factors that determine an individual's continuity, since, in this work, feature type and magnitude of change are held constant while only the direction of change is varied.

In the current chapter, our goal is to examine how people's intuitive theories included in their self-concepts shape their beliefs about what changes will disrupt their self-continuity and transform them into a different person. We discuss recent research on how these intuitive theories guide judgments of change and transformation. Rather than focusing on the absolute magnitude of change or the immutability of specific features, we examine more complex beliefs that are a part of a person's self-concept, relating to past development, causal relationships between personal features, and expectations about the future. Since these theories vary across individuals, we propose people will vary in their perceptions of how disruptive to self-continuity a change will be. That is, a change of the same magnitude to the same feature may be perceived as disruptive to self-continuity by some people and perceived as not disruptive by others.

2. Intuitive Theories Embedded in the Self-Concept Underlie Anticipated Disruption

Lay or intuitive theories are an important part of cognition, influencing knowledge organization, inference, social interactions, and learning. Some researchers suggest that much of cognitive development can be characterized as revision of intuitive theories (Carey, 1985; Gopnik & Wellman 1994; 2012; Wellman & Gelman, 1992). A classic example of theory revision is the development of theory of mind, during which children go from holding a theory that human behavior is driven by the true state of the world to a theory that suggests that behavior is driven by unique internal mental states like desires and beliefs. Thus, while 3-year-olds generally believe that a person will look for a hidden object where it actually is, 5-year-olds have revised these theories and generally believe that a person will look for a hidden object where she thinks it is, regardless of whether the object is actually there (Gopnik & Wellman, 1994; but see Baillargeon, Scott, & He, 2010 and Rhodes & Brandone, 2014 for different perspectives). While there is variation on what is meant by an intuitive theory and its level of specificity, in general, intuitive theories refer to our everyday understanding and explanatory beliefs about the world that often contain beliefs about causal relationships (Carey, 1985; Murphy & Medin, 1985).

Much of the work to date on intuitive theories has been conducted in the domain of conceptual knowledge and has examined what types of changes people perceive as disrupting an item's category membership. For example, based on their intuitive theories about immunology, people encountering someone with a new illness they had never heard of before may think that changes to the patients' symptoms (fever and chills) are less likely to change her diagnosis (categorization) than changes to the virus.

In this chapter, we explore the self-concept by building on two types of theories identified in the conceptual knowledge literature. First, people have theories about how the features of a concept are causally related to each other, which may include the course of development of those features in the past. These theories allow people to pick out the features that are important to a concept. Features that participate in many cause-effect relationships are generally perceived as more defining of a concept, and changes in these features are therefore seen as likely to disrupt category membership (Ahn, Kim, Lassaline, & Dennis, 2000; Rehder & Hastie, 2001). Second, people have domain-specific expectations about how things will develop or change in the future (e.g., see Blok et al., 2005; Rips, 2011; Rips, Block, & Newman, 2006),

and changes that run counter to these expectations are also perceived as disruptive to continuity of the object's category membership.

Whereas these theories have been extremely influential in the study of biological, artifact, and artificial categories, they have only recently been examined in the domain of the self-concept and self-continuity (see Chen, Urminsky, & Bartels, 2016 and Molouki & Bartels, 2017).

Building on this recent research, we suggest that people have theories (similar to those they hold about categories) embedded in their self-concepts that guide their judgments about the degree to which changes will transform them into a different person. More specifically, people have theories about 1) the causal relationships that exist between their personal characteristics or features and, 2) the trajectory of expected or desired change in those features. Changes that are inconsistent with these intuitive theories—changes that 1) disrupt more of these causal relationships or, 2) deviate from the expected trajectory—will be seen as relatively more disruptive to self-continuity (i.e., as more transformative) than equivalent changes that are consistent with these theories.

If people's sense of what changes are transformative is rooted in their intuitive theories, then it may be more relevant to ask for whom a given change will be seen as transformative, rather than attempting to identify which specific changes are uniformly transformative. To the degree that different people have different theories about the causal relationships that exist between features of the self-concept and different beliefs about their expected trajectory of development, this approach allows us to explain why a given change may be more disruptive for some people than for others.

In the next two sections, we provide theoretical background and review experimental evidence that people's intuitive theories about causal relations and expected change guide their beliefs about what will transform them into a different person. We conclude the chapter with some implications of these beliefs for decisions and behavior, and discuss connections to theories of transformative experience discussed in other literature.

Theories About Causal Relationships Among Features of the Self-Concept

We characterize transformative changes as those that disrupt perceived self-continuity. That is, the individual believes she effectively will be transformed into a new person and, to a large degree, will no longer be an instance of the individual she originally was. Jane no longer believes that she would be Jane after a transformative change, but Jane believes that she would

still (largely) be Jane after a less significant change. These judgments about self-continuity are similar to categorization judgments. For example, we think that a duck is no longer a duck when we change a defining feature (its DNA), but we're more confident it remains a duck when we change a less significant feature (its ability to swim). From this perspective, the kinds of intuitive theories that guide beliefs about what features define a category—the features that, when changed, most call into question an item's status as a category member—may also guide what changes people believe will disrupt their self-continuity, when applied to the self-concept.

As in feature- and similarity-based approaches to self-continuity (e.g., Blok, et al., 2005; Gelman, et al., 2007; Haslam et al., 2004; Nichols & Bruno, 2010; Parfit, 1984; Strohminger & Nichols, 2014; 2015), research about category membership more generally has also explored feature type and magnitude of similarity as key determinants. Early ideas of what features define a category included proposals of specific feature types—e.g., shape, in the context of artifact categories (Landau, Smith, & Jones, 1988). Additionally, many models of categorization suggest that items belong to a category to the extent that they are similar to other items in the category or to a summary representation (prototype) of the concept, with (magnitude of) similarity often defined as the number of features that overlap (Medin & Shaffer, 1978; Smith & Medin, 1981).

