Similarity chains in the transformational paradigm

Emmanuel M. Pothos

Department of Psychology, Swansea University.

Ulrike Hahn

School of Psychology, Cardiff University.

Mercè Prat-Sala,

Department of Psychology, The University of Winchester.

in press, European Journal of Cognitive Psychology

Running Head: the transformational paradigm; **Word Count**: 5965

Please address correspondence regarding this article to Emmanuel Pothos, Department of Psychology, Swansea University, Swansea SA2 8PP, UK. Electronic mail may be sent to e.m.pothos@swansea.ac.uk.

Rips's (1989) results with the transformational paradigm have often been cited as supporting accounts of categorization not based on similarity, such as involving necessary or sufficient features (or a belief in such features), which guarantee a categorization outcome once their presence has been established. We discuss a similarity account of the transformational paradigm based on similarity chains, which predicts that when the transformation is more gradual the identity of the transformed object is less likely to change. Conversely, we suggest that an essentialist approach to categorization predicts that essences are more likely to change in gradual transformations, across generations, as is the case with evolutionary change of species. In two experiments we examined the scope of the similarity vs. the essentialist account in the transformational paradigm. With space aliens, the similarity account was superior to the essentialist one, but the converse was true with earth creatures. We suggest that an essentialist mode of categorization is more likely than a similarity one for stimuli which are in better correspondence with our naïve understanding of the world.

Keywords: categorization; critical features; essentialism; general knowledge.

Introduction

A satisfactory account of how humans represent concepts has been elusive despite decades of intense research. Throughout the paper we label as 'similarity accounts' theories such as prototype (Hampton, 1995a) and exemplar theory (Kurtz, 2007; Nosofsky, 1989). Similarity accounts share an assumption that when we categorize a novel object most of its features typically play a role: some features may be more important than others, but no feature is so important that its presence can guarantee or preclude classification into a category[1] (cf. Pothos, 2005). Similarity accounts have led to sophisticated and compelling models. However, they do not appear adequate to cover the range of categorization results.

An excellent example of the (apparent) inadequacy of similarity accounts has been provided from the transformational paradigm, which is the focus of the present research. In Rips's (1989; Keil, 1989) classic study a bird is exposed to toxic waste which leads to changes in its appearance, leaving it looking like an insect. Despite its misfortune, the bird can still successfully mate with other birds ('successfully' in that there were normal bird offspring). The mean categorization ratings provided by Rips's participants indicated the changed bird to be more likely to be a bird, even though it looked nothing like a bird. Therefore, the transformational paradigm *appears* to lead to results which cannot be explained by similarity accounts of categorization (at least for some participants; cf. Hampton, Estes, & Simmons, 2007). Researchers have explored a number of alternative theories to understand such results, for example, the classical view, individually necessary/sufficient features, or essentialism— we refer to these theories as 'beyond similarity' accounts of categorization (Rips, 2001).

Our purpose is to provide a contrast between similarity and beyond similarity accounts of categorization. Our particular emphasis is the transformational paradigm

and essentialism. The key issue is whether results with the transformational paradigm really go beyond similarity accounts.

‘Beyond similarity’ explanations for the transformational paradigm

Participants in Rips’s (1989) study appear to consider ‘mating[2]’ a *sufficient* feature for the category of birds: when the presence of this feature can be established, it guarantees classification into the category of birds. Conversely, *necessary* features are essential before an item is considered a member of a category; for example, being ‘male’ is a necessary feature for the category of bachelors. We refer to necessary and sufficient features as ‘critical’ features, in that a classification outcome can hinge on the presence or absence of such an individual feature. Critical features originate from the classical view of categorization, according to which concepts are psychologically represented as definitions, that is, a set of individually necessary and jointly sufficient features (Katz, 1972).

Definitions imply categorization processes that are discrete, unambiguous, and allow awareness of the actual definitions, but none of these predictions have been empirically supported (Barsalou, 1985; Fodor et al., 1980; Hampton, 1998). Some researchers have examined the evidence for the existence of individual necessary or sufficient features. Rips’s (1989) results have been widely cited as evidence that sufficient features do exist, but other studies have failed to provide support for the existence of critical features (Hampton, Estes, & Simmons, 2007; Larochelle, Cousineau, & Archambault, 2005; Pothos & Hahn, 2000).

Can critical features be understood as ‘very important’ features? Several researchers have provided sophisticated models to predict which features are more important. For example, Sloman, Love, and Ahn (1998; see also Gelman & Wellman,

1991; Rehder, 2003, 2007) suggested that internal features are more important in the representation of creatures, since altering such features would result in several other features being changed (feature mutability); accordingly, internal features would have a greatly enhanced influence on category membership. Probabilistic measures provide another way to quantify a feature's importance, in terms of whether a feature is predictive of category membership and vice versa (Corter & Gluck, 1992). However, a critical feature is qualitatively different from a very important one: the former guarantees a classification outcome, the latter simply has a lot of influence on one. At least, this is the difference between a very important feature and a critical one, *in principle*; some researchers have argued that this is a false distinction (Hampton, 1995b).

Ultimately, one can argue that a critical feature for a creature to be a cat is to have the cat's DNA (for example). However, the majority of people will have a very vague idea of what DNA is. So how can classification be guided by a feature which is not only vague but probably never directly observed as well? This leads us to the intriguing hypothesis of psychological essentialism. According to philosophical essentialism, the members of natural concepts share an essence (hidden, underlying characteristics) that makes them what they are (Putman, 1975; Kripke, 1972). Regarding psychology, although we are typically unaware of what such essences are, we appear to behave in a way that reflects belief in their existence (e.g., Diesendruck, & Gelman, 1999; Gelman, 2004; Gelman & Wellman, 1991; Malt, 1990; Medin & Ortony, 1989; cf. Cleeremans, 2005). Of course, there are more specific versions of essentialist theory (e.g., Rhemtulla, & Xu, 2007; Rips et al., 2006), however, in this work we are only interested in the general assumption of essentialist approaches, as above.

Psychological essentialism has not been without controversy (henceforth, we will refer to psychological essentialism as just essentialism). For example, will something be called ‘water’ if it is known to be composed of H_2O? Malt (1994; cf. Braisby, Franks, & Hampton, 1996) found otherwise. She compiled a list of liquids that contain H_2O and liquids that resemble water (e.g., by being odorless, clear, etc.) but would not contain H_2O. Malt found that belief in whether H_2O was present affected only partly classification (for example, ‘tea’ would not be called water). However, Ahn et al. (2001) suggested that Malt’s findings were the result of ambiguity in the usage of the term ‘water’, which could have confused Malt’s participants and likewise interpretation of her results.

Essentialism provides a compelling account of how the *assumptions* made about an object can affect its classification. Since Murphy and Medin (1985; cf. Heit, 1997; Pickering & Chater, 1995), researchers have been appreciating the importance of general knowledge in categorization. We consider essentialism to be a particular way to formalize general knowledge effects in categorization (although note that there are alternative interpretations to essentialist theory, which suggest no role of general knowledge in essentialist assumptions). Note, finally, that an interesting way to understand essentialism is as a mechanism which determines whether the cognitive system should rely on perceptual information, or other information, when processing new experience.

Any empirical demonstration involving essentialism is complicated by the fact that essences are generally not known. So, how could we contrast similarity accounts of categorization with essentialism? Even though we cannot manipulate essences directly, we can manipulate the circumstances under which we might expect essences

to play a role. Any of the suggestions below are not meant to be 'hard and fast,' but rather reflect typical expectations in a sample of university undergraduates.

First, let's consider the role of familiarity in making assumptions about the essence of a (mammal-like) creature. Consider a cat, a creature about which most of us have quite extensive general knowledge. If we observe the offspring of a particular cat to look nothing like a cat, then we know that something has gone wrong; we would probably assume that the offspring are no longer cats, that is, that their essence has changed. If we encounter a deep-sea fish we have never seen before, and then we observe its offspring, then regardless of how these look we are unlikely to make any inference regarding whether the essence of the creature has changed or not. It could be the case that the offspring typically look nothing like the adult, or that it's a strange creature which changes its form very rapidly from generation to generation; we would have no idea, and, therefore, our lack of familiarity with a creature, would prevent us from making any inferences about whether the essence has changed or not. This example, taken to an extreme, leads us to the manipulation of Experiments 1a and 1b, where we assume that essentialist assumptions would be less likely to be made for transformed space aliens. This is not to say that people do not believe deep-sea fish or space aliens to have essences. Rather, we suggest that making assumptions about whether the essence of a creature has changed or not (e.g., as a result of a certain transformation) is *more likely* when we are more familiar with this creature, compared to when we are not familiar with a creature (cf. Diesendruck & Gelman, 1999; Kalish, 2002; Rips, 2001)

Second, let's consider a typical earth animal and let's assume that our participants will believe that its current form has been determined through evolutionary change. Quoting from Wikipedia's entry for 'evolution', we have:

"Evolution is a result of the cumulative changes that occur in a population of organisms over *succeeding generations*. …Over many generations, the process of the random introduction of new variations and the non-random processes of natural selection can lead to *the formation of new species* from ancestral life forms" (our italics). A change in species corresponds to a change in essence. Therefore, we can state the following implication: we would normally expect that, if at all, the essence of earth creatures would change slowly, across generations, as a result of accumulating evolutionary change. Change of form *together* with gradual change, spanning generations, appear to be the prerequisites for assuming that the essence of an earth creature will change. This is all the more so since there are many instances when creatures might change their appearance abruptly, without any corresponding change in essence, for example, a person losing his hair or a snake shedding its skin. Such changes can be effected as part of the normal life of different creatures (Gelman & Wellman, 1991).

The above two points are the main theoretical assumptions in our work, and their validity is to be assessed against our empirical results. In other words, we suggest that categorization on the basis of essences would be more likely for familiar creatures than unfamiliar ones, and essences are assumed to change through the gradual processes of evolution. These two points represent our attempt to identify implications from the essentialist approach to categorization, which can be directly examined in an empirical way.

Similarity explanations for the transformational paradigm

Can we predict changes in identity in the transformational paradigm from a similarity account? There is evidence suggesting that when an object is transformed gradually it

is more likely to be considered in the same category as the original one, as the gradual transformation establishes a similarity chain (Heit, 1992; cf. Stewart & Brown, 2004). For example, Zaki and Homa (1999) used categories defined either through systematic or random changes in a prototype. The former categories were superior to the latter, across a variety of empirical measures, such as resistance to decay, generalization, and learning rate.

Empirically, the above intuitions can be examined in the following way. Consider a 'fast change' condition, whereby object $A_1A_2A_3A_4$ is transformed to $C_1C_2C_3C_4$ (where the A's and C's refer to features) in one step. Also, in a 'slow change' condition, $A_1A_2A_3A_4$ changes first to $A_1A_2C_3C_4$ and then to $C_1C_2C_3C_4$. A similarity perspective suggests that object identity is *more likely* to be preserved in the slow condition, since in that case a similarity chain can be established between the original and the transformed object. The intermediate object provides a basis for appreciating the compatibility between the original and the (final) transformed object, thus making it more likely that the two will be perceived to be of the same kind.