However, features differ in how central they are to the categorization of an item (Murphy & Medin, 1985). For example, the feature *has a fever* seems less defining of the concept of having the flu than the feature *has the influenza virus*. That is, we're more likely to think a person who does not have the virus (but has a fever and chills) does not have the flu than someone who does not have a fever (but has the virus and the chills). This belief tends to hold even though the magnitude of similarity to a prototypical flu patient is the same in both cases (i.e., both have two out of three of the classic flu features). Thus, a shortcoming of similarity-based approaches to categorization is their lack of ability to systematically determine the weight of features in categorization judgments.

Theory-based approaches address these shortcomings of similarity-based models of categorization. In this view, mental representations of concepts are not simply feature lists, but include intuitive theories about how these features are causally related to one another. These causal relations guide how defining features are of a concept (Murphy & Medin, 1985; Ahn et al., 2000). A seminal finding from this work is that a feature is seen as defining a concept to the

extent that it is *causally central* to the concept (Ahn, 1999; Ahn et al., 2000; Sloman, Love, & Ahn, 1998; Rehder & Hastie, 2001).

Although there are competing perspectives on the definition of causal centrality, in this chapter it is defined as the number of causal relationships that a feature participates in (either as cause or effect) with other features of the concept (see Chen et al., 2016, Experiment 3, for a test of different approaches to causal centrality in the context of the self-concept). For the flu example above, *has the influenza virus* is more defining of the concept of the flu than *has a fever* because the virus is causally linked to two other features (causes both the fever and chills), while *has a fever* is only causally linked to one other feature (it is caused by the virus). Several research streams in cognitive psychology have demonstrated the role of causal beliefs in conceptual centrality. In particular, changing the causal relations that a feature is involved in has been shown to change how important the feature is for determining category membership (e.g., Ahn et al., 2000; Rehder & Hastie, 2001).

In this section, we examine the proposal that people's representations of the self-concept likewise include intuitive theories about causal relationships, and that these subjective causal beliefs influence which personal features are seen as most defining of the self-concept. For example, imagine two people, Jack and Mack. Both have memories of having few friends as children, prefer solitary activities, and are very shy. Jack and Mack both believe that these three features are important to their self-concept; however, they differ in how they believe these features are causally related to one another. Jack believes that it was his memories of having few friends as a child that caused him to develop into a shy person and prefer solitary activities. In contrast, Mack instead believes that it was his shyness that caused him to have only a few friends as a child and prefer solitary activities.

Previous accounts of the self-concept that assume that a particular type of feature (e.g., memories) is most important would predict that since both Jack and Mack have the same set of features, they would view changes the same way. However, a causal centrality account of the self-concept predicts that even though the features of these two men are identical, their self-concepts will be fundamentally different because of their differing beliefs about the causal relationships between these features. That is, shyness will be more defining of Mack's self-concept than Jack's because it is causally linked to both his preferences and his memories (whereas for Jack it is only causally linked to his memories). As a result, they are likely to view

changes differently. In particular, Mack would believe that becoming outgoing (a change to his shyness, a causally central feature for him) would transform him into a different person to a greater extent than Jack would (because shyness is less causally central for him).

To test the hypothesis that changes to more causally central features are perceived as more disruptive to self-continuity, Chen et al. (2016, Experiments 1 and 2) measured people's beliefs about the causal centrality and importance of features to their self-concepts. In Experiment 1, the features were selected from five categories suggested to be important in previous research: memories, moral qualities, personality traits, preferences, and goals and desires. In Experiment 2, participants generated features important to their self-concept. Beliefs about causal relationships were measured by having participants verbally report which features they believed had shaped or influenced the other features of their self-concept. The causal centrality of each feature was then calculated as the number of direct links to other features (either as a cause or effect). Beliefs about how defining features are to the self-concept were measured by asking participants how much a change to each feature would disrupt their self-continuity (i.e., the extent to which a change in a feature would lead them to feel that they were a different person).

Consistent with a causal centrality account of self-continuity, people perceived changes to causally central features of their self-concepts to be more transformative than changes to causally peripheral features. The number of causal relationships that a feature was seen as participating in was correlated with how disruptive to self-continuity a change to that feature was predicted to be. This positive relationship between causal relationships and disruption to self-continuity was observed for the large majority of participants (>75% in both Experiments 1 and 2). These results were replicated when using an alternative method for eliciting causal centrality adapted from Sloman, Love, and Ahn (1998), the concept map task (Chen et al., 2016; Pilot Experiment, Appendix A1). In this task, participants drew these causal links in a concept map (see Figure 1).