Existing research with the transformational paradigm has not considered the potential role of similarity chains (e.g., Rips, 1989; Blok, Newman, & Rips, 2005; Rips, Blok, & Newman, 2006). Thus, we cannot know whether identity changes are driven by the presence (or not) of a critical feature, or by the abruptness of the change.

The key issue is that the (general) similarity model for the transformational paradigm makes the opposite categorization prediction from the (general) essentialist model. According to the similarity model, slow changes make it more likely that a similarity chain will be formed between the initial and the final creature, so that the two will be categorized in the same way. In contrast, according to the essentialist model, slow changes are more likely to correspond to gradual, evolutionary-style

development, so that the final creature will end up having a different essence from the original one.

Experiment 1a

Our objective was to provide a basic test of the relevance of similarity chains in the transformational paradigm. We sought stimuli for which there would be no general knowledge expectations and, therefore, any putative influence of essentialist assumptions should be minimal: our stimuli were (imaginary) space aliens in a distant planet. So as not to completely prejudice a putative essentialist account of our results, these space aliens were described as 'mammal-like', to provide participants some foothold on how to imagine these creatures (i.e., the creatures did not exist as electromagnetic waves or sound patterns, they were very, very broadly analogous to earth mammals). Note also that this is not to say that our participants would consider space aliens not have essences, rather that they would have very little insight about the essences of space aliens, so that when categorizing them essentialist expectations would have correspondingly little influence.

Participants

Participants were 62 undergraduate students at the University of Winchester, who volunteered to take part in the study. In all experiments the design was between participants and participants were experimentally naïve.

Materials

We created six training items, each of which was composed of the same four features (a square enclosed by a circle with two triangles on the side of a circle and a cross on top of it). In each of the six training items the different features were instantiated with different textures (the same applied to intermediate and final test items). For example,

in one training item the texture of the circle could be a thick continuous line while for another it would be a dotted line. Features were selected so that they were roughly equally salient. The pictorial format of the stimuli was necessitated by the fact that it would be hard to verbally describe transformations in unfamiliar space aliens (note that researchers in the transformational paradigm have employed descriptive and pictorial presentation of materials more or less interchangeably, without identifying any corresponding confound; this issue is further considered below).

Both conditions involved a set of six test items that had a form overall analogous to the training items, but for which all individual features were changed. Thus, the test items were made of a large square, circles on the sides, a star on top, and an inner triangle. In the slow condition an additional set of items were created, that were meant to be intermediate to the training items and the final test items. These intermediate items had only two features changed, relative to the training items (e.g., the large circle was changed to a large square and the side triangles were changed to circles). There were three different sets of intermediate items, in which different pairs of features were changed, relative to the training items (each participant received only one set of intermediate items). An example of a training, corresponding intermediate test, and corresponding final test item is shown in Figure 1. Each stimulus was printed individually on an A4 sheet.

----------Figure 1----------

Procedure

Participants were given a sheet with a story of an imaginary space exploration, that they were supposed to be covering as journalists, during which certain mammal-like

life forms, called Tristals, were discovered on a distant planet. Participants read that scientists had devised a schematic representation of the Tristals, that consisted of different features, such as squares, etc. The training items, that participants subsequently received (in a folder), were meant to correspond to examples of the Tristal representations.

In the slow condition, participants were then told that exhaust fumes from the earth space ship had led to some contamination that affected the Tristals. As a result, in the first instance Tristals changed partially. At that point, participants received another folder with the six intermediate test items. Participants were then informed that Tristals changed further so that they looked nothing like they originally did. Nevertheless, in their changed state Oxe-Tristals (this is the name that was given to the changed Tristals) appeared to be able to carry on living. Participants were presented with the final set of test items and were prompted to decide whether they would say, as journalists, that Oxe-Tristals were Tristals; the question was straightforward and did not involve references to notions which may have confused participants (e.g., typicality vs. categorization). The 'journalist' manipulation was added to encourage participants to answer on the basis of their intuition. The fast condition was the same as the slow condition, but there were no intermediate items. In this and the other experiments, no other information was collected from participants, to avoid any contamination of the categorization judgments. Participants were tested individually and the experiment lasted for about five minutes. Verbatim instructions are presented in Appendix 1.

Results and discussion

The results of all experiments are summarized in Table 1. Participants were more likely to consider Oxe-Tristals the same as Tristals in the slow condition than in the fast one: chi square (1) = 4.17, p = .041. Note that, as has been the case with other related research (Hampton et al., 2007), there was quite a lot of division in opinion regarding the categorization of the transformed creatures. A source of such division in opinion could be, for example, that some participants interpreted the space aliens in anthropomorphic terms (or in a way that corresponded to earth life forms). For such participants we would expect more 'same' responses in the fast condition and 'different' responses in the slow condition. While there were many such participants, this was not the general trend in responding: The key empirical result is that there was more of a chance of a 'same' judgment in the slow condition compared to the fast one.

These results provide support for the notion that when a similarity chain is established between an original and a transformed object, the two objects are more likely to be considered the same (Heit, 1992; Zaki & Homa, 1999). Accordingly, a similarity account of categorization appears an adequate explanation for participants' behavior.