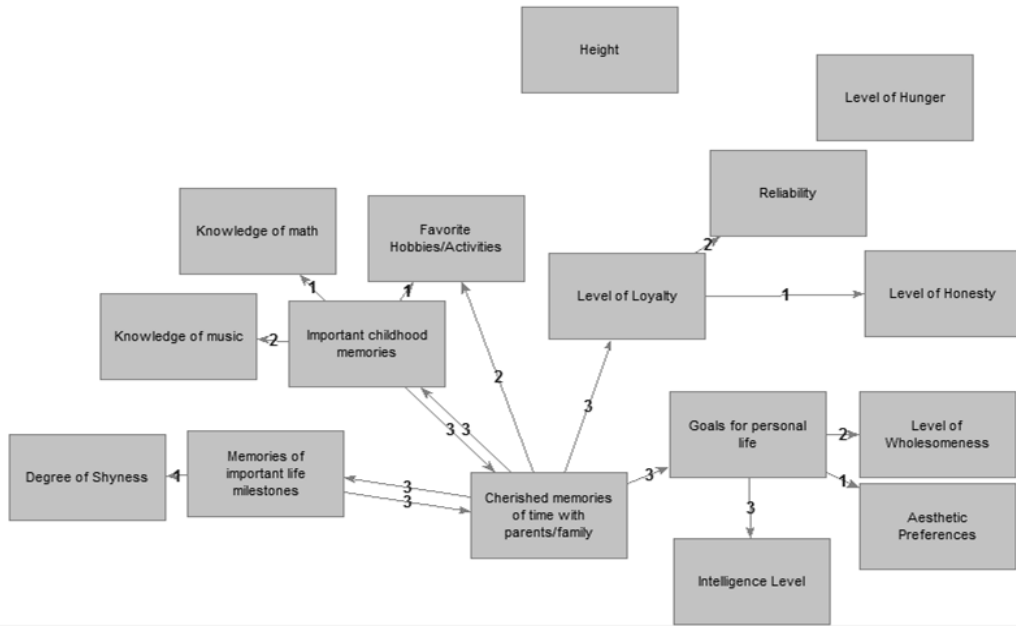


Figure 1. Example of a self-concept map. Each box contains a feature of personal identity. The arrows represent causal relationships between features. The arrow starts at the cause feature and points to the effect feature. The numbers that are on each arrow indicate the strength of the causal relationship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

Further, Chen et al. (2016) found evidence that, similar to the self-concept, people’s concepts of *other* people also include theories about the causal relationships among that person’s features, which affect judgments about the other person’s transformation and continuity. In Experiments 1 and 2 described above, one group of participants was asked for their beliefs about the causal centrality and importance of features of another person who they knew well (rather than the other person’s beliefs about their own features). Participants thought that the other individual was transformed to greater degree when the changed feature was one the participant saw as causally central (vs. causally peripheral) to the other person’s self-concept.

In follow-up experiments, Chen et al. (2016) manipulated the causal centrality of features in participant’s concepts of hypothetical people. The results replicated the previously described correlational results with an experimental manipulation. When participants believed that a feature was causally central, they perceived changes to that feature as more disruptive to a

hypothetical person's individual-continuity¹ than when they believed that the exact same feature was causally peripheral (Chen et al., 2016; Experiment 3 and Supplemental Experiment 4, Appendix A1).

The studies described above demonstrate that people's intuitive theories about how the features of their self-concept are causally interrelated influence their beliefs about which changes will transform them into a different individual. Changes to features that are perceived as causally central—participating in many cause-effect relationships with other features—are perceived as more disruptive to self-continuity than changes to features with fewer cause-effect relationships. This approach explains both why some features tend to be more defining of the self-concept than others, on average (i.e., because they are viewed as casually central for more people) and also why some features matter more to some people than to others (i.e., because people's self-concepts contain different intuitive theories about causal relationships).

Dynamic Theories About Future Trajectories

In addition to their theories about the causal relationships between personal features, people also have ideas about the future trajectories of their features. More specifically, we propose that people have expectations and desires about whether certain personal features will change positively or negatively, which guide their beliefs about self-continuity. Changes that are in line with beliefs will be seen as consistent with the self-concept, whereas changes that are contrary to these expectations and desires will be seen as disrupting self-continuity.

Our focus in this section is on beliefs about the trajectory of change, rather than on which particular feature is changing. This approach contrasts with some research (described in Section 1) that linked perceived self-continuity to the continuity of specific classes of features (e.g., Strohlinger & Nichols, 2014). Instead, we describe how people's beliefs about the direction of expected change (applicable to a wide variety of features) can affect perceptions of self-continuity.

People hold expectations about whether they will improve or decline over time. As described in Section 1, people perceive positive change to be more consistent with their theories about general human development (Newman et al., 2014; Newman et al., 2015; Tobia, 2015).

¹ Note that when assessing the continuity of another person in these particular experiments, participants provided their own opinion about whether the other individual was still the same person, and were not asked about whether the other individual herself still thought she was the same.

Furthermore, when it comes to theories of their own personal development, people are even more likely to endorse beliefs that they will improve in the future. For example, people state that their own improvement on various personality characteristics will be larger than that of their peers (Kanten & Teigen, 2008). In addition, there is recent evidence that people's predictions of their own improvement are in fact overoptimistic compared to their actual change over the same time period on various measures of personality, values, and cognitive performance (Molouki, Bartels, & Urminsky, 2017). This imbalance occurs because people tend to anticipate positive future change but neglect the possibility of negative change in the future (Molouki et al., 2017). Because people expect an improving future trajectory, it is likely that the idea of general positive change is incorporated into their self-concept. In line with this idea, we propose that people would consider many types of positive changes in themselves to be consistent with their theories about their future trajectory (regardless of the type of feature changing) and thus not perceive these changes as disruptive to self-continuity. In contrast, negative changes would be seen as inconsistent with these theories and would be perceived as relatively more disruptive to self-continuity.

The influence of valence of change in perceived personal transformation. Molouki and Bartels (2017) tested the hypothesis that perceptions of one's own self-continuity are determined by congruence with a trajectory of improvement across various types of mental features. Participants were asked to imagine that each of a list of personal characteristics (presented in Table 1) would either improve, worsen, or change (valence unspecified) in the future. For each characteristic, they provided their perception of the extent to which the specified change would disrupt their self-continuity, i.e., whether they would still be themselves after the change. If people incorporate an overall trajectory of improvement into their self-concept, then positive change would be viewed as less disruptive than negative change across all types of features.

Morality		Personality	
Morality Values Degree of Honesty Beliefs Level of Humility Religious or Spiritual Faith Level of Selflessness		Personality Sense of Humor Impulsiveness Temperament Curiosity Dominance Level of Calmness Confidence Level Imagination Reliability Degree of Independence Level of Friendliness Self-Awareness Level of Happiness Intelligence Level Level of Disorganization Helpfulness	
Preferences	Experiences	Memories	
Professional Goals Preferences/Favorite Things Major Likes and Dislikes Goals for Your Personal Life	Life Experiences Friendships Everyday Activities Health Ability to Feel Pain Occupation	Cherished Memories of Time Spent with Family Memories of Time Spent Commuting to Work Knowledge of How to Ride a Bike Knowledge of How to Play the Piano Knowledge of Math Bad Memories	

Table 1. Listing of stimuli used to measure the effects of positive and negative change across various feature types. Characteristics were selected and categorized based on pre-tests (Molouki & Bartels, 2017).