----------Table 1----------

Experiment 1b

Experiment 1a is important in that it illustrated that similarity chains can be formed, and can affect categorization, in the transformational paradigm. If similarity chains could not influence categorization even in Experiment 1a, then there would be no point in further investigating a putative contrast between similarity chains and

essentialist assumptions in the transformational paradigm. Experiment 1b is as closely matched to Experiment 1a as possible, but for the following difference: in Experiment 1b, changes as a result of the toxic accident took place more slowly and were evident in subsequent generations. Moreover, the changed creatures were not assigned a unique name (which may have reinforced the perception of their being in a different category), but rather they were called the 'offspring of the afflicted Tristals'. This scenario is more consistent with putative essentialist assumptions, since changes in subsequent generations could be interpreted as resulting from changes in the creatures' DNA and, therefore, changes in the creatures' essence. Thus, Experiment 1b involved a scenario of (effectively) evolutionary change, albeit a much accelerated one (as a result of the toxic accident). Accordingly, with Experiment 1b we sought a more critical examination of the contrast between similarity chains and essentialist assumptions in the categorization of space aliens.

Participants

Participants were 84 Swansea University (mostly undergraduate) students who took part in the study for a small payment. The design was between participants.

Materials and Procedure

Materials and procedure were identical to these in Experiment 1a, but for the following change: the creatures of interest were the *offspring of the affected Tristals*, which could change either in one generation (fast condition) or in two generations (slow condition). In the slow condition, participants were presented with the intermediate stimuli as in Experiment 1a.

Results and Discussion

The results of Experiment 1b are nearly identical to the results of Experiment 1a (Table 1). Participants were more likely to consider the offspring of the affected

Tristals the same as Tristals in the slow condition than in the fast one: chi square (1) = 7.55, p = .006.

Experiment 1b was meant to be a more conservative test of the role of similarity chains in the transformational paradigm: we did provide a scenario of evolutionary change, but the creatures were still space aliens (general knowledge assumptions were not expected). The results were clearly in favor of the similarity account in the transformational paradigm: where it was possible to establish a similarity chain (in the slow change condition), the transformed and the original creature were categorized in the same way.

Experiment 2a

In Experiments 1a and 1b we have verified the importance of similarity chains in the transformational paradigm. The key reason why we expected the similarity explanation to be the dominant one in these experiments is that for space aliens essentialist assumptions should be less likely to be made. In other words, the average participant would not have enough familiarity with space aliens to allow assumptions about their putative essences to influence categorization decisions. We examine the validity of this conclusion in Experiment 2a (and 2b).

The aim of Experiment 2a was to create a task analogous to that of Experiment 1a, but where essentialist assumptions about category membership could more obviously bring to bear. This was achieved by eliminating the scenario involving space aliens and introducing a common earth animal, specifically crows. Crows were selected because they are a basic level category all our participants would be familiar with, and because they do not undergo any particularly drastic transformations in their

lives, short of what is typically expected of birds in general (cf. Diesendruck & Gelman, 1999; Estes, 2003).

Participants

Participants were 46 undergraduate students of Cardiff University, who took part in the study for course credit or a small payment; some of the participants were tested in other, unrelated experiments in the same session. The design was between participants.

Materials

The materials were a hypothetical scenario about how a toxic accident in South America (cf. Rips, 1989; Appendix 2) affected the crows in the area. Participants were told that the effects of the toxic accident resulted in changes in the offspring of the afflicted crows, either in one generation (fast condition) or over several generations (slow condition). In this case, we presented the changes of the creatures verbally and relied on our participants to imagine these; the complexity of the changes was such that a convincing pictorial presentation would have been unworkable (this was directly verified with a pilot study, outlined below).

No special name was given to the transformed crows, they were just referred to as 'the offspring of the afflicted crows' (as was the case in Experiment 2b). Once the descendants of the afflicted crows stopped changing, in both the fast and the slow condition, they could happily go on with their lives in their altered states. The changes themselves were such that the transformed creature would look nothing like the original one.

Procedure

We tried to replicate the procedure of Experiments 1a/1b. Participants were given A4 sheets with the story of the toxic waste accident and its subsequent effects on the

crows in the area. The only difference between the fast and the slow conditions was that in the former case the crows changed over one generation, while in the latter case the changes were gradual and took place over several generations. Note that we did not attempt to ensure that the changes between the original and the intermediate creature were exactly equivalent to these between the intermediate and the final creature; we were only interested in whether *exactly the same* overall changes would lead to differences in responding between the slow and the fast condition.

Once participants had read the story, they were asked to imagine how the transformed creatures looked and to classify them as crows or not on the basis of their intuition. Participants were told that there were no correct answers and no justification had to be provided, so as to prevent them from trying to guess a 'correct' answer. Participants indicated their response by circling a corresponding YES or NO in the sheets they received. They were tested individually and the experiment lasted for about five minutes.

Results and discussion

The results of this experiment are summarized in Table 1. Participants were more likely to consider the transformed crows as crows in the *fast* condition than in the slow one: chi square (1) = 4.39, p = .036.

The result of Experiment 2a is opposite from the results of Experiments 1a/1b. The most prominent difference between Experiment 2a and Experiments 1a/1b is that in the former case an earth creature was considered, while in the latter space aliens. Accordingly, we suggest that an essentialist mode of categorization is more likely to be adopted with earth creatures, and one based on straightforward similarity overlap with creatures for which little is known. Of course, there are other differences between Experiments 1a/1b and 2b, for example, the fact that in one case the stimuli

were presented pictorially, in the other verbally. We do not believe such a difference to be confounding. First, as noted, several researchers have used descriptive and pictorial stimuli interchangeably, without identifying any confounding influence on participants' performance (e.g., Rips et al., 2006; Blok et al., 2005). Second, there is considerable evidence from the imagery literature that the process of imagining a stimulus (as our participants were asked to do for the changed crows) is equivalent to the process of actually perceiving the stimulus (e.g., Ishai & Sagi, 1995). Third, arguably, it is not possible to fully counterbalance format of presentation, without destroying the manipulation. When it comes to transformations of real, earth creatures, we believed that the only effective way to communicate such information would be in a verbal way (so that participants could imagine the changes in a semi-plausible way). By contrast, when it comes to space aliens, we felt that presenting information about fantastical creatures in a verbal way would have been too cumbersome and awkward.

To partly assess this issue, we ran a pilot study matched to Experiment 2a, but where the stimuli were presented in a pictorial way. Briefly, participants were given the same scenario of a toxic accident. The original creatures were crows, and participants were initially shown a picture of a crow. Using morphing software (Morpheus photo morpher), we created an 'imaginary' creature as a blend between a dog and a bear (the transformed creature could not have been another recognizable earth creature). Then, we created two intermediate blends between the crow and the dog-bear (we used two intermediate blends in order to make the relation between the initial and the final creature more salient). Participants in the slow condition saw first the two intermediate blends and subsequently the final dog-bear. Participants in the fast condition just saw the dog-bear. Regardless of condition, the overwhelming

majority of participants decided that the final creature was not the same as the original one (in the slow condition, 31 ‘different’ 4 ‘same’, in the fast condition 33 ‘different’, 3 ‘same’). This finding can be interpreted in several ways. First, we would argue that when participants saw a picture of a transformed earth crow, the picture was so strange that they all decided the original and the transformed crow had nothing to do with each other. By contrast, the verbal version of the experiment allowed participants to imagine the transformation in a more plausible way and so that the ceiling effects obtained in the pilot were avoided. In other words, the most effective presentation format for Experiment 2a is the verbal format we adopted. A second interpretation is that the difference in results was due to the fact that in Experiment 2a the transformation took place in terms of discrete feature changes, but in the pilot in terms of the morphing (and distortion) of continuous dimensions. Accordingly, in the pilot study, the transformed creature effectively involved new, novel dimensions, so different to those of the original crow that participants considered the two creatures different. In conclusion, even though we think the first interpretation is more plausible, additional experimental work is needed before this issue can be more convincingly settled.

In Experiment 2b we tried harder to ensure that there would be enough scope for a similarity chain to be established between the original and the transformed creatures. This was achieved by elaborating the procedure to emphasize the intermediate creature in the slow change condition.

Experiment 2b

In this experiment we tried to encourage the formation of a similarity chain by asking participants, in the slow condition, to imagine and categorize the intermediate creatures, as well as the final ones.

Participants

Participants were 102 undergraduate students of Cardiff University, who took part in the study for course credit; some of the participants were tested in other, unrelated experiments in the same session. The design was between participants.

Materials

As before, participants received a hypothetical scenario about how a toxic waste accident in South America resulted in transformations in the local population of crows.

Procedure

We followed the procedure of Experiment 2a, but for the following difference: in the slow condition, after the first set of changes in the offspring of the afflicted crows (across two generations), participants were asked to imagine the transformed crows and to decide whether they would be called crows or not; at that point, participants did not know that they would have to provide a second classification judgment. Subsequently, once the offspring of the afflicted crows underwent their second (and final) set of changes (across several generations), participants were asked again whether the (further) transformed crows would be called crows or not. By having participants carry out an intermediate judgment we hoped to encourage the formation of a similarity chain between the initial and final creature. As in Experiment 2a, participants were asked to indicate their responses by circling either YES or NO in the sheets they received. Participants were tested individually and the experiment lasted for about five minutes.

Results and discussion

In the slow condition participants provided two classification judgments, one for the intermediate creatures and one for the final creatures. We first examined responses to the intermediate creatures. Out of the 51 participants in the slow condition, only one responded that the intermediate creature was not a crow. We eliminated the data from this participant and proceeded to compare the final responses in the slow condition with the ones in the fast condition. The results are shown in Table 1. As in Experiment 2a, participants were more likely to consider the transformed crows as crows in the fast condition than in the slow one: chi square (1) = 24.5, $p < .0005$.

In Experiment 2b, our intention was to create a methodology which would force participants to think more of the intermediate creatures; the intermediate classification in the slow condition aimed to achieve exactly this objective. Moreover, the logic of similarity chains seems fairly straightforward: if there is an intermediate representation between an initial and a final one, the initial and the final one would be more likely to be categorized in a similar way. Of course, in considering the logic of Experiment 2b, the effect of the intermediate judgment could be more complex than the formation of an intermediate representation. The key point is that, despite such complexities, the results of Experiment 2b were qualitatively identical to the results of Experiment 2a. Since participants still considered the slow change as more likely to lead to a change in identity, it looks like their categorizations were predominantly influenced by essentialist assumptions regarding the crows, as opposed to similarity chains. An essentialist account appears essential.

General discussion

The transformational paradigm has been widely used to support 'beyond similarity' theories of categorization, such as critical features and essentialism. Can one provide a similarity account for the transformational paradigm? Our experimental design was based on the notion of similarity chains (Heit, 1992) and the intuition that creature essences are more likely to change in a slow, evolutionary-style way, rather than abruptly. In our design, generalization on the basis of similarity chains would be different from generalization on the basis of essentialism. We consider the merit of the present work exactly this, that it allows a specific contrast of similarity and essentialist approaches in categorization, in a way that is independent of the implementation details of specific essentialist or similarity theories.