In terms of feature type, it was found that changes in items related to morality were perceived as the most disruptive to self-continuity, followed in turn by changes in items related to personality, preferences, experiences, and memories (see Figure 2). These results for people’s beliefs about their own change are generally consistent with findings demonstrating an influence

of the type of feature changing on judgments of the individual-continuity of third parties (Strohminger & Nichols, 2014).

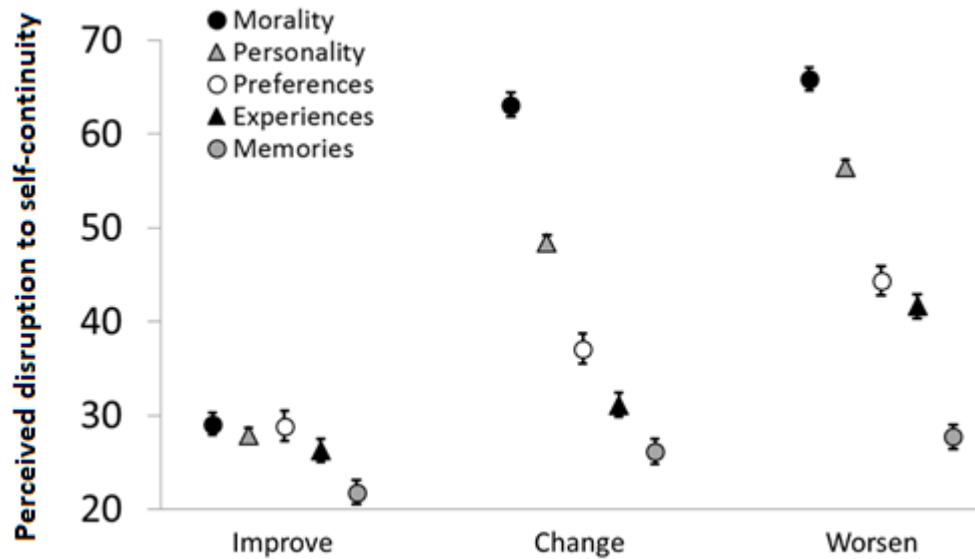


Figure 2. Perceived disruption to self-continuity based on valence and category of change. Error bars represent +/- 1 SE (Figure reproduced with permission from Molouki & Bartels, 2017).

However, importantly, the degree of perceived disruption to self-continuity also depended on the valence of the change. On average, positive change was perceived to be less disruptive than either negative change or unspecified change for all categories of characteristics, suggesting that beliefs about self-continuity are not solely guided by the type of feature changing. Thus, although changes to morality and personality generally led to the greatest perceptions of discontinuity, change that was explicitly labeled as positive was an exception--disruption to self-continuity related to improvement was quite low for all types of features (Figure 2).

These results suggest that the valence of a change influences whether that change is seen as disruptive to self-continuity. Perceiving improvement as consistent with theories about future trajectory may be common to most individuals as part of the general concept of a “person”, as this perception may stem from intuitive beliefs about positive human development over time

(Newman et al., 2014; Tobia, 2015), or be related to common aspirations for self-improvement (Kanten & Teigen, 2008). However, the impact of valence on perceived self-continuity may also depend on the person's specific expectations about her individual trajectory of personal change—i.e., how she expects her own particular characteristics to change in the future, as defined by her unique view of her own self-concept. Beliefs about one's own self-concept may be a more richly elaborated version of a general mental schema of how an individual person develops over time – however, the specific relations between these two types of mental representations remain to be further examined in future research.

The effect of individual expectations about specific changes on perceived personal transformation. Research on how people categorize objects suggests that beliefs about continuity depend on expectations of change, which vary across different types of objects (Blok et al., 2005; Gutheil, Bloom, Valderrama, & Freedman, 2004; Gutheil, Gelman, Klein, Michos, & Kelaita, 2008; Gutheil & Rosengren, 1996; Rips, 2011; Rips, et al., 2006). For example, people believe that the process of getting smaller is consistent with the identification of an object as an iceberg (see Rips et al., 2006), whereas the process of growing taller would be consistent with the identification of a human child. Observing, for example, an iceberg getting larger over time or a child getting shorter over time would be unexpected, and might suggest that the object does not fit into the proposed category or that a disruption to its continuity as the original object type has occurred (see Blok, Newman, & Rips, 2007; Nozick, 1981; Rips, 2011; Rips & Hespos, 2015; Rips et al., 2006; Sagi & Rips, 2014). Much as a person might use specific beliefs about the natural course of development of a certain type of object to assess whether it has maintained its category membership over time, she might also contrast a personal change with her specific beliefs about her own natural development to assess whether she has maintained self-continuity over time.

Intuitive theories of positive personal development could help explain why positive change in general tends not to lead to perceived disruption of self-continuity, since positive change tends to be consistent with most people's expectations and desires (e.g., Bench, Schlegel, Davis, & Vess, 2015; Busseri, Choma, & Sadava, 2009; Haslam, Bastian, Fox, & Whelan, 2007; Molouki et al., 2017; Newby-Clark & Ross, 2003; Wilson & Ross, 2001). Whereas people may have already incorporated the idea of positive change into their self-concept, negative change is more likely to be seen as inconsistent with both expectations and desires for their personal

development over time, and therefore more disruptive. However, specific expectations about a given characteristic (e.g., reasons to expect that it may change negatively) might also guide the perceived effects of change in this characteristic on self-continuity.

To test this idea, Molouki and Bartels (2017) orthogonally manipulated both valence of change and expectations of change to isolate the influence of these factors. Participants selected features for which they held strong expectations about change, and then were asked to imagine that each of these features in turn would improve, worsen, or stay the same. If individual expectations play a role in judgments related to the self-concept, any type of change, regardless of valence, that is in alignment with a person's expectations for her own future will result in relatively lesser disruption to perceived self-continuity than a change that is misaligned with expectations.