With stimuli for which there would be few general knowledge expectations (space aliens), the similarity account was sufficient to explain our results. With earth creatures, results were consistent with an essentialist account of classification. We conclude that essentialism has to remain an integral part of our understanding of categorization processes, especially with real life concepts (cf. Estes, 2003; Kalish, 2002).

Other influential 'beyond similarity' accounts are Rips et al.'s (2006) causal continuity hypothesis and sortalism (e.g., Rhemtulla & Xu, 2007). According to causal continuity, whether the identity of a transformed object is preserved or not depends on establishing a *causal* connection between the original and transformed object. A sortal is a collection of objects which are created, exist, and are destroyed in similar ways. Both causal continuity and sortalism are related to essentialism, however, our results are not specific enough to tease apart such possibilities. For

example, a crucial distinction in related empirical tests is between unique individuals and categories—however, we considered only the latter in our research.

How defensible is the transformational paradigm in general? Blok et al. (2005, p.20) summarize the justification in an eloquent way: "Our objective in using the transformation paradigm was to investigate what is necessarily true of object identity, not just what is customarily true; so some departure from business-as-usual is essential in order to check our hypotheses at all. Moreover, although we hope our participants have not experienced such transformations first hand, we believe that speculation about the effects of similar changes is quite natural and familiar. Closely related scenarios figure prominently in fiction and even in non-fiction." We agree entirely.

More generally, the nature of the transformational paradigm (and essentialist theory) is such that it is difficult to create studies controlled to the same degree of rigor and sophistication as studies, e.g., involving prototype and exemplar theory (e.g., cf. Nosofsky & Zaki, 2002). The recent exchange between Malt and Sloman (2007) and Bloom (2007) amply illustrates this issue. It is indeed difficult to fully match conditions involving, on the one hand, space aliens in a distant planet and, on the other hand, earth crows. Nonetheless, certain considerations should argue in favor of the robustness of our conclusions. First, we employed specific features, avoiding possible problems with instantiations of abstract features (Markman & Maddox, 2003). Second, we considered generalization only from positive exemplars, avoiding complications relating to generalization from negative evidence (cf. Kalish & Lawson, in press; Chin-Parker & Ross, 2004). Third, we were careful to consistently use in all experiments classes and not a mixture of classes/individuals, since several

investigators predict differences in generalization from individuals vs. classes (Gutheil & Rosengren, 1996; Rips et al., 2006).

In sum, we hope to have provided a reasonably well-controlled examination of the scope of applicability of essentialist theory, relative to similarity-based accounts.

Acknowledgements

This research was partly supported by ESRC grant R000222655 and EC Framework 6 grant contract 516542 (NEST). We would like to thank Nick Chater, James Close, Zach Estes, James Hampton, Serge Larochelle, Amotz Perlman, and Ilias Tzimkas for their help/ comments at various stages of this project. Thanks also to Darren Edwards and Liv Kosnes for collecting the data for the pilot study exploring a pictorial version of Experiment 2a.

References

Ahn, W., Kalish, C., Gelman, S. A., Medin, D. L., Luhmann, C., Atran, S., Coley, J. D., & Shafto, P. (2001). Why essences are essential in the psychology of concepts. Cognition, 82, 59-69.

Barsalou, L. W. (1985). Ideals, central tendency and frequency of instantiation as determinants of graded structure in categories. Journal of Experimental Psychology: Learning, Memory and Cognition, 11, 629-654.

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, DC: American Psychological Association.

Bloom, P. (2007). More than words: A reply to Malt and Sloman. Cognition, 105, 649-655.

Braisby, N., Franks, B., & Hampton, J. (1996). Essentialism, word use, and concepts. Cognition, 59, 247-274.

Chin-Parker, S. & Ross, B. H. (2004). Diagnosticity and prototypicality in category learning: A comparison of inference learning and classification learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 30, 216-226.

Cleeremans, A. (2005). Computational correlates of consciousness. Progress in Brain Research, 150, 81-98.

Corter, J. E. & Gluck, M. A. (1992). Explaining Basic Categories: Feature Predictability and Information. Psychological Bulletin, 2, 291-303.

Diesendruck, G. & Gelman, S. A. (1999). Domain differences in absolute judgments of category membership: Evidence for an essentialist account of categorization. Psychonomic Bulletin & Review, 6, 338-346.

Estes, Z. (2003). Domain differences in the structure of artifactual and natural categories. Memory & Cognition, 31, 199-214.

Fodor, J. A., Garrett, M. F., Walker, E. C., & Parkes, C. H. (1980). Against definitions. Cognition, 18, 263-367.

Gelman, S. A. (2004). Psychological essentialism in children. Trends in Cognitive Sciences, 8, 404-409.

Gelman, S. A. & Wellman, H. M. (1991). Insides and essences: Early understanding of the non-obvious. Cognition, 38, 213-244.

Gutheil, G. & Rosengren, K. S. (1996). A rose by any other name: preschooler's understanding of individual identity across name and appearance change. British Journal of Developmental Psychology, 14, 477-498.

Hampton, J. A. (1995a). Similarity-based categorization: The development of prototype theory. Psychologica Belgica, 35, 103-125.

Hampton, J.A. (1995b). Testing the prototype theory of concepts. Journal of Memory and Language, 34, 686-708.

Hampton, J. A. (1998). Similarity-based categorization and fuzziness of natural categories, Cognition, 65, 137-165.

Hampton, J. A., Estes, Z., & Simmons, S. (2007). Metamorphosis: essence, appearance, and behavior in the categorization of natural kinds. Memory & Cognition, 35, 1785-1800.

Heit, E. (1992). Categorization using chains of examples. Cognitive Psychology, 24, 341-380.

Heit, E. (1997). Knowledge and Concept Learning. In K. Lamberts & D. Shanks (Eds.), Knowledge, Concepts, and Categories (pp. 7-41). London: Psychology Press.

Ishai, A. & Sagi, D. (1995). Common Mechanisms of Visual Imagery and Perception, Science, 268, 1772-1774.

Kalish, C. W. (2002). Gold, Jade, and Emeruby: The value of naturalness for theories of concepts and categories. Journal of Theoretical and Philosophical Psychology, 22, 45-56.

Kalish, C. W. & Lawson, C. A. (in press). Negative evidence and inductive generalization. Thinking and Reasoning.

Katz, J. (1972). Semantic theory. New York: Harper & Row.

Keil, F. C. (1989). Concepts, kinds, and cognitive development. Cambridge, MA: MIT Press.

Kripke, S. A. (1972). Naming and necessity. Cambridge, MA: Harvard University Press.

Kurtz, K. J. (2007). The divergent autoencoder (DIVA) model of category learning. Psychonomic Bulletin & Review, 14, 560-576.

Larochelle, S., Cousineau, D., & Archambault, A. (2005). Definitions in categorization and similarity judgments. In H. Cohen & C. Lefebvre (Eds.) Handbook of Categorisation in Cognitive Science. Amsterdam: Elsevier, p. 278-303.

Malt, B. C. (1990). Features and Beliefs in the Mental Representations of Categories. Journal of Memory and Language, 29, 289-315.

Malt, B. C. (1994). Water is not H2O. Cognitive Psychology, 27, 41-70.

Malt, B. C. & Sloman, S. A. (2007). More than just words, but still not categorization. Cognition, 105, 656-657.

Markman, A. B. & Maddox, W. T. (2003). Classification of exemplars with single- and multiple-feature manifestations: The effects of relevant dimension variation and category structure. Journal of Experimental Psychology: Learning, Memory, and Cognition, 29, 107-117.

Medin, D. L. & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou and D. L. Ortony (Eds.) Similarity and analogical reasoning. Cambridge: Cambridge University Press.

Murphy, G. L. & Medin, D. L. (1985). The Role of Theories in Conceptual Coherence. Psychological Review, 92, 289-316.

Nosofsky, R. M. (1989). Further tests of an exemplar-similarity approach to relating identification and categorization. Journal of Experimental Psychology: Perception and Psychophysics, 45, 279-290.

Nosofsky, R. B. & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. Journal of Experimental Psychology: Learning, Memory, & Cognition, 28, 924-940.

Pickering, M. & Chater, N. (1995). Why cognitive science is not formalized folk Psychology. Minds and Machines, 5, 309-337.

Pothos, E. M. (2005). The rules versus similarity distinction. Behavioral & Brain Sciences, 28, 1-49.

Pothos, E. M. & Hahn, U. (2000). So concepts aren't definitions, but do they have necessary *or* sufficient features?. British Journal of Psychology, 91, 439-450.

Putnam, H. (1975). The meaning of 'meaning'. In H. Putnam (Ed.) Mind, language and reality: Philosophical papers, vol. 2. Cambridge, UK: Cambridge University Press.

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

Rehder, B. (2007). Essentialism as a generative theory of classification. In A. Gopnik, & L. Schultz, (Eds.), Causal learning: Psychology, philosophy, and computation, pp. 190-207. Oxford, UK: Oxford University Press.

Rhemtulla, M. & Xu, F. (2007). Sortal concepts and causal continuity: Comment on Rips, Blok, and Newman (2006). Psychological Review, 114, 1087-1095.

Rips, L. J. (1989). Similarity, typicality and categorization. In S. Vosniadou and A. Ortony (Eds.) Similarity and analogical reasoning. Cambridge, UK: Cambridge University Press.

Rips, L. J. (2001). Necessity and natural categories. Psychological Bulletin, 127, 827-852.

Rips, L. J., Blok, S., & Newman, G. (2006). Tracing the identity of objects. Psychological Review, 113, 1-30.

Sloman, S. A., Love, B. C., & Ahn, W. (1998). Feature Centrality and Conceptual Coherence. Cognitive Science, 22, 189-228.

Stewart, N. & Brown, G. D. A. (2004). Sequence effects in the categorization of tones varying in frequency. Journal of Experimental Psychology: Learning, Memory, and Cognition, 30, 416-430.

Tversky, A. (1977). Features of Similarity. Psychological Review, 84, 327-352.

Zaki, S. R. & Homa, D. (1999). Concepts and transformational knowledge. Cognitive Psychology, 39, 69-115.

Footnotes

Footnote 1: What we mean here is that there is no mechanism in similarity accounts whereby a single feature logically entails the classification of an object, regardless of the information available from the other features. In other words, however important a single feature is in the classification of an object, there would typically be contributions from other features as well. Models of similarity where individual features/ dimensions are weighted typically make this assumption (Nosofsky, 1989; Tversky, 1977). Of course, it is possible that the weight of all features would be reduced to zero apart from one, but in such cases it is debatable as to whether we still have an overall similarity process (Pothos, 2005).

Footnote 2: Strictly speaking, this feature would be ‘mating with another bird to produce bird offspring’.