Molouki and Bartels (2017) found that people's individual expectations indeed influenced their perceptions of self-continuity. Although perceived disruption to self-continuity were generally influenced by the overall valence of change (with improvements seen as less disruptive), changes that were consistent with a participant's specific expectations about future change also led to relatively lower perceived disruption to self-continuity (see circled means in Figure 3). In contrast, change in any characteristic that ran counter to one's theory of expected change was associated with increased reports of disruption. In other words, decline was viewed as relatively more disruptive to self-continuity when a person specifically expected improvement in that characteristic, whereas improvement was seen as relatively more disruptive when a person specifically expected decline in that characteristic.

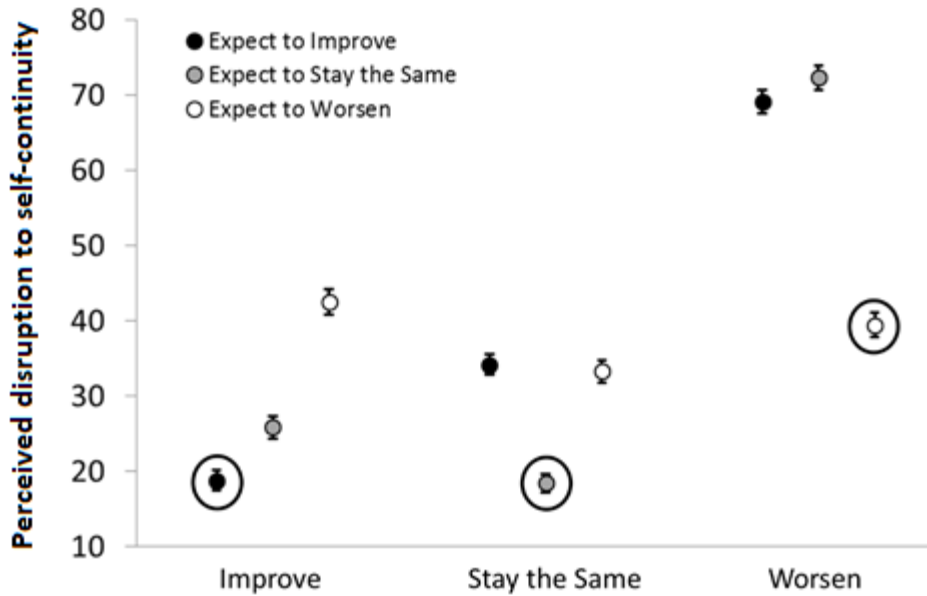


Figure 3. Perceived disruption to self-continuity based on imagined change and expected change. Note: Circled values are those for which imagined change matches expected change. Error bars represent +/- 1 SE. (Figure reproduced with permission from Molouki & Bartels, 2017).

The effect of individual desires for specific changes on perceived personal transformation. People tend to expect improvements for their future change, but it is not clear whether these ideas about future improvement stem from specific personal desires, or are instead guided by broader lay theories about how people, in general, change over time (e.g., Newman et al., 2014). This is an important distinction because if people evaluate whether a change is disruptive to their self-continuity by consulting their own individual desires, it suggests that the self-concept contains a specific aspirational component that will vary from individual to individual based on their personal goals.

Molouki and Bartels (2017) found that congruence of a change with people's specific desires matters for self-continuity judgments. Changes in a given feature that were desired (see circled means in Figure 4) led to less perceived disruption to self-continuity than undesired changes. Furthermore, the effect of desires on continuity judgments was distinct from the main effect of valence. In other words, people did report less perceived disruption to self-continuity on average in response to improvements than in response to worsening in any characteristic (an overall effect of valence), but a given participant was also more likely to report feeling that her

self-continuity was disrupted when her specific desires for a characteristic were not met (an effect of individual desires).

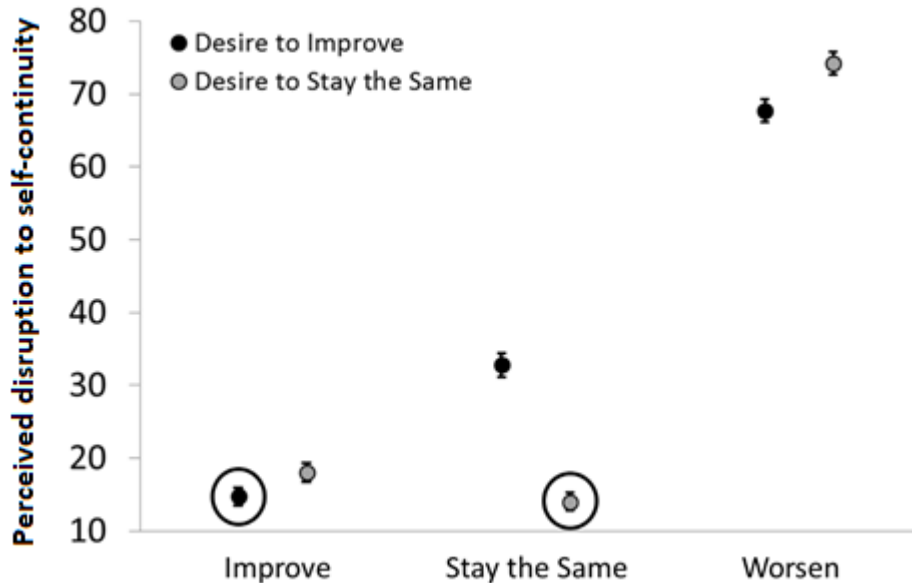


Figure 4. Perceived disruption to self-continuity based on type of change and desire for change. Features where worsening was desired were not identified in this experiment, as it was expected that most people would not hold such desires. Note: Circled values are those where imagined change matches desired change. Error bars represent +/- 1 SE (Figure reproduced with permission from Molouki & Bartels, 2017).