Appendix 1. The materials for Experiment 1a.

Welcome to this study! Please ask questions while reading the instructions if anything is not clear.

Space exploration has recently been marked by a major breakthrough: Mammal-like life forms have been discovered on the distant planet ZIR-33, thus, providing an answer to one of humanities' oldest questions. You are a journalist who has accompanied the space mission, with a view to providing media coverage for these exciting events.

The creatures have been called Tristals, after the astronaut who first discovered them. While other noteworthy creatures have subsequently been found on ZIR-33, they have not spurred the same level of enthusiasm as Tristals have, perhaps due to the Tristals' resemblance to some earth mammals.

Biologists have been very excited about Tristals, working day and night to analyse and classify them. They have developed a highly efficient schematic representation of the creatures, that consists of different features, such as squares, circles etc. The features are meant to characterise Tristals' properties. You will shortly see a few examples of these representations: They correspond to six of the first Tristals that have been labelled for further study.

Please study the Tristals as carefully as possible. When you have finished, please let the experimenter know.

A disaster has befallen the area where astronauts first landed on ZIR-33. Possibly as a result of the engine emissions of the earth spaceship, the composition of atmosphere in the landing area has been severely affected, making it very apparently toxic for the Tristals there. Scientists have tried to keep track of the effects of this

accident on Tristals, to find that, over a short period of time, Tristals have changed quite dramatically. The afflicted Tristals have been called Oxe-Tristals (Oxe was the code name for the spaceship the astronauts came in).

[Continue with the following passage for the Slow Condition.]

In the first instance, Oxe-Tristals only partially changed. Please ask the experimenter for a folder showing how Oxe-Tristals looked in the first phase of their affliction. When you have finished seeing the Oxe-Tristals, please let the experimenter know.

Subsequently, over the course of only a few days on ZIR-33, the Oxe-Tristals further changed, so that they were almost nothing like what they originally looked. After that, however, the Oxe-Tristals did not appear to further change and, moreover, it looked as if they could carry on living in their altered state.

Please ask the experimenter for some examples of the final state of Oxe-Tristals, and let him / her know when you have finished looking at them.

[Now skip to the Final Passage.]

[Continue with the following passage for the Fast Condition.]

Over the course of only a few days on ZIR-33, the Oxe-Tristals changed, so that they were almost nothing like what they originally looked. After that, however, the Oxe-Tristals did not appear to further change and, moreover, it looked as if they could carry on living in their altered state.

Please ask the experimenter for some examples of the final state of Oxe-Tristals, and let him / her know when you have finished looking at them.

[Now skip to the Final Passage.]

[Final Passage.]

All in all, scientists were very mystified by this situation. The Oxe-Tristals changed so much over a period of a few days that they looked nothing like what they originally did. The question for you is the following: Would you, as a journalist, say that Oxe-Tristals are in fact Tristals, or not?

Appendix 2

The materials for Experiment 2a.

Slow condition.

Welcome to this study! The study is very simple: You will shortly read a short story. Subsequently, you will be asked a simple question relating to the story. You do not have to memorise anything, there is no correct or wrong answer, and the story sheets will be available to you while you respond. When you have read these instructions, please turn to the next sheet.

[Next sheet] United Chemicals' latest factory was built in central South America, as it was easy to find cheap land far away from cities and towns. Unfortunately, shortly after the start of the factory's operation, there was an accident and the entire ecosystem near the factory was polluted. The company was able to rapidly neutralize the toxic substances. However, concerns that the contamination may have affected the genetic profile of some indigenous species led biologists from a local university to monitor the area for a number of subsequent years. What they found was that crows in the area did not undergo any obvious changes immediately following the accident; however, observing subsequent generations of these crows made it clear that the accident's effects were in fact quite severe. Initially, in the first generation, the offspring of the afflicted crows hardly had any feathers. The second generation crows appeared with thicker and much narrower wings. By the third generation, the crows' wings looked almost like legs. Eventually, over a few more generations, the birds were seen to move around like cats or dogs. At that point, the

offspring of the afflicted crows changed no further and they appeared to live happily within their new abilities and physique.

Try to imagine this process of change and how the offspring of the afflicted crows eventually turned out to be. Would your first intuition be to still call these creatures crows or not? Clearly, there is no correct or wrong answer and you do not have to justify your response. Just circle "Yes" or "No": The creatures are crows:

YES NO

Fast condition. As above, but for the following difference:

…What they found was that crows in the area did not undergo any obvious changes immediately following the accident; however, observing the crows' offspring made it clear that the accident's effects were in fact quite severe. The children of the afflicted crows hardly had any feathers. Moreover, they appeared with thicker and much narrower wings, that looked like legs and which they used to move around like cats or dogs. No other changes were observed in the offspring of the afflicted crows, who appeared to live happily within their new abilities and physique. …

Tables

Table 1

Table 1. The results from all experiments. In each experiment participants had to decide whether an original and transformed creature are the same or not. Experiments varied in how the original and the transformed creature were called, as follows: In Experiment 1a we had Tristals vs. Oxe-Tristals, in Experiment 1b Tristals vs. offspring of the affected Tristals, and in Experiments 2a, 2b crows vs. offspring of the afflicted crows.

	Slow same	Slow different	Fast same	Fast different
Experiment 1a	21	10	13	18
Experiment 1b	33	14	14	23
Experiment 2a	10	13	17	6
Experiment 2b	9	41	34	17

Figures

Figure 1. An example of a training, corresponding intermediate, and final test item in Experiment 1a.