The results of this study suggest that individual desires play a significant role in perceptions of self-continuity, above and beyond the general effect of valence of change. For example, rather than uniformly reporting that improvements would be the most consistent with maintaining self-continuity, people actually reported somewhat greater *discontinuity* when a trait they desired to stay the same improved, compared to when it stayed the same. Overall, the effect of specific desires on self-continuity judgments suggests that people incorporate their idiosyncratic personal desires and goals into theories of development within their self-concept.

3. Conclusions

In this chapter, we have summarized recent empirical findings which suggest that people's intuitive theories determine whether a change is more likely to be perceived as transforming them into a different individual. Based on these findings, the most transformative changes are likely to be those which are unexpected, unwelcome, and involve aspects of the self-concept that are most causally linked to other features. However, these are not mutually necessary conditions. For example, a positive change can be seen as transformative (i.e., as causing a discontinuity in the self) when the change is highly unexpected, when a negative change was expected, or if the change applies to a highly causally central aspect of the self-concept. Our framework also predicts that unexpected and unwelcome changes will be less disruptive when they involve less causally central features (vs. more causally central features). However, specific predictions regarding the interaction of different intuitive theories (e.g., determining the effect of a change by simultaneously considering theories of development and theories of causal centrality) have not yet been empirically tested.

The concept of perceived disruption to self-continuity that we have introduced in this chapter provides a new way of understanding how people think about transformative experiences. Although the phrase "transformative experience" has previously been discussed in the context of specific life events (e.g., having a child, experiencing a religious conversion, or even becoming a vampire), the defining nature of such experiences is that they involve a change, often unexpected a priori, to fundamental preferences, beliefs, or ways of seeing the world (Paul, 2014). The experiments we discussed explored how people's intuitive theories about themselves relate to how they perceive these types of changes (e.g., to preferences, beliefs, and other personal characteristics). Regardless of the nature of the changes that ultimately occur, a variety of life experiences can be *subjectively perceived* as transformative to varying degrees, depending on people's theories about personal features that are changing.

In many cases, these individual perceptions may differ from normative definitions of what would constitute a transformative experience. For example, some descriptions classify the birth of a child as a transformative change, since it brings about unanticipated developments to one's fundamental ways of thinking and functioning as a person that could not be fathomed prior to the change (Paul, 2014; 2015). However, individuals themselves will likely vary on the degree to which they personally view having a child as a transformative experience. Though some individuals believe that the birth of a child will transform who they are as a person, others might

not anticipate that this event would disrupt who they are because it is viewed as a) consistent with the causal theories embodied in their self-concepts, b) a positive change, and/or c) an expected or desired change. If a person sees having a biological child in the future as desirable, expected, and causally linked to many other aspects of the self-concept, then it may in fact be the *failure* to conceive a child that would be assessed as far more disruptive to self-continuity, regardless of the degree of change to fundamental preferences that in fact occurs when becoming a parent.

We further note that one central aspect of transformative change, as defined in previous work, is that due to changes in fundamental preferences and personal characteristics, one is unable to understand the phenomenal quality of the experience or outcome (Paul, 2014; 2015). We believe that many of the changes people reported as disrupting self-continuity in our experiments may indeed be characterized by the fact that the resulting outcome could not be fully understood prior to the change. However, we did not explicitly assess a) the degree to which the resulting experience of the transformed self could or could not be understood by the pre-transformation individual or b) whether people anticipated that they would be unable to predict the experience of the transformed self after any given change. Thus, lay beliefs about the specific aspect of transformative experience that relates to unknowable outcomes remain an interesting area for future study.

Theories of Personal Transformation and Decision Making

The work reviewed in this chapter suggests that intuitive theories included in the self-concept, which specific to an individual, influence the degree to which a change is seen as transformative. Variations across individuals' intuitive theories influence the degree to which an anticipated change makes the person feel that they would be transformed into a different person. Such beliefs may be an important consideration in decision making in several different ways.

As one example, people may choose to behave in ways that are shaped by anticipation of possible disruption to self-continuity and a resultant desire to avoid this outcome. If people are generally motivated to preserve their self-continuity, they may be more likely to make choices that are consistent with the causally central features of their self-concept (e.g., using green products in order to be consistent with being an environmentalist; Chen & Urminsky, 2017). As another illustration of behavior aimed at preserving self-continuity, people who perceive that a political social category (e.g., Democrat or Republican) is causally central in their self-concept

are more likely to act in ways that are consistent with that feature (e.g., vote for the candidate nominated by their political party) than those who believe that the same social category is causally peripheral (Chen & Urminsky, 2018). Understanding a person's intuitive theories about the causal relationships among features of their self-concept may thus provide unique insight into likely future behaviors that are motivated by the desire to preserve certain features. However, in other cases, people may not be motivated to maintain self-continuity. For example, when people anticipate future negative outcomes, they may instead be motivated to distance themselves from these by acting in ways that disrupt their self-continuity, or at least signal a change in the self (Yang & Urminsky, 2015).

When change is inevitable, beliefs about transformation may also shape how people cope with that change. For example, Parfit (1984) points out that if one perceives the future self to be only weakly connected to the current self, this may reduce one's current concern about the fate of the future self. Given that a transformative experience will tend to weaken the perceived connection between current and future self, the very act of perceiving an experience as transforming the self could sometimes actually mitigate concern and anxiety about the future outcomes of that change. Nichols, Strohinger, Rai, and Garfield (2018) discuss a similar idea in a study of religious beliefs and thoughts about the self. Among a sample of participants of various religions, Tibetan Buddhists, who endorse the idea that there is no such thing as a continuous, enduring self, specifically reported using this belief to cope with thoughts of death (perhaps the most extreme transformative experience possible). However, these Buddhist participants nevertheless reported high levels of anxiety about self-annihilation at death, suggesting that feelings or intuitions about self-continuity may persist despite explicit religious beliefs to the contrary. An interesting topic for future research is the relationship between beliefs about personal transformation and strategies for coping with a range of unexpected or undesired changes of a less extreme nature than death.

Beliefs about transformation and disruption may have broad implications for understanding people's motivation for future-oriented decision making more generally (Bartels & Rips, 2010). When changes on the horizon are seen as transformative, in the sense of reducing people's subjective sense of similarity to their future self, they are likely to feel less motivated to sacrifice in the present for the benefit of that future self. This can have important consequences for decision-making, leading to undesired outcomes such as greater financial impatience (Bartels

& Urminsky, 2011; Ersner-Hershfield, Garton, Ballard, Samanez-Larkin, & Knutson, 2009), reduced motivation (Peetz, Wilson, & Strahan, 2009; Dai, Milkman, & Riis, 2015), more willingness to spend discretionary funds even when considering future consequences (Bartels & Urminsky, 2015), and the commission of unethical acts (Hershfield, Cohen, & Thompson, 2012). On the other hand, disruption to self-continuity and the resulting reduced concern about the future self can also sometimes prompt positive behaviors such as heightened generosity to future others (Bartels, Kvaran, & Nichols 2013).

Lastly, beliefs about change may shape how people make decisions about the change itself. An individual who does not consider childbirth to be a transformative experience may be confident (or even overconfident) in her abilities to make decisions about this event, believing that she will remain mostly the same person with the same preferences, beliefs, desires, etc. after having a child. In contrast, someone for whom becoming a parent is less consistent with the theories embedded in her self-concept might not be as confident in making decisions about how to parent. Because she is more likely to think that the experience will transform her into a different individual, she may feel that she does not fully grasp how to evaluate and respond to the outcomes of this change, and may therefore choose to defer key decisions.

Evaluating a given individual's perceptions about what constitutes a personally transformative experience is an important supplement to research on normative theories of what constitutes such experiences, because people's tendencies to employ decision-making strategies will vary based on their personal beliefs about transformation and continuity. Regardless of whether or not a given individual's theory of what defines a transformative experience agrees with or diverges from reality, it is ultimately the individual's own evaluation that will guide her feelings and behaviors surrounding the experience.

References

- Ahn, W. K. (1999). Effect of causal structure on category construction. *Memory & Cognition*, 27(6), 1008-1023.
- Ahn, W. K., Kim, N. S., Lassaline, M. E., & Dennis, M. J. (2000). Causal status as a determinant of feature centrality. *Cognitive Psychology*, 41(4), 361-416.
- Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, 115(3), 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.
- Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, 14(3), 110-118.
- Bartels, D. M., Kvaran, T., & Nichols, S. (2013). Selfless giving. *Cognition*, 129(2), 392-403.
- Bartels, D. M., & Rips, L. J. (2010). Psychological connectedness and intertemporal choice. *Journal of Experimental Psychology: General*, 139(1), 49-69.
- Bartels, D. M., & Urminsky, O. (2011). On intertemporal selfishness: How the perceived instability of identity underlies impatient consumption. *Journal of Consumer Research*, 38(1), 182-198.
- Bartels, D. M., & Urminsky, O. (2015). To know and to care: How awareness and valuation of the future jointly shape consumer spending. *Journal of Consumer Research*, 41(6), 1469-1485.
- Bench, S. W., Schlegel, R. J., Davis, W. E., & Vess, M. (2015). Thinking about change in the self and others: The role of self-discovery metaphors and the true self. *Social Cognition*, 33(3), 169-185.
- Blok, S., Newman, G., & Rips, L. J. (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 lab* (pp. 127-149). Washington, D.C.: American Psychological Association.
- Blok, S. V., Newman, G. E., & Rips, L. J. (2007). Out of sorts? Some remedies for theories of object concepts: A reply to Rhemtulla and Xu (2007). *Psychological Review*, 114(4), 1096-1102.
- Busseri, M. A., Choma, B. L., & Sadava, S. W. (2009). Functional or fantasy? Examining the implications of subjective temporal perspective “trajectories” for life satisfaction. *Personality and Social Psychology Bulletin*, 35(3), 295-308.

- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: Bradford Books, MIT Press.
- Chen, S. Y., & Urminsky, O. (2018). The role of causal beliefs in political identity and voting. Working paper.
- Chen, S. Y., & Urminsky, O. (2017). We are what we think: Representations of the self-concept and identity-based choice. Working paper.
- Chen, S. Y., Urminsky, O., & Bartels, D. M. (2016). Beliefs about the causal structure of the self-concept determine which changes disrupt personal identity. *Psychological Science*, 27(10), 1398-1406.
- Dai, H., Milkman, K. L., & Riis, J. (2015). Put your imperfections behind you: Temporal landmarks spur goal initiation when they signal new beginnings. *Psychological Science*, 26(12), 1927-1936.
- Ersner-Hersfield, H., Garton, M. T., Ballard, K., Samanez-Larkin, G. R., & Knutson, B. (2009). Don't stop thinking about tomorrow: Individual differences in future self-continuity account for saving. *Judgment and Decision Making*, 4(4), 280.
- Gelman, S. A., Heyman, G. D., & Legare, C. H. (2007). Developmental changes in the coherence of essentialist beliefs about psychological characteristics. *Child Development*, 78(3), 757-774.
- Gopnik, A., & Wellman, H. M. (1994). The "theory theory". In L. Hirschfield and S. Gelman (Eds.) *Domain specificity in culture and cognition* (pp. 257–293). New York: Cambridge University Press.
- Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, 138, 1085–1108.
- Gutheil, G., Bloom, P., Valderrama, N., & Freedman, R. (2004). The role of historical intuitions in children's and adults' naming of artifacts. *Cognition*, 91(1), 23-42.
- Gutheil, G., Gelman, S. A., Klein, E., Michos, K., & Kelaita, K. (2008). Preschoolers' use of spatiotemporal history, appearance, and proper name in determining individual identity. *Cognition*, 107(1), 366-380.
- Gutheil, G., & Rosengren, K. S. (1996). A rose by any other name: Preschoolers' understanding of individual identity across name and appearance changes. *British Journal of Developmental Psychology*, 14(4), 477-498.

- Haslam, N., Bastian, B., & Bissett, M. (2004). Essentialist beliefs about personality and their implications. *Personality and Social Psychology Bulletin*, 30(12), 1661-1673.
- Haslam, N., Bastian, B., Fox, C., & Whelan, J. (2007). Beliefs about personality change and continuity. *Personality and Individual Differences*, 42(8), 1621-1631.
- Heiphetz, L., Strohminger, N., & Young, L. L. (2017). The role of moral beliefs, memories, and preferences in representations of identity. *Cognitive Science*, 41(3), 744-767.
- Hershfield, H. E., Cohen, T. R., & Thompson, L. (2012). Short horizons and tempting situations: Lack of continuity to our future selves leads to unethical decision making and behavior. *Organizational Behavior and Human Decision Processes*, 117(2), 298-310.
- Hogg, M. A., Terry, D. J., & White, K. M. (1995). A tale of two theories: A critical comparison of identity theory with social identity theory. *Social Psychology Quarterly*, 58(4), 255-269.
- Kanten, A. B., & Teigen, K. H. (2008). Better than average and better with time: Relative evaluations of self and others in the past, present, and future. *European Journal of Social Psychology*, 38(2), 343-353.
- Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, 3(3), 299-321.
- LeBoeuf, R. A., Shafir, E., & Bayuk, J. B. (2010). The conflicting choices of alternating selves. *Organizational Behavior and Human Decision Processes*, 111(1), 48-61.
- Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist*, 41(9), 954-969.
- Markus, H., & Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology*, 38(1), 299-337.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85(3), 207-238.
- Molouki, S., & Bartels, D. M. (2017). Personal change and the continuity of the self. *Cognitive Psychology*, 93, 1-17.
- Molouki, S., Bartels, D. M., & Urminsky, O. (2017). Neglecting decline: Remembered and predicted personal development diverge from actual longitudinal change. University of Chicago Working Paper.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92(3), 289-316.

- Newby-Clark, I. R., & Ross, M. (2003). Conceiving the past and future. *Personality and Social Psychology Bulletin*, 29(7), 807-818.
- Newman, G. E., Bloom, P., & Knobe, J. (2014). Value judgments and the true self. *Personality and Social Psychology Bulletin*, 40(2), 203-216.
- Newman, G. E., De Freitas, J., & Knobe, J. (2015). Beliefs about the true self explain asymmetries based on moral judgment. *Cognitive Science*, 39(1), 96-125.
- Nichols, S., & Bruno, M. (2010). Intuitions about personal identity: An empirical study. *Philosophical Psychology*, 23(3), 293-312.
- Nichols, S., Strohminger, N., Rai, A., & Garfield J. (2018). Death and the self. *Cognitive Science*, 42(51), 314-332.
- Nozick, R. (1981). *Philosophical explanations*. Cambridge, MA: Harvard University Press.
- Parfit, D. (1971). Personal identity. *The Philosophical Review*, 80(1), 3-27.
- Parfit, D. (1984). *Reasons and persons*. Oxford University Press.
- Paul, L. A. (2014). *Transformative experience*. Oxford University Press.
- Paul, L. A. (2015). What you can't expect when you're expecting. *Res Philosophica*, 92(2), 149-170.
- Peetz, J., Wilson, A. E., & Strahan, E. J. (2009). So far away: The role of subjective temporal distance to future goals in motivation and behavior. *Social Cognition*, 27(4), 475-495.
- Reed, A. (2004). Activating the self-importance of consumer selves: Exploring identity salience effects on judgments. *Journal of Consumer Research*, 31(2), 286-295.
- Reed, A., Forehand, M. R., Puntoni, S., & Warlop, L. (2012). Identity-based consumer behavior. *International Journal of Research in Marketing*, 29(4), 310-321.
- 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(3), 323-360.
- Rhodes, M., & Brandone, A. C. (2014). Three-year-olds' theories of mind in actions and words. *Frontiers in Psychology*, 5, 263.
- Rips, L. J. (2011). Split identity: Intransitive judgments of the identity of objects. *Cognition*, 119(3), 356-373.
- Rips, L. J., Blok, S., & Newman, G. (2006). Tracing the identity of objects. *Psychological Review*, 113(1), 1-30.

- Rips, L. J., & Hespos, S. J. (2015). Divisions of the physical world: Concepts of objects and substances. *Psychological Bulletin*, *141*(4), 786-811.
- Ross, L., Greene, D., & House, P. (1977). The “false consensus effect”: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, *13*(3), 279-301.
- Sagi, E., & Rips, L. J. (2014). Identity, causality, and pronoun ambiguity. *Topics in Cognitive Science*, *6*(4), 663-680.
- Sloman, S. A., Love, B. C., & Ahn, W. K. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, *22*(2), 189-228.
- Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.
- 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*(9), 1469-1479.
- Swann, W. B. (1983). Self-verification: Bringing social reality into harmony with the self. *Social Psychological Perspectives on the Self*, *2*, 33-66.
- Tobia, K. P. (2015). Personal identity and the Phineas Gage effect. *Analysis*, *75*(3), 396-405.
- Urminsky, O. (2017). The Role of Psychological Connectedness to the Future Self in Decisions Over Time. *Current Directions in Psychological Science*, *26*(1), 34-39.
- Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology*, *43*(1), 337-375.
- Wilson, A. E., & Ross, M. (2001). From chump to champ: People's appraisals of their earlier and present selves. *Journal of Personality and Social Psychology*, *80*(4), 572-584.
- Yang, A. X., & Urminsky, O. (2015). The foresight effect: Local optimism motivates consistency and local pessimism motivates variety. *Journal of Consumer Research*, *42*(3), 361-377.